Published in final edited form as:

Infancy. 2013 ; 18(4): 534–553. doi:10.1111/j.1532-7078.2012.00135.x.

Increased focus on the mouth among infants in the first year of life: A longitudinal eye-tracking study

Elena J. Tenenbaum, Rajesh J. Shah, David M. Sobel, Bertram F. Malle, and James L. Morgan

Brown University Department of Cognitive, Linguistic and Psychological Sciences. The Center for the Study of Children at Risk, Warren Alpert Medical School at Brown University, Women & Infants Hospital

Abstract

The present study examines face-scanning behaviors of infants at 6, 9, and 12 months as they watched videos of a woman describing an object in front of her. The videos were created to vary information in the mouth (speaking vs. smiling) and the eyes (gazing into the camera vs. cueing the infant with head turn or gaze direction to an object being described). Infants tended to divide their attention between the eyes and the mouth, looking less at the eyes with age and more at the mouth than the eyes at 9 and 12 months. Attention to the mouth was greater on speaking trials than on smiling trials at all three ages, and this difference increased between 6 and 9 months. Despite consistent results within subjects, there was considerable variation between subjects. This raises the question of whether a developmental “norm” of face scanning in infancy ought to be pursued. Rather, these data add to emerging evidence suggesting that individual differences in face scanning might reliably predict aspects of later development.

Keywords

Infants; face scanning; eye tracking

Numerous researchers have documented infants’ interest in faces (Fantz, 1963; Goren, Sarty, & Wu, 1975; Johnson, Dziurawiec, Ellis, & Morton, 1991; Turati, Valenza, Leo, & Simion, 2005). Researchers have also asked *where* on the face infants look, particularly during communicative acts. During communication and interaction, infants’ gaze tends to focus on the eyes in early infancy (Haith, 1977; Maurer & Salapatek, 1976), but on the mouth in later infancy and toddlerhood (Frank, Vul, & Saxe, 2011; Hunnius & Geuze, 2004; Lewkowicz & Hansen-Tift, 2012; Nakano et al., 2010; von Hofsten, Uhlig, Adell, & Kochukhova, 2009), perhaps due to infants’ emerging interest in language (Hunnius & Geuze, 2004; Lewkowicz & Hansen-Tift, 2012). One question that has emerged is how age and interlocutor behavior might interact in face-scanning behaviors. The goal of this research is to explore how attention to faces changes as a function of age and interlocutor behavior between 6 and 12 months of age and the extent to which individual infants differ in their face scanning behaviors across infancy.

Early studies of face scanning showed that within the first few months of life, infants look less at external features and more at internal features of the face, and within the face, tend to focus on the eyes (e.g. Haith, Bergman, & Moore, 1977; Maurer & Salapatek, 1976).

Recently, researchers have identified increases in attention to the mouth over the first year of life. Hunnius and Geuze (2004) showed that infants peak in focus on the eyes at approximately 3.5 months before focusing more on the mouth by 6 months. Chawarska, Shic, Bradshaw, Macari and Klin (2009) found a similar increase in attention to the mouth between 6 and 12 months. Oakes and Ellis (2011) also found that infants move from attending to the eyes towards a more distributed scanning of the face in the second half of the first year of life.

Though Lewkowicz and Hansen-Tift (2012) recently demonstrated that a return to focus on the eyes begins occurring by 12 months, others have demonstrated a preference for the mouth in the second year (Frank et al., 2011), with some demonstrating greater than 90% of trial time spent fixating this area (Nakano et al., 2010; von Hofsten et al., 2009). By two years, lack of attention to the eyes has been associated with pathological development and specifically with severity of social impairment in toddlers diagnosed with autism spectrum disorders (Jones, Carr, & Klin, 2008).

The effect of age on face scanning is somewhat difficult to interpret from these results, however, because the types of stimuli used varied substantially across studies. Evidence for attention to the mouth has been largely demonstrated in videos of people talking (Frank et al., 2011; Hunnius & Geuze, 2004; Lewkowicz & Hansen-Tift, 2012; Nakano et al., 2010; von Hofsten et al., 2009) where as evidence for attention to the eyes tends to emerge for static images or still face videos (Chawarska & Shic, 2009; Oakes & Ellis, 2011). In some cases, a preference for the upper half of the face has been shown for videos in which the speaker's mouth was moving, but no sound was presented (Liu et al., 2011; Wheeler et al., 2011). This pattern was also observed by Chawarska, Shic, Bradshaw, Macari & Klin (2009). Chawarska and colleagues addressed the interaction between age and stimulus type by showing 3-, 6-, 9- and 12-month-olds videos of someone speaking, dynamic videos without speaking, and static images. Across the age ranges, infants looked less at the eyes for speaking than for non-speaking dynamic videos and less at the eyes for dynamic videos than for static images (and vice versa for attention to the mouth). This study also found decreased attention to the mouth between 9 and 12 months for typically-developing infants, but no such pattern among unaffected siblings of autistic infants.

Because of inconsistencies in the published literature, we are left with several unanswered questions regarding the effects of age and interlocutor behavior on face scanning. Specifically, how do age and interlocutor behavior interact in face scanning, and how consistent are face scanning behaviors within and between infants? To address these questions, the current study used a longitudinal design in which 6-, 9- and 12-month-old infants saw four types of video stimuli. The stimuli were counter balanced for presence or absence of linguistically relevant information in the mouth (speaking vs. smiling) and eyes (gaze shift towards an object or gaze into camera). This design provided a range of stimuli from highly dynamic (speaking and shifting attention towards a target object) to nearly still (smiling and gazing at the camera while an audio track played from a speaking trial).

Two sets of stimuli were used in this experiment. In one, the speaker shifted her gaze to the object on gaze to object trials while in the other, she shifted her head and gaze to the object (thus making the gaze shift a redundant and less relevant cue). This design allowed us to explore specific contributions of four factors on face scanning behaviors in infancy: presence of information in the mouth, presence of information in the eyes, relevance of the gaze-following cue, and age. We expected that infants' face scanning would increasingly optimize the likelihood of extracting rich information available in the speaker's face. Specifically, we expected infants to look more at the mouth when the speaker was speaking because the moving mouth is both salient and informative and more at the eyes of the

speaker when she was indicating the target location with her gaze because this would help inform their understanding of the situation. We expected greater attention to the eyes for the gaze shift only condition when that would be the sole cue to the target location. Finally, because infants are producing more complex and phonologically language-specific utterances over the second half of the first year (Adamson, 1995) and because the mouth carries information about how particular speech sounds are formed, we hypothesized that infants' attention to the mouth would increase between 6 and 12 months.

The use of face-scanning and gaze-following measures in this design was intended to examine correlations among these behaviors and their relation to later language development. These results are presented elsewhere (Tenenbaum, Sobel, Malle, Sheinkopf, & Morgan, 2012). Here, we focus on the results of infants' face scanning to consider how such scanning behaviors change within and across infants.

Method

Participants

Infants were recruited from public birth records. All were born full-term to monolingual English speaking families and had no known developmental or hearing disabilities. Families were compensated with a toy, book, or t-shirt at each visit. Families were informed that the study would involve visit at 6, 9, and 12 months, but were not excluded from participating if they could not commit to a longitudinal study or if data collection had failed at a previous visit.

At 6 months, 125 infants visited the lab. Data were collected successfully for 97 infants ($M = 198$ days, $SD = 11$ days; 53 male, 44 female). Additional infants were tested but excluded from the analyses for the following reasons: could not calibrate eye tracker ($n = 9$), fewer than half the trials successfully tracked ($n = 16$), non-native speaker of English ($n = 1$), experimenter error ($n = 2$). At 9 months, data were collected successfully for 65 infants ($M = 293$ days, $SD = 12$ days; 36 male, 29 female). Additional infants were tested but excluded from analysis for the following reasons: equipment failure ($n = 10$)¹, could not calibrate ($n = 14$), fewer than half the trials successfully tracked ($n = 16$). At 12 months, data were collected from 62 infants ($M = 396$ days, $SD = 12$ days; 33 male, 29 female). Additional infants were tested but excluded from analysis for the following reasons: could not calibrate ($n = 6$), fewer than half the trials successfully tracked ($n = 7$), equipment failure ($n = 3$). A trial was considered successfully tracked if data were collected for at least 4 of the 8 seconds of trial time. This cutoff was established to exclude trials on which the infant was not attending to the stimuli. By the same logic, to be included in the analysis, infants had to contribute data on at least 8 of the 16 trials.

Although there was substantially more data loss in this study than in other comparable longitudinal eye-tracking studies (e.g. Gredeback, Ficke & Melinder 2010), the rates are inline with other investigations (Lewkowicz & Hansen-Tift, 2012). A number of factors, including the use of independent head tracking through a probe affixed to the infant's head and our requirement that infants attend to at least 4 of the 8 seconds of trial time on at least 8 trials, may have contributed to a slightly higher attrition rate.

¹A flood in the laboratory during this period of data collection damaged the cable that controlled the illuminator on the camera. This made it extremely difficult to calibrate infants' eyes to the system.

Stimuli

Infants saw 16 trials per test session; four trials from each of the four trial types. In each clip, a woman seated behind a table with two objects in front of her described one of the objects in highly animated infant-directed speech. The four trial types were counter balanced for presence or absence of information in the eyes and mouth (see Figure 1). In *Gaze to Object and Speak* trials (Figure 1a), the woman gazed toward the object while describing it. In *Gaze to Object and Smile* trials (Figure 1b), the woman gazed toward the object and smiled while an audio track from a speaking trial was played. In *Direct Gaze and Speak* trials (Figure 1c), the woman fixated the camera while describing the object and in *Direct Gaze and Smile* trials (Figure 1d), the woman fixated the camera and smiled while an audio track from a speaking trial was played.

Two stimulus sets were used across subjects. In one set, the head-turn stimuli, the woman shifted her head and gaze toward the target object on gaze to object trials. This resulted in a redundant cue to object location. To eliminate this redundant information and isolate the relevance of the information in the eyes, a second set of stimuli were constructed. In the gaze-shift stimuli, the speaker shifted only her gaze to the target object while maintaining head position facing forward on the two types of gaze-to-object trials (i.e., smiling and speaking). In order to allow the speaker to see the objects while shifting only her gaze, objects had to be moved closer to her face. This resulted in a zooming in on the scene. Infants were assigned to one of the two stimulus types and saw only stimuli of that type at each experimental visit. Thus, if an infant saw head-turn stimuli at 6-months, they also saw those same stimuli at 9- and 12-months. Each infant saw four trial types of stimuli from either the head-turn stimuli or four trials types from the gaze-shift stimuli at each session. Figure 2 shows screen captures from the two sets.

At the start of each 8-second trial, the speaker smiled into the camera in order to demonstrate engagement with the infant (following a procedure by Senju and Csibra, 2008). She then described one of the objects in front of her (hereafter, the target object). During speaking trials, the woman described the target object while on smiling trials, she simply smiled while the infants heard an audio track from an analogous speaking trial. The object descriptions began with a labeling phrase such as “Would you like to hear about a *modi*?” and were followed by a descriptive sentence that was not intended to identify the target object: “It’s one little piece of a bigger garden we can make.” The lack of uniquely descriptive information meant that target objects could be identified only on gaze-to-object trials. On all trials, the onset of the target word occurred 2 seconds from the start of the trial. At that point, on gaze-to-object trials, the speaker shifted her gaze from the camera toward one of the objects. She maintained fixation on that object for the remainder of the trial.

The set of objects were selected to be unfamiliar to the infants and were matched on relative salience through pilot testing. Thirty-two object pairs were used to create the stimuli. Four pairs were excluded because the objects proved to be excessively salient or unbalanced in relative salience. Infants saw 16 of 24 pairs of objects at each test session. Objects thus overlapped between the ages, but infants did not see exactly the same set of objects at each visit. Target object and position were counter balanced with trial type across infants and trials. Presentation of the 16 trials was randomized for each infant.

Apparatus

Infants were seated on a caregiver’s lap in a sound-treated testing room, approximately 32 inches from the computer screen on which the videos were displayed. The display at this distance subtended approximately $30^\circ \times 20^\circ$ visual angle. The woman’s face on the video screen covered approximately $7^\circ \times 10^\circ$ (for the gaze-shift stimuli, this increased to $10^\circ \times$

14°). The objects covered an average visual angle of $3.5^\circ \times 3.5^\circ$ ($5^\circ \times 5^\circ$ in the gaze-shift stimuli). Two speakers sat on either side of the screen. Below the computer screen was an Applied Science Laboratories (ASL) Pan-Tilt 5000 eye-tracker (sampling rate of 60 hertz), which was connected to a control unit running updated 6000-level software in the adjacent room. The control unit was connected to a PC in the testing booth which ran ASL software for eye tracking. Three video capture boards allowed the experimenter to view (1) the wide angle view of infant and caregiver, (2) the infant's eye and (3) a simultaneous view of the screen infants were watching overlain with crosshairs indicating the running average of the infant's current fixation point and the three previous fixation points. Videos were captured at a rate of 30 frames per second and compressed using a LEAD MCMP/MJPEG Video Codec.

Procedure

Testing sessions began with a two-point calibration sequence. Calibration was followed by a video of a moving ball. Experimenters determined online whether the infant was successfully calibrated based on the apparent tracking of the motion of the ball (within an estimated 1–2° visual angle was accepted). At the start of each trial, an animated figure appeared in one of five positions around the screen (thus randomizing the initial fixation point for each trial). This fixation also allowed the experimenter to reconfirm calibration before each trial. If calibration was deemed inaccurate (greater than 1–2° visual angle), the infant could be recalibrated at any point during the experiment. Once fixation was verified, the experimenter initiated the trial. This was repeated for the 16 test trials (four trials of each of the four trial types). The video of the infant's view overlain with crosshairs indicating fixation was then retained for off-line coding of the infant's focus across the trial.

Coding

Using custom-designed software, four dynamic regions of interest (ROIs) (eyes, mouth, target, and distracter) were established for each stimulus video. The remaining areas of the screen were designated “away” or, if crosshairs were not successfully recorded, “no track.” ROIs were adjusted to fit the changing size of the speaker's face as she spoke and turned her head throughout the trials. Care was taken to equate the eyes and mouth regions. When the two regions differed slightly in overall size, the ROI surrounding the eyes was made larger than the mouth. The ROIs surrounding the mouth and eyes each accounted for approximately 2% of the scene (4% for the gaze-shift stimuli). The objects each covered approximately 3% of the scene (5% for gaze-shift stimuli).

Crosshairs were then extracted frame-by-frame (at a rate of 30 frames/second) from the videos recorded during the eye tracking sessions and cross-referenced by trial with established regions of interest. Crosshairs were extracted using software written in Python. A series of image filters was used to identify the location of the horizontal and vertical lines on the screen. For the 640×480 pixel scene, the estimated error for the X position was less than 3 pixels and for the Y position, less than 1 pixel. Because the sampling rate of the ASL camera did not match that of the video capture board, some frames had what appeared to be two fixation points. In these cases, we calculated the mean of the horizontal and vertical coordinates. Dependent measures were calculated based on the raw cumulative tracked time spent fixating each region (no post-processing or fixation analysis was completed).

A subset of approximately 10% of all collected data (24 of 225 test sessions) was hand coded, frame-by-frame by a naïve rater. Inter-rater reliability between the automated scoring and that completed by hand was 90.5%, Cohen's Kappa = .90.

Results

Trials were considered usable if fixations were successfully tracked for at least half the trial time (4 of the 8 seconds). Data for a given session were included in the analysis if the infant was successfully tracked on at least half of the test trials (8 of the 16 trials such that a mean could be calculated for each trial type). For sessions with sufficient data, an average of 13.68 of the 16 trials ($SD = 2.23$ trials) were successfully tracked. Within those usable trials, gaze was tracked successfully for an average of 6.42 of the 8 seconds of trial time ($SD = 0.71$ sec) across the three ages.

This study was designed to address four factors in face scanning behaviors: presence of information in the mouth, presence of information in the eyes, relevance of the gaze-following cue, and age. We had predicted that attention to mouth would be greater when information was available in the mouth (i.e. speaking trials) and that attention to eyes would be greater when information was available in the eyes (i.e. gaze-to-object trials) and especially when there were no other cues to the target location (i.e. gaze-shift stimuli). To test this, we ran 2 (Gaze: at object or at camera) $\times$ 2 (Speaking: speaking or smiling) $\times$ 3 (Region: eyes, mouth, objects) repeated measures ANOVAs on the proportion of time spent fixating each region with a between subjects factor of stimulus type (head-turn or gaze-shift stimuli) at each of the three ages. In the case of significant interactions, we followed up with alpha-corrected orthogonal contrasts. To address the potential effects of gender, we ran a preliminary analysis with gender as a between subjects factor. There was no main effect of gender nor any significant interactions at 6, 9 or 12 months. We thus collapsed across this factor for the remaining analyses.

Table 1 shows the number of infants who saw each stimulus type, and the mean proportion of time spent fixating the eyes, mouth, and objects at the three ages tested, collapsed across trial types for both the head-turn and gaze-shift stimuli. Figures 4 and 5 illustrate the effects of gaze and speaking at each of the three ages.

At 6 months, with 97 infants contributing usable data, there were significant main effects of region, $F(2, 94) = 226.83$, $p < .01$, partial $\eta^2 = .83$, speaking, $F(1, 95) = 4.77$, $p < .05$, partial $\eta^2 = .05$ and gaze, $F(1, 95) = 5.13$, $p < .05$, partial $\eta^2 = .05$. Infants at 6 months looked more at the face than the objects, $F(1, 95) = 451.17$, $p < .01$, partial $\eta^2 = .83$, but showed no significant difference in their overall distribution of attention to the mouth vs. eyes, $F(1, 95) = .06$, *ns*. At this age, infants looked more at the stimuli overall when the speaker was speaking and less overall when she was looking at the object than when she was looking into the camera.

There was a significant interaction between speaking and region, $F(2, 94) = 44.18$, $p < .01$, partial $\eta^2 = .49$, reflecting an increase in attention to the mouth and decrease in attention to the eyes on speaking trials relative to smiling trials, $F(1, 95) = 67.78$, $p < .01$, partial $\eta^2 = .42$, as well as a decrease in attention to the objects relative to the face when the speaker was speaking, $F(1, 95) = 20.17$, $p < .01$, partial $\eta^2 = .18$. There was also a significant interaction between gaze and region, $F(2, 94) = 9.24$, $p < .01$, partial $\eta^2 = .16$. This was due to increased attention to the eyes when the speaker was looking straight into the camera. There were no other main effects nor any significant interactions at 6 months.

Thus at 6 months, as predicted, attention to the mouth was greater on speaking trials than on smiling trials. Attention to the eyes did not increase for gaze to object trials, even for gaze-shift stimuli on which the eyes were the only cue to the target location. Of note, there was no significant difference in attention between the target and distracter objects for either stimulus set (head turn: $t(48) = .50$, *ns*; gaze shift: $t(47) = .32$, *ns*) indicating that infants were not reliably following the speaker's gaze at this age.

At 9 months, with 65 infants contributing usable data, there were significant main effects of region, $F(2, 62) = 59.28, p < .01$, partial $\eta^2 = .66$, and speaking, $F(1, 63) = 6.84, p < .01$, partial $\eta^2 = .10$, but no significant effect of gaze, $F(1, 63) = .00, ns$. At 9 months, infants looked more at the face than the objects, $F(1, 63) = 102.74, p < .01$, partial $\eta^2 = .62$, and also more at the mouth overall than at the eyes, $F(1, 63) = 7.92, p < .01$, partial $\eta^2 = .11$. Infants at 9 months also looked more at the stimuli overall when the speaker was speaking. There was a significant interaction between speaking and region, $F(2, 62) = 81.35, p < .01$, partial $\eta^2 = .72$, reflecting an increase in attention to the mouth and decrease in attention to the eyes on speaking trials, $F(1, 63) = 156.17, p < .01$, partial $\eta^2 = .71$, as well as a decrease in attention to the objects relative to the face when the speaker was speaking, $F(1, 63) = 23.97, p < .01$, partial $\eta^2 = .28$. There was also a significant interaction between gaze and region, $F(2, 62) = 6.51, p < .01$, partial $\eta^2 = .17$, reflecting increased attention to the eyes when the speaker was looking straight into the camera and increased attention to the speaker's mouth when she was looking at the object $F(1, 63) = 9.13, p < .01$, partial $\eta^2 = .13$.

While there was no main effect of stimulus type at 9 months, nor significant interactions between stimulus type and speaking or gaze, there were significant three-way interactions between speaking, region, and stimulus type, $F(2, 62) = 12.55, p < .01$, partial $\eta^2 = .29$ as well as between gaze, region, and stimulus type, $F(2, 62) = 8.63, p < .01$, partial $\eta^2 = .22$. For head-turn stimuli, even when the speaker was smiling, infants were still more likely to look at her mouth than her eyes. For the gaze-shift stimuli, infants showed the reverse pattern, focusing more on the speaker's eyes than her mouth for smiling trials, $F(1, 63) = 25.32, p < .01$, partial $\eta^2 = .29$. Infants also demonstrated an exaggerated preference for the mouth on the head-turn stimuli by focusing more on the speaker's mouth than on her eyes for gaze to camera trials while those watching the gaze-shift stimuli distributed their attention fairly evenly between the eyes and the mouth on those trial types, $F(1, 63) = 10.89, p < .01$, partial $\eta^2 = .15$. There were no other main effects nor any other significant interactions at 9 months.

Thus at 9 months, as was true at 6 months, attention to the mouth was greater on speaking trials than on smiling trials. Attention to the eyes did not increase for gaze to object trials, even for gaze-shift stimuli on which the eyes were the only cue to the target location. Of note, there was no significant difference in attention between the target and distracter objects for either stimulus set (head turn: $t(29) = .38, ns$; gaze shift: $t(34) = .52, ns$) indicating that infants were not reliably following the speaker's gaze at this age. In contrast to the results at 6 months, at 9 months infants looked more at the mouth of the speaker overall than at her eyes.

At 12 months, with 62 infants contributing usable data, there were significant main effects of region, $F(2, 59) = 18.74, p < .01$, partial $\eta^2 = .39$, and speaking, $F(1, 60) = 8.65, p < .01$, partial $\eta^2 = .13$ but no significant effect of gaze, $F(1, 60) = 1.48, ns$. At 12 months, infants looked more at the face than at the objects, $F(1, 60) = 20.63, p < .01$, partial $\eta^2 = .26$, and more at the mouth than at the eyes, $F(1, 60) = 12.73, p < .01$, partial $\eta^2 = .18$. Infants at 12 months also looked more at the stimuli overall when the speaker was speaking. There was a significant interaction between speaking and region, $F(2, 59) = 80.93, p < .01$, partial $\eta^2 = .73$, reflecting an increase in attention to the mouth and decrease in attention to the eyes on speaking trials, $F(1, 60) = 139.48, p < .01$, partial $\eta^2 = .70$, as well as a decrease in attention to the objects relative to the face when the speaker was speaking, $F(1, 60) = 29.73, p < .01$, partial $\eta^2 = .33$. There was also a significant interaction between gaze and region, $F(2, 59) = 4.54, p < .05$, partial $\eta^2 = .13$, reflecting increased attention to the eyes when the speaker was looking straight into the camera and increased attention to the speaker's mouth when she was looking at the object $F(1, 60) = 7.24, p < .01$, partial $\eta^2 = .11$.

While there was not a significant main effect of stimulus type at 12 months, there was a significant interaction between stimulus type and gaze, $F(1, 60) = 1.48$, $p < .05$, partial $\eta^2 = .07$. Infants watching the head-turn stimuli looked at the stimuli less overall when the speaker's attention was focused on the object than when she looked into the camera. Infants who saw the gaze-shift stimuli looked at both gaze to camera and gaze to object trials roughly the same amount. There were also significant three-way interactions between speaking, region, and stimulus type, $F(2, 59) = 6.44$, $p < .01$, partial $\eta^2 = .18$ and gaze, region, and stimulus type, $F(2, 59) = 3.53$, $p < .05$, partial $\eta^2 = .11$. For head-turn stimuli, even when the speaker was smiling, infants were still more likely to look at her mouth than her eyes. For the gaze-shift stimuli, infants showed the reverse pattern, focusing more on the speaker's eyes than her mouth for smiling trials, $F(1, 60) = 10.50$, $p < .01$, partial $\eta^2 = .15$. Infants also demonstrated an exaggerated preference for the mouth on the head-turn stimuli by focusing more on the speaker's mouth than on her eyes for gaze to camera trials while those watching the gaze-shift stimuli distributed their attention fairly evenly between the eyes and the mouth on gaze to camera trials, $F(1, 60) = 6.84$, $p = .01$, partial $\eta^2 = .10$. Aside from a significant 4-way interaction between gaze, speak, region and stimulus type, which exceeds the scope of this experiment, there were no other significant main effects nor any significant interactions at 12 months.

Once again, at 12 months, attention to the mouth was greater on speaking trials than on smiling trials. Attention to the eyes did not increase for gaze to object trials, however, even for gaze-shift stimuli on which the eyes were the only cue to the target location. Of note, there was no significant difference in attention between the target and distracter objects for the gaze-shift stimulus set, $t(34) = 1.4$, *ns*, indicating that infants were not reliably following the speaker's gaze at this age for this stimulus set. For the head-turn stimuli, infants at 12 months were reliably looking more towards the target object than the distracter, $t(26) = 2.24$, $p < .05$. As was true at 9 months, infants at 12 months looked more at the speaker's mouth overall than at her eyes.

Thus, at each of the three ages, attention to the mouth did increase for speaking trials relative to smiling trials. Attention to the eyes, however, did not increase when information was available in that region of the face. Possible explanations for this are discussed below.

To explore the effect of age with all usable data (6 months: $n = 97$; 9 months: $n = 65$; 12 months, $n = 62$), we used a restricted maximum likelihood linear mixed model, examining the effects of gaze, speaking, and age. We used a linear mixed model because it is robust to missing data. Because the results were generally consistent across the stimulus sets, we collapsed across this factor, thereby also increasing the power of the analysis, but retained stimulus type as a covariate. For the dependent measure, we calculated the ratio of time spent fixating the mouth to time spent fixating the mouth Reyes. We then transformed the resulting ratio using an arcsine transformation. There were significant main effects of speaking $F(1, 738.71) = 75.50$ and of age $F(2, 485.57) = 13.35$, $p < .001$. The effect of gaze was not significant, $F(1, 738.71) = 2.33$, $p = .13$, and there were no significant interactions.² The main effect of speaking reflects an increase in attention to the mouth on speaking trials. The main effect of age reflects the shift towards the mouth between 6 and 9 months.

Thus far, we have been describing the data as group patterns. However, the variability we observed across infants was quite striking. At all three ages, scores on the mouth to face index covered the range from 0 (focus on the eyes) to 1 (focus on the mouth). This variability raises the question of whether infants' face scanning varied within individuals as

²Because SPSS uses a Satterthwaite approximation to compute the denominator degrees of freedom when fitting Linear Mixed Models, the denominator degrees of freedom are not whole numbers.

much as it did across infants. Individual differences across ages were in fact quite stable: within subjects, the ratio of mouth to face across the three ages was significantly correlated, Pearson correlation coefficients for 6 to 9 months: $r(41) = .42, p < .01$; 9 to 12 months: $r(42) = .49, p < .01$; 6 to 12 months: $r(44) = .51, p < .01$. Scatter plots for these correlations are shown in Figure 6.

Discussion

This study explored the effects of four factors on distribution of attention to faces in infancy: information in the mouth, information in the eyes, relevance of the gaze cue, and age. We did this by longitudinally tracking the fixation patterns of infants at 6, 9 and 12 months as they watched video stimuli counter balanced for presence or absence of information in the eyes and mouth. This age range coincides with the emergence of language production.

As we had anticipated, when the speaker was speaking rather than smiling, the infants focused their attention predominantly on her mouth at all three ages. This result reflects the importance of information in the mouth for infants in the second half of the first year of life. This also reflects the highly salient nature of mouth movement. This result is consistent with recent studies indicating that young children focus a good deal of attention on the mouth, especially for stimuli involving speech (Frank et al., 2011; Lewkowicz & Hansen-Tift, 2012; Nakano et al., 2010; von Hofsten et al., 2009). Further, Lewkowicz & Hansen-Tift's finding that infants continue to focus on the mouth longer in infancy for non-native speech than for their native language suggests that the attention to the mouth at this stage is related to the linguistic content and not simply to the saliency of the moving mouth.

We also found that attention to the eyes decreased over the same period. This pattern was likely due to two factors: increased looking to the mouth and a trend towards increased interest in the objects. Contrary to our expectations, infants did not look more at the eyes when the speaker was using her gaze to indicate the target object. Though there was some increase in attention to the target objects in this condition, this shift did not account for the magnitude of the decrease in attention to the eyes. In retrospect, what we called gaze-informative trials (i.e., when the woman was cueing the infant with her gaze direction) could also be described as 'gaze averted': Although the speaker is shifting her attention to the object, she is also shifting her attention *away* from the infant. From birth, infants prefer to look at a person making eye contact over one who is averting her gaze (Farroni, Menon, & Johnson, 2006).

It is possible that the presence of an audio track playing the speaker's voice on trials in which the speaker was smiling may have influenced the patterns of attention in this study. The inclusion of sound was intended to ensure that all stimuli had auditory and visual information. Future work should address the potential effects of this design.

With the exception of the head-turn stimuli at 12 months, infants were not successfully following the gaze of the speaker to the target object in this study. This failure to replicate previous findings that infants will follow gaze by 6 months in an eye tracking paradigm (Senju & Csibra, 2008) may simply be the result of our specific stimuli. It is possible that our novel objects were smaller and less interesting than those used in other studies, making it less obvious to infants that this was a gaze following task. The lack of successful gaze following makes it impossible to determine how infants might have directed their attention if they had recognized the gaze-following component to this task. Had the gaze following task been more salient, infants might have exploited the information in the eyes more readily, especially when this was the only cue to the target location. Further work will be necessary to address this.

One additional interesting result of this study is that there was significant variation between infants on face scanning behaviors. This variation in combination with the consistency of face scanning patterns for infants across the three ages suggests that some infants may be attending to and benefiting from information available in the scene while others are missing out. Gredebäck, Eriksson, Schmitow, Laeng, and Stenberg (2012) indicated that individual factors such as the time infants have spent with their parents can affect the way they scan faces. Young, Merin, Rogers and Ozonoff (2009) demonstrated that infants who looked more at the mouth at 6 months showed superior vocabulary development at age 2. Similarly, Falck-Ytter, Fernell, Gillberg, and Von Hofsten (2010) found that patterns of face scanning also predict social-cognitive skills in autistic and typically-developing children.

Much research has focused on determining where groups of infants tend to look at faces and social scenes. This approach to exploring face scanning may be ignoring significant variability between infants in the way they scan these images. The results of the current study add to the growing body of evidence suggesting that individual patterns of face scanning in infancy may reliably predict future patterns of development. Future studies should examine further correspondences between sensitivity to information in the social scene and other developmental factors.

Acknowledgments

This work was supported by NIH grant HD32005 to JLM and a graduate research award from the Brain Sciences Program at Brown University to EJT and DMS. Thanks go to Lori Rolfe for substantial effort in recruiting and testing, Halie Rando for her face and the participant families for all of their efforts. We also thank Scott P. Johnson at UCLA for the attention attracting video used to calibrate the infants in this study.

References

- Chawarska K, Shic F. Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-Year-old children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. 2009; 39(12):1663–1672. [PubMed: 19590943]
- Chawarska, K.; Shic, F.; Bradshaw, J.; Macari, S.; Klin, A. Atypical Face Scanning in Unaffected High-Risk Infant Siblings. Paper presented at the International Meeting for Autism Research; Chicago, IL. 2009.
- Falck-Ytter T, Fernell E, Gillberg C, Von Hofsten C. Face scanning distinguishes social from communication impairments in autism. *Developmental Science*. 2010; 13(6):864–875. [PubMed: 20977557]
- Fantz RL. Pattern vision in newborn infants. *Science*. 1963; 140:296–297. [PubMed: 17788054]
- Farroni T, Menon E, Johnson MH. Factors influencing newborns' preference for faces with eye contact. *Journal of Experimental Child Psychology*. 2006; 95(4):298–308. [PubMed: 17030037]
- Frank MC, Vul E, Saxe R. Measuring the Development of Social Attention Using Free-Viewing. *Infancy*. 2011:1–21.
- Goren CC, Sarty M, Wu PYK. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*. 1975; 56(4):544. [PubMed: 1165958]
- Gredebäck G, Eriksson M, Schmitow C, Laeng B, Stenberg G. Individual Differences in Face Processing: Infants' Scanning Patterns and Pupil Dilations are Influenced by the Distribution of Parental Leave. *Infancy*. 2012; 17(1):79–101.
- Haith MM. Eye contact and face scanning in early infancy. *Science*. 1977; 198(4319):853–855. [PubMed: 918670]
- Haith MM, Bergman T, Moore MJ. Eye contact and face scanning in early infancy. *Science*. 1977; 198:853–855. [PubMed: 918670]
- Hunnus S, Geuze RH. Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy*. 2004; 6(2):231–255.

- Johnson M, Dziurawiec S, Ellis H, Morton J. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*. 1991; 40(1):1–19. [PubMed: 1786670]
- Jones W, Carr K, Klin A. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Arch Gen Psychiatry*. 2008; 65(8):946–954. [PubMed: 18678799]
- Lewkowicz DJ, Hansen-Tift AM. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*. 2012
- Liu S, Quinn PC, Wheeler A, Xiao N, Ge L, Lee K. Similarity and difference in the processing of same-and other-race faces as revealed by eye tracking in 4-to 9-month-olds. *Journal of Experimental Child Psychology*. 2011; 108(1):180–189. [PubMed: 20708745]
- Maurer D, Salapatek P. Developmental changes in the scanning of faces by young infants. *Child Development*. 1976; 47(2):523–527. [PubMed: 1269319]
- Nakano T, Tanaka K, Endo Y, Yamane Y, Yamamoto T, Nakano Y, et al. Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society B: Biological Sciences*. 2010; 277(1696):2935–2943.
- Oakes LM, Ellis AE. An Eye-Tracking Investigation of Developmental Changes in Infants' Exploration of Upright and Inverted Human Faces. *Infancy*. 2011:no–no.
- Senju A, Csibra G. Gaze following in human infants depends on communicative signals. *Current Biology*. 2008; 18(9):668–671. [PubMed: 18439827]
- Tenenbaum E, Sobel DM, Malle BF, Sheinkopf SJ, Morgan JL. Gaze following and attention to faces in infancy predict language development. 2012 Manuscript submitted for publication.
- Turati C, Valenza E, Leo I, Simion F. Three-month-olds' visual preference for faces and its underlying visual processing mechanisms. *Journal of Experimental Child Psychology*. 2005; 90(3):255–273. [PubMed: 15707862]
- von Hofsten C, Uhlig H, Adell M, Kochukhova O. How children with autism look at events. *Research in Autism Spectrum Disorders*. 2009; 3(2):556–569.
- Wheeler A, Anzures G, Quinn PC, Pascalis O, Omrin DS, Lee K. Caucasian Infants Scan Own-and Other-Race Faces Differently. *PLoS ONE*. 2011; 6(4):e18621. [PubMed: 21533235]
- Young GS, Merin N, Rogers SJ, Ozonoff S. Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*. 2009; 12(5):798–814. [PubMed: 19702771]

Figure 1.

Trial Types: A. Look at Object and Speak: The woman gazed toward the object while speaking about it. B. Look at Object and Smile: The woman gazed toward the object and smiled while an audio track from a speaking trial was played. C. Look at Camera and Speak: The woman fixated the camera and spoke about the object. D. Look at Camera and Smile: The woman fixated the camera and smiled while an audio track from a speaking trial was played.

Figure 2.
Scene captures from the head-turn (left) and gaze-shift (right) stimuli.

Figure 3.

A scene capture from the custom software used to define dynamic ROIs for the eyes, mouth, target, and distracter objects. The lines that extend to the edges of the screen indicate the infant's fixation. In this case, the frame would be coded as "Mouth."

Figure 4.

Effect of the speaker's direction of attention (left panel: looking at camera, right panel: looking at object) on infants' fixations patterns by age and stimulus set (top panel: gaze-shift stimuli, bottom panel: head-turn stimuli). Attention to the eyes increased overall for the gaze-shift stimuli over head-turn stimuli. While the focus of the speaker's attention made little difference in patterns of fixation for the head-turn stimuli, infants who saw the gaze-shift stimuli were more likely to look at the speaker's eyes when she was looking at the camera than when she was looking at the object for all three ages. Tracked trial time unaccounted for in these proportions was spent fixating other areas of the screen.

Figure 5.

Effect of smiling (left) or speaking (right) on distribution of attention for all infants tested by age and stimulus set (gaze-shift stimuli in top panel, head-turn stimuli in bottom panel) collapsed across the speaker's direction of attention. Increases in attention to the mouth between 6 and 9 months are apparent for both stimulus sets but only on speaking trials. The remaining tracked trial time was spent fixating other areas of the screen.

Figure 6. Scatter plots of the mouth per face ratios at 6 and 9 months (top left), 9 and 12 months (top right), and 6 and 12 months (bottom).

Overall mean proportion of tracked trial times spent fixating the eyes, mouth and objects by age and stimulus type, collapsed across the four trial types. Totals on the left reflect means for the set of infants who saw the head-turn stimuli and totals on the right were for the infants who saw the gaze-shift stimuli.

Table 1

Head-Turn Stimuli			Gaze-Shift Stimuli		
	Eyes	Mouth	Objects	Eyes	Objects
6 months					
Mean	.25	.31	.05	.32	.25
SD	(.20)	(.22)	(.06)	(.20)	(.17)
n	49	49	49	48	48
9 months					
Mean	.19	.36	.11	.25	.29
SD	(.16)	(.19)	(.06)	(.17)	(.12)
n	30	30	30	35	35
12 months					
Mean	.15	.32	.15	.21	.29
SD	(.12)	(.17)	(.10)	(.17)	(.13)
n	27	27	27	35	35