

Free Viewing Gaze Behavior in Infants and Adults

John M. Franchak<sup>1</sup>, David J. Heeger<sup>2</sup>, Uri Hasson<sup>3</sup>, Karen E. Adolph<sup>2</sup>

<sup>1</sup> Department of Psychology, University of California, Riverside

<sup>2</sup> Department of Psychology and Center for Neural Science, New York University

<sup>3</sup> Department of Psychology and the Neuroscience Institute, Princeton University

Author Note

This research was supported by NICHD R37-HD033486 to KEA, NIMH R01-MH094480 to UH, and NIDA R21-DA024423 to DH.

WORD COUNT: 8615

### Abstract

The current study investigated age differences in free viewing gaze behavior. Adults and 6-, 9-, 12-, and 24-month-old infants watched a 60-s *Sesame Street* video clip while their eye movements were recorded. Adults displayed high inter-subject consistency in eye movements; they tended to fixate the same places at the same. Infants showed weaker consistency between observers and inter-subject consistency increased with age. Across age groups, the influence of both bottom-up features (fixating visually-salient areas) and top-down features (looking at faces) increased. Moreover, individual differences in fixating bottom-up and top-down features predicted whether infants' eye movements were consistent with those of adults, even when controlling for age. However, this relation was moderated by the number of faces available in the scene, suggesting that the development of adult-like viewing involves learning when to prioritize looking at bottom-up and top-down features.

*Keywords:* eye movements, vision, attention, face perception, saliency, inter-subject correlation

### **Free Viewing Gaze Behavior in Infants and Adults**

Parsing a visual scene requires moving the eyes to particular points of interest. Observers must move their eyes frequently to acquire high-resolution information about relevant areas of a scene because the distribution of photoreceptors in the fovea is denser than in the periphery of the eye. When viewing a static scene such as a photograph, observers can scan the image at their leisure to explore its features. But with a dynamic scene such as a movie, observers must prioritize where to look from moment to moment as the events unfold. In real life, observers must move their heads and bodies in addition to their eyes to select what is in view (Land, 2004). As such, many factors influence visual selection: perceptual characteristics of stimuli (Borji & Itti, 2013; Itti, 2005; Mital, Smith, Hill, & Henderson, 2011; Parkhurst, Law, & Niebur, 2002; 't Hart et al., 2009), motor constraints on looking behavior (Franchak, Kretch, Soska, & Adolph, 2011; Kretch, Franchak, & Adolph, 2014), task demands (Hayhoe & Rothkopf, 2011; Land & Hayhoe, 2001; Yarbus, 1967), the presence of socially-relevant stimuli (Amso, Haas, & Markant, 2014; Franchak et al., 2011; Frank, Vul, & Johnson, 2009; Frank, Vul, & Saxe, 2011), and scene comprehension (Kirkorian, Anderson, & Keen, 2012; Pempek et al., 2010). Accordingly, characterizing the development of gaze behavior is complex.

What factors differentiate the eye movements of infants and adults, and how do infants eventually achieve adult-like gaze behavior? Only a handful of studies report infants' free viewing of complex, dynamic stimuli (Frank et al., 2009; Frank et al., 2011; Kirkorian et al., 2012; Kretch & Adolph, 2015). Instead, most studies rely on static images or videos depicting simple events such as bars moving behind occluders—presumably to avoid confounds among various influences on infants' eye movements. Although such studies provide valuable information about developmental changes in particular aspects of gaze, they cannot inform on

age-related changes in eye movements when infants watch complex, dynamic scenes. Here, we build on prior work by investigating the development of free viewing gaze behavior in infants. In particular, we focus on age-related changes in eye movement consistency and how those changes relate to bottom-up and top-down influences (i.e., stimulus salience versus semantic relevance) on eye movements. Moreover, we bridge the literatures on infant and adult visual behavior by asking what age-related changes make infants' viewing patterns more adult-like.

### **Eye movement consistency during free viewing**

Adults' eye movements are highly consistent when freely viewing dynamic stimuli—observers tend to look at the same location at the same time (Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Hasson, Yang, Vallines, Heeger, & Rubin, 2008; Mital et al., 2011; Shepherd, Steckenfinger, Hasson, & Ghazanfar, 2010; Smith & Mital, 2013; 't Hart et al., 2009; Wang, Freeman, Merriam, Hasson, & Heeger, 2012). Eye movement time series are correlated between multiple observers looking at the same stimulus and between repeated stimulus presentations to the same observer. Such consistency, however, is not obligatory. For example, idiosyncratic eye movements in autistic adults watching a movie result in poor inter-subject correlation (Hasson et al., 2009). Moreover, high inter-subject correlations in typically-developing adults do not mean that observers' eye movements are identical. Rather, the correlations indicate a high degree of overlap among observers' eye movements. Furthermore, different types of stimuli vary in how successfully they evoke consistent gaze among observers. For example, Hollywood movies evoke greater consistency in eye movements compared to homemade, “naturalistic” movies (Dorr et al., 2010; Hasson, Landesman, Knappmeyer, Vallines, Rubin, & Heeger, 2008; Hasson et al., 2008).

Are infants' eye movements during free viewing consistent among observers? Previous work yielded varying results regarding whether eye movement consistency increases over development. Bivariate ellipse area analysis, which estimates the spatial spread of fixation locations within age groups, showed increasing consistency across age in 1-year-olds, 4-year-olds, and adults viewing an episode of *Sesame Street* (Kirkorian et al., 2012). Similarly, entropy analysis showed increasing eye movement consistency among observers within age groups in 3-, 6- and 9-month-olds and adults watching short clips from *A Charlie Brown Christmas* (Frank et al., 2009). However, a more recent study of 200 infants from the same researchers (Frank et al., 2011) found no age difference in entropy in younger (3- to 12-month-olds) and older infants (12- to 30-month-olds). The study also reported more consistent eye movements for each age group when viewing scenes containing a single agent compared to scenes containing multiple agents.

### **What accounts for eye movement consistency?**

Common strategies for distributing eye movements would account for consistency in adults' eye movements. Two types of influences on adults' gaze selection—bottom-up and top-down features—have been studied extensively in free viewing tasks (Henderson, 2007; Tatler, Hayhoe, Land, & Ballard, 2011). If eye movement consistency does indeed increase with age, the extent to which bottom-up and top-down features account for eye movement consistency in adults would provide a basis for understanding its development.

Bottom-up influences are characterized by how well the salience of low-level stimulus features accounts for eye movements. Quantitative models of low-level stimulus salience have been proposed that approximate early visual processing to calculate the prominence of image features based on luminance, contrast, color, orientation, and motion (for review, see Borji & Itti, 2013). The individual channels are combined to form an overall saliency map that predicts the

likelihood that different areas of the image will be fixated. Indeed, observers' eye movements when viewing dynamic stimuli are influenced by bottom-up saliency (Borji & Itti, 2013; Itti, 2005; Mital et al., 2011; Smith & Mital, 2013; 't Hart et al., 2009). In particular, motion information may account for eye movement consistency in free viewing (Mital et al., 2011; Smith & Mital, 2013). However, observers' eye movements do not correlate with the most salient location in the image while watching Hollywood movies (Shepherd et al., 2010).

Top-down factors also influence adults' eye movements. Changing observers' task affects eye movements when viewing static images (Yarbus, 1967), dynamic stimuli (Smith & Mital, 2013), and when performing natural actions (Franchak & Adolph, 2010; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land, Mennie, & Rusted, 1999). Even in the absence of an explicit task, top-down factors influence free viewing by prioritizing semantically-relevant stimuli such as objects and faces. Faces attract observers' gaze when viewing static images (Cerf, Harel, Einhauser, & Koch, 2007; Yarbus, 1967) or dynamic movies (Foulsham, Cheng, Tracy, Henrich, & Kingstone, 2010; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Shepherd et al., 2010). Thus, the tendency to look at faces may contribute to eye movement consistency among observers (Shepherd et al., 2010; Smith & Mital, 2013).

Bottom-up and top-down factors influence eye movements differently in infants compared to adults. Although young infants prefer to look at faces in static image arrays over other types of stimuli (Gliga, Elsabbagh, Andravizou, & Johnson, 2009; Gluckman & Johnson, 2013; Libertus & Needham, 2011), the proportion of time spent fixating faces in static images (Amso et al., 2014) and dynamic displays (Frank et al., 2009) starts at a modest level before increasing gradually over development. Bottom-up features may capture young infants' attention more readily than top-down features: A saliency model accounted for free viewing patterns in 3-

month-olds whereas a face model better predicted gaze in 9-month-olds and adults (Frank et al., 2009). Over development, an increase in the tendency to fixate faces and a decrease in the tendency to fixate salient features may account for increasing eye movement consistency.

Bottom-up and top-down influences are not independent, and thus are difficult to disentangle. Correlating eye movements with information about bottom-up and top-down features cannot reveal causal effects of features on gaze behavior if those features are inter-related (Henderson, 2003). In static images, salient regions are more likely to be objects (Einhauser, Spain, & Perona, 2008; Elzary & Itti, 2008), and child and adult observers fixate salient faces more often than non-salient faces (Amso et al., 2014). In dynamic scenes, agentive action creates motion contrast. For example, a person walking across a static background generates motion salience. If an observer fixates the person, is the fixation the result of low-level salience or the presence of a social agent? Designers of children's television may exploit the power of saliency to draw attention to social agents: Flicker and feature congestion predict the location of the speaking character's face in toddler-directed but not adult-directed television programs (Wass & Smith, 2015).

Similarly, top-down features are difficult to separate from overall scene comprehension. Infants' inability to comprehend and learn from screen-based media has been extensively documented (Anderson & Pempek, 2005; Barr & Hayne, 1999). Only at 24 months of age do children detect glaring inconsistencies such as scrambling the order of scenes in a television program (Pempek et al., 2010). Infants and children are unlikely to watch media that exceeds their understanding such as adult television programs (Valkenburg & Vroone, 2004). Different comprehension of media over development may modulate eye movements to top-down features such as faces. For example, if there are multiple faces in a scene, which one should observers

fixate? In this case, a simple “fixate faces” strategy would not lead to consistent gaze across observers. Whereas adults might fixate the face of an agent who is speaking or participating in a key action, infants may fixate faces haphazardly if they fail to follow the narrative content. Lacking a reason to select a particular face, infants might fall back on a bottom-up viewing strategy.

### **Current Study**

The purpose of the current study was to assess the development of eye movement consistency during free viewing of dynamic stimuli and to test how bottom-up and top-down factors influence eye movements. Participants watched a short clip from *Sesame Street* while their eye movements were recorded. To integrate our findings with prior work on free viewing (Frank et al., 2009; Frank et al., 2011; Kirkorian et al., 2012), we tested 6- to 24-month-old infants and compared their eye movements with those of adult observers (Hasson, Yang, et al., 2008; Smith & Mital, 2013; Wang et al., 2012).

A primary aim was to describe the developmental course of eye movement consistency. Prior work suggests an increase in consistency within age groups from infants to adults (Frank et al., 2009; Kirkorian et al., 2012). Such an increase might result from observers adopting a single, adult-like viewing strategy or from eye movements that become more similar within age groups but are dissimilar between age groups. For example, infants might fixate the location with the most movement whereas adults fixate a face. By measuring eye movement consistency between infants and adults in addition to measuring consistency within each age group, we tested whether increasing consistency over infancy results from increasingly adult-like viewing patterns.

A second aim was to assess what factors predict adult-like gaze. Prior work suggests that bottom-up and top-down features relate to eye movement consistency in adults (Mital et al.,



2011; Shepherd et al., 2010; Smith & Mital, 2013) and that these influences change over development (Amso et al., 2014; Frank et al., 2009). But it is unclear whether and how changes in eye movement consistency are related over development to changes in fixation of bottom-up versus top-down features. We analyzed the extent to which individual observers' gaze patterns are influenced by bottom-up (salient) and top-down (faces) factors at different ages. By measuring inter-subject correlations between infants' and adults' eye movements, we addressed whether fixating salient regions or faces predicts adult-like free viewing.

Third, we asked whether eye movement consistency and the influence of bottom-up and top-down features vary with changes in scene content. Prior work showed that eye movements were less consistent when multiple agents were present in the scene compared to only a single agent or face (Frank et al., 2011). By analyzing segments of the stimulus that contained either a single agent or multiple agents, we asked how different age groups respond to changes in scene content. Specifically, we determined whether infants' eye movements are less consistent when multiple faces compete for their attention. Furthermore, we assessed how scene-content related changes in eye movement consistency are manifest in differential gaze to bottom-up and top-down features.

## **Method**

### **Participants**

Participants were six 6-month-olds (6.0 to 6.2 months, 5 male), six 9-month-olds (8.9 to 9.2 months, 4 male), six 12-month-olds (11.8 to 12.4 months, 2 male), six 24-month-olds (23.7 to 24.8 months, 4 male), and 6 adults (20.6 to 32.6 years, 2 male). Infants were healthy and full-term with normal vision. Adults had normal or corrected-to-normal vision. Families were recruited from hospitals in the greater New York City metropolitan area; most infants were White

and middle-class. Six additional adults (20.4 to 23.3 years of age, 2 male) were recruited to serve as a comparison group (see Data Analysis for more details). An additional 5 infants were recruited but did not contribute data due to poor eye tracking data quality (1 infant) and failure to watch experimental stimuli due to fussiness or distraction (4 infants). Infants' families received a small gift for their participation and adult participants received course credit.

### **Procedure and apparatus**

Infants' and adults' eye movements were recorded while they watched a 60 s video clip from *Sesame Street* with the audio track played through computer speakers. The stimulus video showed a human actor singing a song about counting to four. Two times of interest (TOIs) were defined based on scene content: A 16 s segment at the beginning of the video featuring only the human actor singing ("One Agent," Figure 1A) was immediately followed by a 21 s segment that contained 4 Muppets singing and dancing with the actor ("Multiple Agents," Figure 1B). The entire 60 s video was composed of a single shot (i.e., no cuts to different scenes) to avoid the tendency of observers to fixate the center of the display following cuts (Kirkorian et al., 2012; Wang et al., 2012).

Participants sat 65 cm away from a 55.9 cm (diagonal) widescreen LCD monitor equipped with a 120 Hz SMI RED eye tracker (SensoMotoric Instruments). The stimulus video was presented full-screen at 30 Hz (visual angle of the stimulus =  $40^\circ \times 26^\circ$ ). The monitor was mounted on an adjustable arm to accommodate participants' height. Infants sat in a highchair with shoulder straps to reduce body movement. Parents sat behind infants and did not interact with them during the experiment.

After adjusting the position of the monitor, the experimenter ran the SMI calibration routine using a 2-point calibration for infants and a 9-point calibration for adults. The stimulus

video was shown following a successful calibration. At the end of the session, a 4-point validation determined the accuracy of the calibration. Spatial accuracy averaged  $0.92^\circ$  (horizontal) by  $1.1^\circ$  (vertical) and did not vary by age. Eye tracking data were sampled at 120 Hz.

### **Data analysis**

Each participant's eye tracking data were extracted as horizontal and vertical time series and analyzed in Matlab. Periods when eye tracking data were unavailable (observers turned away, closed their eyes, or looked offscreen) were excluded from analyses.

**Eye movement consistency.** Measures used in prior infant work, such as entropy (Frank et al., 2009) and bivariate ellipse area analysis (Kirkorian et al., 2012), result in a single score that represents consistency within a group. Other metrics, such as root mean square (Gredebäck, Eriksson, Schmitow, Laeng, & Stenberg, 2012), describe the spatial spread of eye movements for an individual. To analyze eye movement consistency between and within age groups, we needed a different sort of analysis. So we calculated *inter-subject correlations* (ISCs) between pairs of observers as a metric of similarity between eye movement time series. For a pair of observers, ISC was calculated by: (1) calculating the correlation coefficient among observers' horizontal eye movement time series, (2) calculating the correlation coefficient for observers' vertical eye movement time series, and (3) averaging the resulting horizontal and vertical correlation coefficients from steps 1 and 2 (Hasson, Yang, et al., 2008). Figure 2A shows an example of two adult observers' horizontal time series and the corresponding horizontal ISC (the result of step 1).

Each observer's *within-age correlation* ( $ISC_w$ ) was obtained by averaging ISCs from that observer paired with each of the others in the same age group (Figure 2B). For example, a 9-month-old's  $ISC_w$  was determined by computing ISC between that observer and each of the five

other 9-month-olds, and then averaging the five resulting numbers. Each observer's *between-age* ISCs were obtained by averaging ISCs from that observer paired with each of the observers in another age group (Figure 2C-D). We calculated between-age ISCs for each observer compared with each participant in every other age group. For example, a 6-month-old's between-age correlation with 12-month-olds was obtained by averaging the ISCs of the 6-month-old paired with each of the six 12-month-olds. Likewise, the 6-month-old's inter-subject correlation with adults ( $ISC_a$ ) was obtained by calculating the average ISC of the 6-month-old paired with each of the six adult observers.

The factors that accounted for observers' correlation with adult participants were of key interest; thus, we needed to compare infants'  $ISC_a$  to adults'  $ISC_a$ . However, calculating ISCs between pairs of adults in the same group would yield a within- rather than a between-group comparison. To address this issue, we recruited six adults to serve as an independent comparison group.  $ISC_a$  values for every age group were calculated against the comparison group of adults; data from comparison adults were not used in other analyses. The outcomes of statistical tests did not change when the two groups of adults were swapped.

Within- and between-age correlations were tested for statistical significance using randomization tests with time-randomized baselines, replicating procedures used previously (Shepherd et al., 2010), to assess whether correlations were stronger than what would be expected by chance. To preserve the sequential information in a time series while randomizing the time stamps, each time series was realigned to start at a random time. The last sample in the time series "looped" to the first sample so that all time stamps from the original time series were used in the randomized time series. Inter-subject correlations and the resulting within- and between-age correlations were recalculated using 1000 randomized time series derived from

each observer to create null distributions to compare to our results (according to the null hypothesis, any correlations in the data were due to statistics of the eye movements, independent of the video). If a within- or between-age correlation fell beyond the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of the corresponding baseline distribution, we considered the correlation to be significant at the .05 level.

**Saliency analyses.** The stimulus video was converted to a sequence of image frames and imported into Matlab for saliency analyses. An implementation of the standard Itti, Koch, and Neibur algorithm (1998) was downloaded (<http://www.vision.caltech.edu/~harel/share/gbvs.php>), and used to calculate the relative salience of each pixel based on five biologically-inspired channels (color, contrast, orientation, flicker, and motion). Values from each channel were weighted equally to determine an overall saliency map for each video frame. As in prior work (Kretch & Adolph, 2015), a percentile rank was calculated for each pixel relative to the other pixels in the frame. The most salient pixel (Figure 1, blue crosses) was ranked 100 and the least salient pixel was ranked 1.

We calculated *gaze saliency* for each participant to determine the overall saliency of areas where the observer directed gaze. For each frame, we averaged the percentile ranks of the pixels within a 1.5-degree radius of gaze location. Averaging the mean saliency ranks across frames yielded a single gaze-saliency metric for each participant. Larger gaze-saliency values indicated that participants tended to fixate relatively more salient regions of the video.

**Looking at faces.** To determine the proportion of time that participants' gaze tracked the faces of the human actor and four Muppets, we conducted dynamic area of interest (AOI) analyses using SensoMotoric Instruments' BeGaze software. For each frame of the video, the experimenter drew an elliptical AOI around each face (Figure 1, white ellipses). On average,

AOIs were  $59.7 \text{ deg}^2$  in area but varied slightly over time and between faces to accommodate changes in size due to faces moving in depth. The proportion of time spent fixating the human actor's face (*face looking*) was measured for the entire duration of the video because it was always on screen. We calculated looking at the Muppets' faces in a separate analysis because they were only present during the Multiple Agents TOI. Because observers had different amounts of missing data in each TOI, each observer's face-looking scores were calculated by dividing the amount of time spent fixating faces by the amount of valid eye tracking data available for that participant. Conflicts between overlapping AOIs were resolved by attributing gaze to the front-most AOI based on depth order.

## Results

We conducted analyses in three stages. First, we measured age-related changes in eye movement consistency and tested for bottom-up and top-down influences on eye movements using data from the entire video. Second, we investigated whether individual differences in fixation to bottom-up and top-down features predict eye movement consistency. Third, we repeated our analyses for two distinct segments of the video (Figure 1) to determine whether eye movement consistency and its relation to bottom-up and top-down features depended on scene content (one face versus multiple faces). We examined age effects with one-way ANOVAs and follow-up tests of linear trends. With only 6 participants per group and 5 age groups, pairwise post hoc comparisons were underpowered, so we could not test whether each age group was reliably different from the others.

### Eye movement consistency increases with age

The within-age eye movement correlations (as exemplified in Figure 2B) were statistically significant (compared to time-randomized baselines) at every age, indicating that

even the youngest infants showed some degree of inter-subject consistency while watching the video. However, correlations were not equal across age groups (Figure 3A) and became stronger with age (Table 1). The ANOVA confirmed age differences in within-age correlations,  $F(4, 25) = 5.00, p = .004$ ; a linear trend confirmed the age-related increase in correlations,  $F(1, 25) = 18.487, p < .001$ .

Infants' eye movements becoming more adult-like might drive increasing within-age correlations. Alternatively, different patterns of looking might underlie within-age correlations in the different age groups. We addressed the latter possibility by examining between-age correlations (as in Figure 2D). Table 1 shows the average between-age correlations for every pair of age groups (adults' eye movements were correlated against the comparison sample of adults); every between-age correlation was statistically significant, suggesting some degree of consistency in eye movements across age groups. Of key interest was to what degree infants' eye movements were *adult-like*, as measured by calculating inter-subject correlations with adults' eye movements ( $ISC_a$ ). Figure 3B shows that younger infants' eye movements were weakly correlated with adults' and illustrates that those correlations increased with age. The ANOVA confirmed that  $ISC_a$  differed by age,  $F(4, 25) = 8.14, p < .001$ , and a linear trend confirmed that  $ISC_a$  increased with age,  $F(1, 25) = 24.49, p < .001$ .

### **Bottom-up and top-down influences on looking patterns**

Across ages, participants fixated salient areas (Figure 3C). On average, the pixels around the point of gaze ranked in the top quartile ( $M = 80.7\%, SD = 4.79$ ). The ANOVA did not show an effect of age on gaze saliency,  $F(4, 25) = 2.43, p = .093$ , but a linear trend indicated a modest increase in fixating salient areas with age,  $F(1, 25) = 5.94, p = .022$ .

With age, observers spent more time fixating the human actor's face (Figure 3D). For this analysis, we computed proportion of looking at the human actor's face (excluding the Muppet faces) over the entire video (collapsing across times when one face and multiple faces are present). Six-month-olds and 9-month-olds spent the least amount of time fixating the human actor's face,  $M = .220$  ( $SD = 0.228$ ) and  $M = .183$  ( $SD = 0.120$ ), respectively. Twelve-month-olds and 24-month-olds spent relatively more time fixating the actor's face,  $M = .386$  ( $SD = 0.084$ ) and  $M = .423$  ( $SD = 0.058$ ), respectively, and adults spent the most time,  $M = .656$  ( $SD = 0.148$ ). The ANOVA confirmed age differences in face looking,  $F(4,25) = 10.84$ ,  $p < .001$ , and a linear trend confirmed that face looking increased with age,  $F(1,25) = 37.717$ ,  $p < .001$ .

Faces were more salient than other areas of the video that participants fixated. We compared the saliency rank of pixels around the point of gaze when participants fixated faces (of the human actor and Muppets) compared to when they fixated non-face regions. Regardless of age, gaze saliency was greater for face regions ( $M = 84.2\%$ ,  $SD = 3.96$ ) compared to non-face regions ( $M = 76.7\%$ ,  $SD = 5.34$ ). A 5 (age group)  $\times$  2 (region of interest: face, non-face) ANOVA revealed only a main effect of region of interest,  $F(1,25) = 93.45$ ,  $p < .001$ .

### **Looking at bottom-up and top-down features predicts adult-like viewing**

Individual observers' gaze-saliency and face-looking values were related to their correlation with adults' gaze. Across age groups, gaze saliency and  $ISC_a$  values were positively correlated,  $r(28) = .597$ ,  $p < .001$  (Figure 4A). Similarly, face looking was significantly correlated with  $ISC_a$ ,  $r(28) = .672$ ,  $p < .001$  (Figure 4B). To see if this pattern held only for infants, we recalculated the correlations after excluding adult observers; significant correlations remained when considering only infant observers for both gaze saliency,  $r(22) = .593$ ,  $p = .002$ , and face looking,  $r(22) = .474$ ,  $p = .019$ .



Because gaze saliency, face looking, and age were inter-correlated, we used hierarchical linear regression to assess whether each factor could explain unique variance in infants'  $ISC_a$  values. We entered predictors one-by-one into the regression to test whether each predictor accounted for additional variance ( $R^2$  change) after accounting for the effects of the previous variables. We chose to enter face looking first due to its predominance in the developmental literature on free viewing, then entered gaze saliency and age to determine whether each could account for additional unique variance. Table 2 shows  $R^2$  and  $R^2$  change in the model after entering each predictor. Face looking accounted for 22.5% of the variance in infants'  $ISC_a$  values. Adding gaze saliency to the model accounted for an additional 18.9% of variance. However, age at test did not explain additional variance (1.9%) and was excluded from the model. Together, face looking and gaze saliency accounted for 41.3% of the variance in infants'  $ISC_a$  values,  $F(2,23) = 7.40, p = .004$ .

### **Scene content affects eye movement consistency**

Eye movement consistency decreased when multiple faces were present compared to when only a single face was present (Figure 5A-B). We recalculated  $ISC_w$  and  $ISC_a$  during the “One Agent” and “Multiple Agents” times of interest. For  $ISC_w$ , a 5 (age group)  $\times$  2 (TOI: One Agent, Multiple Agents) ANOVA revealed only a main effect of TOI,  $F(1,25) = 11.16, p = .003$ ; both  $ISC_w$  and  $ISC_a$  were lower in the Multiple Agents TOI compared to the One Agent TOI. However, a linear trend contrast showed that  $ISC_w$  increased with age across both TOIs,  $F(1,25) = 11.16, p = .003$ . For  $ISC_a$ , a 5 (age group)  $\times$  2 (TOI: One Agent, Multiple Agents) ANOVA confirmed only main effects of TOI,  $F(1,25) = 93.16, p < .001$ , and age,  $F(4,25) = 5.98, p = .002$ . A significant linear trend indicated that  $ISC_a$  increased with age for both TOIs,  $F(1,25) = 93.16, p < .001$ .

Eye movement consistency was affected by scene content but gaze saliency and face looking were not ( $p > .05$ ) (Figure 5C). Proportion of time looking at the human actor's face did not vary across times of interest but did vary by age (Figure 5D). A 5 (age group)  $\times$  2 (TOI: One Agent, Multiple Agents) ANOVA revealed only an effect of age on looking at the human actor's face,  $F(4,25) = 12.83, p < .001$ .

**Predicting adult-like viewing in the “One Agent” segment.** When only one face was present in the scene, gaze saliency but not face looking predicted infants' correlations with adults' gaze:  $ISC_a$  values in the One Agent TOI were positively related to gaze saliency for infant observers,  $r(22) = .558, p = .005$ . Although face looking and  $ISC_a$  values were positively related, the correlation was not statistically significant,  $r(22) = .338, p = .106$ . We employed the same hierarchical regression procedure as before to test the unique contributions of face looking, gaze saliency, and age to  $ISC_a$  values within the One Agent TOI. Table 2 shows that looking at the human actor's face accounted for 11.4% of the variance but failed to reach significance. Gaze saliency accounted for a significant 19.7% of the variance in infants'  $ISC_a$  values after controlling for face looking. Adding infants' age into the model did not significantly change the explained variance.

When only a single face is present, increased salience of the face compared to other regions might account for why gaze saliency predicts  $ISC_a$ . Indeed, saliency ranks were much greater when participants looked at the human actor's face ( $M = 89.8, SD = 3.22$ ) compared to non-face regions ( $M = 72.9, SD = 4.14$ ). A 5 (age group)  $\times$  2 (ROI: face, non-face) ANOVA revealed only a main effect of ROI on saliency,  $F(1,23) = 578.1, p < .001$ .

**Predicting adult-like viewing in the “Multiple Agents” segment.** Top-down features predicted adult-like viewing better than bottom-up features when there were multiple, visually-

salient faces present in the scene. In the Multiple Agents TOI, the correlation between gaze saliency and  $ISC_a$  was not statistically significant,  $r(22) = -.003$ ,  $p = .998$ , but face looking and  $ISC_a$  were positively correlated,  $r(22) = .401$ ,  $p = .052$ . Table 2 shows  $R^2$  and  $R^2$  change statistics for a hierarchical regression predicting  $ISC_a$  from gaze saliency, face looking, and age. Looking at the human actor's face accounted for a significant portion of the variance in  $ISC_a$  values (16.1%). However, in the Multiple Agent TOI, gaze saliency failed to predict variance in  $ISC_a$  values (0.1%), and entering age did not significantly increase the variance explained (8.7%).

Infants spent less time fixating faces in general and more time fixating the Muppet faces compared to older children and adults (Figure 5D, gray triangles), which might have contributed to lower  $ISC_a$  values. A two-way repeated-measures ANOVA with age group and agent (Muppets vs. human) as factors found a main effect of age on face looking,  $F(4, 25) = 8.12$ ,  $p < .001$ . A significant linear contrast on age indicated that overall looking at faces increased with age,  $F(1, 25) = 22.70$ ,  $p < .001$ . Furthermore, a significant agent  $\times$  age interaction revealed that the proportion of looking at Muppet and human faces changed with respect to age,  $F(4, 25) = 7.061$ ,  $p = .001$ . Follow-up Sidak-corrected pairwise comparisons showed that 6- and 9-month-old infants looked longer at the faces of the Muppets compared to the human actor ( $ps < .01$ ), whereas adults spent a greater proportion of time looking at the human actor's face ( $p < .01$ ). Twelve- and 24-month-olds did not look significantly more at either Muppet or human faces.

Adult-like gaze in the Multiple Agents TOI was predicted by looking more often at the human actor's face in spite of the fact that the Muppet faces were more salient. Possibly, increased salience of Muppet faces attracted gaze in younger infants and distracted them from fixating the human actor's face. We calculated the average saliency rank of the pixels around the point of gaze when participants fixated Muppet and human faces. Pixels around the point of gaze

were more salient when participants fixated Muppet faces ( $M = 82.5\%$ ,  $SD = 3.97$ ) compared to the human actor's face ( $M = 78.4\%$ ,  $SD = 5.61$ ), and this difference held across age groups. A 5 (age group)  $\times$  2 (ROI: human face, Muppet faces) ANOVA confirmed only a main effect of region,  $F(1,25) = 12.31$ ,  $p = .002$ .

### **Alternative saliency analyses display the same pattern of results**

The literature contains many methods for calculating visual saliency (Borji & Itti, 2013). Although our results might differ if we used a different model, an exhaustive test of all available saliency calculations would be impractical. Instead, we repeated our analyses and replaced the saliency values from the combined channel model with values derived from two single channel models, flicker and local contrast. To test models of flicker and local contrast, we converted each video frame to the CIE L\*a\*b colorspace to obtain luminance values for each pixel. Flicker was calculated by determining the magnitude of luminance change for each pixel between successive video frames. Similar to our original gaze-saliency calculation, we averaged the flicker values within a 1.5-degree radius around the point of gaze for each frame. Local contrast was defined as the standard deviation of luminance within the 1.5-degree region around the point of gaze. We recalculated the hierarchical regressions predicting  $ISC_a$  using flicker and local contrast instead of gaze saliency. The same pattern of results in Table 2 held when substituting either local contrast or flicker for our original gaze-saliency measurement.

We also asked whether a different method of calculating gaze saliency from the combined-channel model would lead to a different pattern of results. Instead of calculating the mean of saliency ranks in the region around the point of gaze, we asked how consistently the observer's eye movements tracked the most salient area of the image (Shepherd et al., 2010). Gaze-saliency correlations were calculated between an observer's eye gaze location and the

location of the most salient pixel (Figure 1, blue cross) in a manner similar to inter-subject correlations. This method led to the same pattern of results in Table 2 with two exceptions. First, age accounted for additional variance beyond face-looking and gaze-saliency correlations in the overall video. Second, both face-looking and gaze-saliency correlations accounted for significant variance in the Multiple Agents TOI. However, the association between gaze-saliency correlations and  $ISC_a$  was *negative*—infants who closely tracked the most salient pixel were less likely to show adult-like eye movements.

### Discussion

The current study investigated eye movement consistency in free viewing of dynamic stimuli in infants and adults. As in prior work (Frank et al., 2009; Kirkorian et al., 2012), eye movements became more consistent within age groups over development. Eye movement consistency was high between adult observers (Dorr et al., 2010; Hasson, Yang, et al., 2008; Mital et al., 2011; Shepherd et al., 2010; Smith & Mital, 2013; 't Hart et al., 2009), but young infants were more likely to “do their own thing” by looking at different regions of the stimulus. However, even the youngest infants (6-month-olds) showed some degree of consistency when viewing a complex scene, as evidenced by within-group correlations that exceeded chance levels. Moreover, we found that increasingly adult-like gaze patterns are responsible for gains in within-group consistency over development. With age, infants’ eye movements become more similar to those of adults.

#### What factors predict adult-like free viewing?

Infants display adult-like free viewing behavior when they prioritize visual features in the same way as do adults. As in past work, we found age-related changes in looking at bottom-up (salient regions) and top-down (faces) features (Amso et al., 2014; Frank et al., 2009). Adults

fixated the human actor's face more frequently than did infants and adults were marginally more likely than infants to fixate more visually salient regions. We took a novel approach by asking whether individual differences in infants' looking to these two features predicted adult-like gaze patterns. Across the entire video, infants who spent more time looking at the human actor's face and at salient regions showed more consistency with adult eye movements. Infants' age did not explain additional variance, indicating that face looking and gaze saliency accounted for most of the age-related variance.

Saliency was a strong predictor of adult-like free viewing. The current study and past work link gaze saliency and inter-subject consistency (Mital et al., 2011; Smith & Mital, 2013). Because saliency is confounded with higher-level features, the extent to which saliency predicts adults' eye movements may depend on the degree to which saliency is related to semantically informative regions of the stimulus. Indeed, gaze saliency was greater when fixating faces compared to non-face regions, which is characteristic of children's television (Wass & Smith, 2015). But if gaze saliency were simply the result of looking at faces, why did saliency account for variance in viewing patterns after accounting for face looking? One possibility is that the bottom-up features predicted both face and non-face regions that attracted eye movements. Adults spent the majority of the time looking at faces, but still spent considerable time looking at other areas, which may have been selected due to high saliency.

The roles of saliency and faces in predicting adult-like free viewing varied according to scene content. When the human actor was the only agent present, saliency was the lone predictor of adult-like gaze, presumably because saliency was high both when looking at the actor's face and when selecting other points of interest. However, when multiple agents were present, looking at the human actor's face predicted adult-like gaze, but saliency did not. The gaze-

saliency correlations suggested that adults might have avoided looking at the most salient region of the video. In contrast, young infants spent a great deal of time looking at the visually-salient faces of the Muppet characters, resulting in poor consistency with adults' gaze.

### **How does adult-like free viewing develop?**

The development of free viewing depends on where to look and when to look. At first glance, our results suggest that increased looking at both salient regions and faces account for the development of adult-like free viewing. But on closer inspection, it is clear that simply looking *more* to low- and high-level features is only part of the story. The degree to which low- and high-level features accounted for older infants' and adults' gaze depended on scene content. High salience sometimes leads to semantically informative areas but at other times can draw the eyes away from informative areas. Thus, infants need to learn *when* to look (or avoid looking) at different types of features. Many aspects of infants' visual processing of static stimuli become adult-like by about 6 months: eye movements while scanning simple shapes (Bronson, 1994), perception of object features (Colombo, 1990), and configural processing of faces (Cashon & Cohen, 1994). Yet, endogenous attention—sustaining attention and inhibiting attention shifts while distracted—follows a slower developmental course (Colombo, 2001). Susceptibility to distraction from visually salient areas in dynamic displays may relate to infants' immature endogenous attention skills. Free viewing may improve as infants become better able to inhibit attention to salient areas that lack relevance in the scene.

Moreover, better comprehension of narrative content may contribute to the development of free viewing of dynamic stimuli, especially when multiple top-down features compete for attention. When multiple faces were present in the current study, infants had to select *which* face to look at. Even adults displayed lower inter-subject consistency when multiple faces were

present. Most likely, comprehension of the narrative content aids mature observers in prioritizing where to look (Kirkorian et al., 2012; Pempek et al., 2010). Our stimulus might have been too short in duration (60 s) to present a clear story line like other investigations that used longer clips (3 minutes: Hasson, Yang, et al., 2008; Shepherd et al., 2010; 4-8 minutes: Goldstein, Woods, & Peli, 2007; 20 minutes: Kirkorian et al., 2012). Still, one minute of video is likely enough for mature observers to glean some useful top-down information that could guide looking: the context of singing a song, the meanings of the words in the song, and identifying which person is singing. Adults and older infants recognized that the human actor was the main character in the scene and continued to look at her face even when other, more salient faces were present in the image. An intriguing possibility is whether a more “infant-oriented” stimulus that contains cues that are more recognizable to infants would result in greater consistency among infant but not adult observers. Regardless, long-term accumulation of narrative content cannot be the only explanation for eye movement consistency because inter-subject correlations are observed in clips as short as 1-s in duration (Wang et al., 2012). Even without supporting context, visual features of a scene evoke reliable gaze in adults.

It is not possible to fully separate the influences of narrative comprehension, faces, and saliency on participants’ eye movements in the current study because they were inter-related. The human actor was the main focus of the video, and her face was often among the most salient regions in the stimulus. Although this makes it difficult to interpret the exact role of each factor, the high degree of overlap between these factors is likely to be representative of real life. The relation between attention and comprehension over development presents an intriguing challenge for future research. Do children attend to informative regions because they comprehend the content, or is their understanding enriched by a more adult-like strategy of parsing the scene? If



lower-level information is predictive of important, higher-level features, saliency may be a scaffold that young infants can exploit to learn how to parse a scene. Similar statistical learning mechanisms have been documented in other perceptual domains, such as language comprehension (Saffran, Aslin, & Newport, 1996).

### **Future directions**

It is important to consider how free viewing of television programs might generalize to real world gaze behavior. Measuring inter-subject consistency of eye movements required a screen-based experimental protocol to present the same stimulus to each observer. However, real world vision depends on coordinating the eyes, head, and body to actively select what is in view. Eye movements contribute to performing tasks in everyday life, not just to watching events. As in other screen-based tasks (Foulsham et al., 2010; Frank et al., 2009; Frank et al., 2011), we observed that observers spent large amounts of time looking at faces. But recent studies that used head-mounted eye tracking to measure infants' eye movements show that face looking is rare in natural tasks (Bambach, Franchak, Crandall, & Yu, 2014; Franchak et al., 2011; Yu & Smith, 2013). Spatial differences in viewing position, such as infants' short stature relative to adults, mean that faces are not always available in infants' field of view (Jayaraman, Fausey, & Smith, 2015; Kretch et al., 2014). And when infants are able to actively engage with their surroundings, manipulable objects draw a substantial amount of attention (Yu & Smith, 2013). Some real world contexts are more comparable to screen-based tasks: When infants are carried on the caregiver's chest in a forward-facing infant carrier, spatial differences are eliminated and infants look at faces with greater frequency (Kretch & Adolph, 2015). And, similar to the current study, saliency at the point of gaze is higher when infants fixate faces in this context. Future work should

integrate naturalistic and screen-based approaches to determine the influences on children's gaze behavior in different contexts and across development.

### References

- 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.
- Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist*, 48, 505-522.
- Bambach, S., Franchak, J. M., Crandall, D. J., & Yu, C. (2014). Detecting hands in children's egocentric views to understand embodied attention during social interaction. *Proceedings of the 36th annual meeting of the Cognitive Science Society*.
- Barr, R., & Hayne, H. (1999). Developmental changes in imitation from television during infancy. *Child Development*, 70, 1067-1081.
- Bronson, G. (1994). Infants' transitions toward adult-like scanning. *Child Development*, 65, 1243-1261.
- Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modelin. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 185-207.
- Cashon, C. H., & Cohen, L. B. (2004). Beyond U-shaped development in infants' processing of faces: An information-processing account. *Journal of Cognition and Development*, 5, 59-80.
- Cerf, M., Harel, J., Einhauser, W., & Koch, C. (2007). Predicting human gaze using low-level saliency combined with face detection. *Advances in Neural Information Processing Systems*, 20, 241-248.
- Colombo, J. (2001). The development of visual attention in infancy. *Annual Reviews in Psychology*, 52, 337-367.

- Colombo, J., Mitchell, D. W., Coldren, J. T., & Atwater, J. D. (1990). Discrimination learning during the first year of life: Stimulus and positional cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 98-109.
- Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, 10, 1-17.
- Einhauser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, 8, 1-19.
- Einhauser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, 8, 1-26.
- Elzary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, 8, 1-15.
- Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, 117, 319-331.
- Franchak, J. M., & Adolph, K. E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research*, 50, 2766-2774.
- 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.
- Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, 110, 160-170.
- Frank, M. C., Vul, E., & Saxe, R. (2011). Measuring the development of social attention using free-viewing. *Infancy*, 17, 355-375.
- Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy*, 14, 550-562.

Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants.

*Frontiers in Psychology*, 4, 1-7.

Goldstein, R. B., Woods, R. L., & Peli, E. (2007). Where people look when watching movies: Do

all viewers look at the same place? *Computers in Biology and Medicine*, 37, 957-964.

Gredebäck, G., Eriksson, M., Schmitow, C., Laeng, B., & Stenberg, G. (2012). Individual

differences in face processing: Infants' scanning patterns and pupil dilations are

influenced by the distribution of parental leave. *Infancy*, 17, 79-101.

Hasson, U., Avidan, G., Gelbard, H., Vallines, I., Harel, M., Minshew, N., & Behrmann, M.

(2009). Shared and idiosyncratic cortical activation patterns in autism revealed under

continuous real-life viewing conditions. *Autism Research*, 2, 220-231.

Hasson, U., Yang, E., Vallines, I., Heeger, D. H., & Rubin, N. (2008). A hierarchy of temporal

receptive windows in human cortex. *Journal of Neuroscience*, 28, 2539-2550.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger D. J. (2008)

Neurocinematics: The neuroscience of films. *Projections: The Journal for Movies and*

*Mind*, 2, 1-26.

Hayhoe, M. M., & Rothkopf, C. A. (2011). Vision in the natural world. *WIREs Cognitive*

*Science*, 2, 158-166.

Hayhoe, M. M., Shrivastava, A., Mruczek, R. E. B., & Pelz, J. B. (2003). Visual memory and

motor planning in a natural task. *Journal of Vision*, 3, 49-63.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in*

*Cognitive Sciences*, 7, 498-504.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, 16,

219-222.

- Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, 12, 1093-1123.
- Itti, L., Koch, B., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 1254-1259.
- Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PLoS ONE*, 10, e0123780.
- Kirkorian, H. L., Anderson, D. R., & Keen, R. (2012). Age differences in online processing of video: An eye movement study. *Child Development*, 83, 497-507.
- Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, 59, 809-816.
- Kretch, K. S., & Adolph, K. E. (2015). Active vision in passive locomotion: Real-world free viewing in infants and adults. *Developmental Science*, 18, 206-218.
- Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, 85, 1503-1518.
- Land, M. F. (2004). The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research*, 159, 151-160.
- Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41, 3559-3565.
- Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28, 1311-1328.

- Libertus, K., & Needham, A. (2011). Reaching experience increases face preference in 3-month-old infants. *Developmental Science*, 14, 1355-1364.
- Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, 3, 5-24.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42, 107-123.
- Pempek, T. A., Kirkorian, H. L., Richards, J. E., Anderson, D. R., Lund, A. F., & Stevens, M. (2010). Video comprehensibility and attention in very young children. *Developmental Psychology*, 46, 1283-1293.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-olds infants. *Science*, 274, 1926-1928.
- Shepherd, S. V., Steckenfinger, S. A., Hasson, U., & Ghazanfar, A. A. (2010). Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Current Biology*, 20, 649-656.
- Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, 13, 1-24.
- 't Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., Konig, P., & Einhauser, W. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition*, 17, 1132-1158.
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11, 1-23.
- Valkenburg, P. M., & Vroone, M. (2004). Developmental changes in infants' and toddlers' attention to television entertainment. *Communication Research*, 31, 288-311.

- Wang, X. H., Freeman, J., Merriam, E. P., Hasson, U., & Heeger, D. H. (2012). Temporal eye movement strategies during naturalistic viewing. *Journal of Vision*, 12, 1-27.
- Wass, S. V., & Smith, T. J. (2015). Visual motherese? Signal-to-noise ratios in toddler-directed television. *Developmental Science*, 18, 24-37.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.
- 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.



*Table 1.* Mean inter-subject correlations within and between each age group. Correlations along the diagonal (in bold) are within-age correlations ( $ISC_w$ ). Correlations between each age group and comparison adults (bottom row) were used to calculate the  $ISC_a$  measure. All correlations were significant at the .05 level using time-randomized baseline tests.

	6 months	9 months	12 months	24 months	Adults	Comparison Adults
6 months	<b>.381</b>					
9 months	.380	<b>.396</b>				
12 months	.400	.326	<b>.520</b>			
24 months	.328	.357	.455	<b>.527</b>		
Adults	.318	.332	.486	.495	<b>.598</b>	
Comparison Adults	.374	.318	.517	.487	.597	<b>.600</b>

*Table 2.* Hierarchical regressions predicting infants' ISC<sub>a</sub> values from face looking, average saliency rank at fixated locations, and age across the entire video, One Agent time of interest, and Multiple Agents time of interest.  $R^2$  change statistics show the additional variance accounted for as each predictor is entered into the model. Predictors in italics were excluded from the final model.

Predictor	Overall		One Agent		Multiple Agents	
	$R^2$	$R^2$ change	$R^2$	$R^2$ change	$R^2$	$R^2$ change
Looking at human actor's face	.225	.225*	.114	.114	.161	.161†
Gaze saliency	.413	.189*	.311	.197*	.162	.001
Age at testing	.432	.019	.340	.029	.248	.087
Final model	$F(2,23) = 7.40, p = .004$		$F(2,23) = 4.74, p = .020$		$F(1,23) = 4.32, p = .052$	

†  $p < .10$ , \*  $p < .05$ , \*\*  $p < .01$

### Figure Captions

*Figure 1.* Example frames from the (A) “One Agent” and (B) “Multiple Agents” times of interest. Fixation locations are displayed for each individual 6-month-old (orange +’s) and adult (green +’s) observer. White ellipses indicate regions of interest for human and Muppet faces. The most salient pixel is marked by a blue cross.

*Figure 2.* (A) Example horizontal eye movement time series for two adults, A1 and A2, resulting in a strong inter-subject correlation (ISC). (B) Averaging ISCs from each pair of observers within an age group (each line represents an ISC) yielded a measure of within-group correlation ( $ISC_w$ ). Thick lines highlight ISCs calculated for one observer, A1, paired with observers A2-A6. Dashed circles highlight the observers used in panel (A), A1 and A2. (C) Example horizontal ISC between a 6-month-old (B1) and an adult (A1) resulting in a weak correlation. (D) Averaging ISCs from pairs of observers in two different age groups yielded a measure of between-group correlation. Thick lines illustrate ISCs calculated for one observer, B1, paired with observers A1-A6 of a different age group. Dashed circles highlight the observers used in panel (C), B1 and A1.

*Figure 3.* (A) Within-age eye movement consistency ( $ISC_w$ ), calculated across the entire stimulus video, and averaged across individuals in each age group. (B) Eye movement consistency between each age group and adults ( $ISC_a$ ), averaged across individuals in each age group. (C) Gaze saliency averaged across individuals in each age group. (D) Face looking averaged across individuals in each age group.

*Figure 4.* (A) Relation between gaze saliency and eye movement consistency with adults ( $ISC_a$ ), calculated across the entire stimulus video. (B) Relation between face looking and  $ISC_a$ . Each symbol shows data from one infant (open symbols) or adult (closed symbols).

*Figure 5.* (A) Within-age eye movement consistency ( $ISC_w$ ), calculated separately for the “One Agent” (white symbols) and “Multiple Agents” (gray symbols) times of interest, and averaged across individuals in each age group. (B) Eye movement consistency between each age group and adults ( $ISC_a$ ), averaged across individuals in each age group. (C) Gaze saliency averaged across individuals in each age group. (D) Face looking averaged across individuals in each age group. Gray triangles indicate the proportion of time looking at Muppet faces during the “Multiple Agents” time of interest.

A. One agent time of interest (TOI)

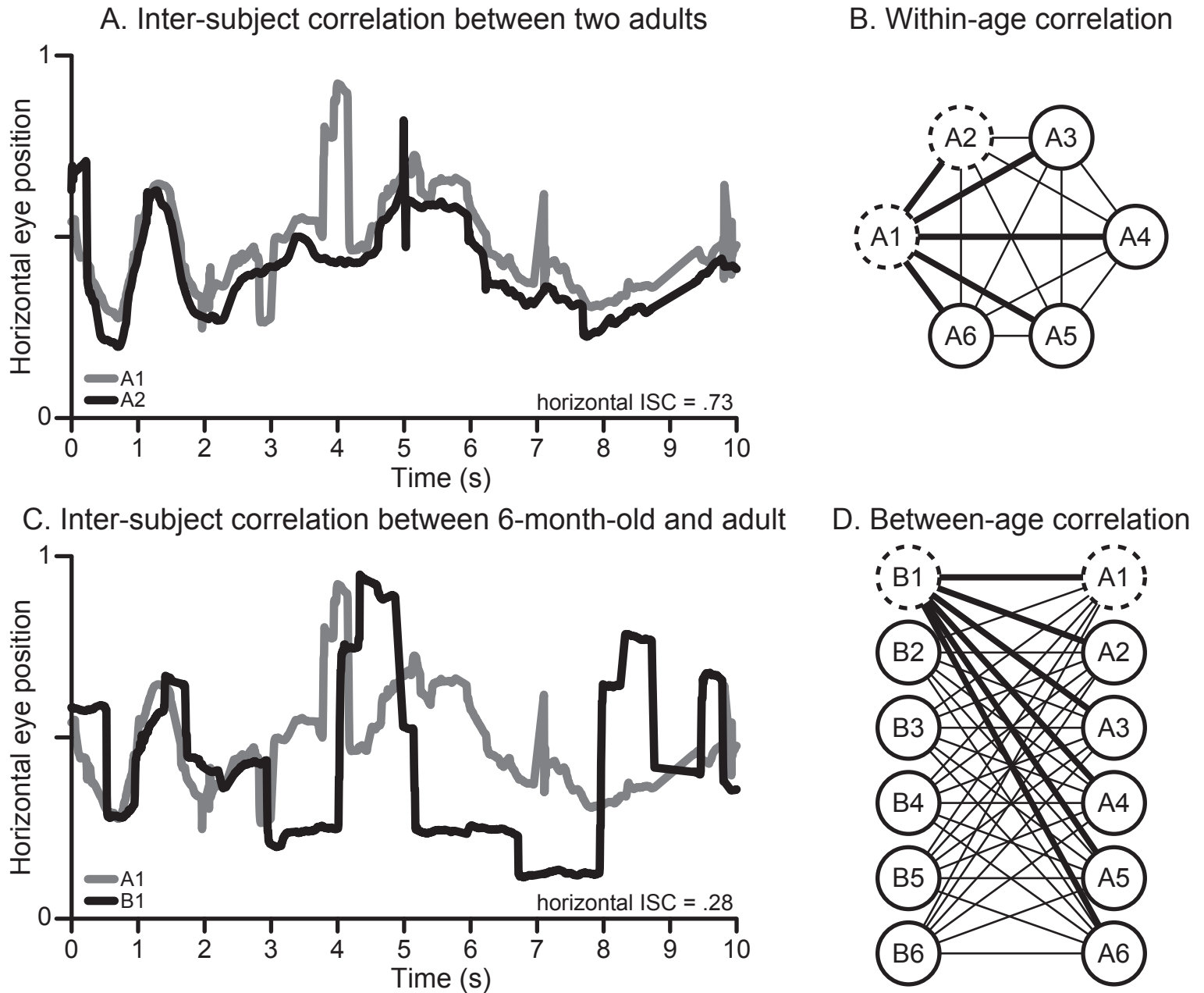


B. Multiple agents time of interest (TOI)

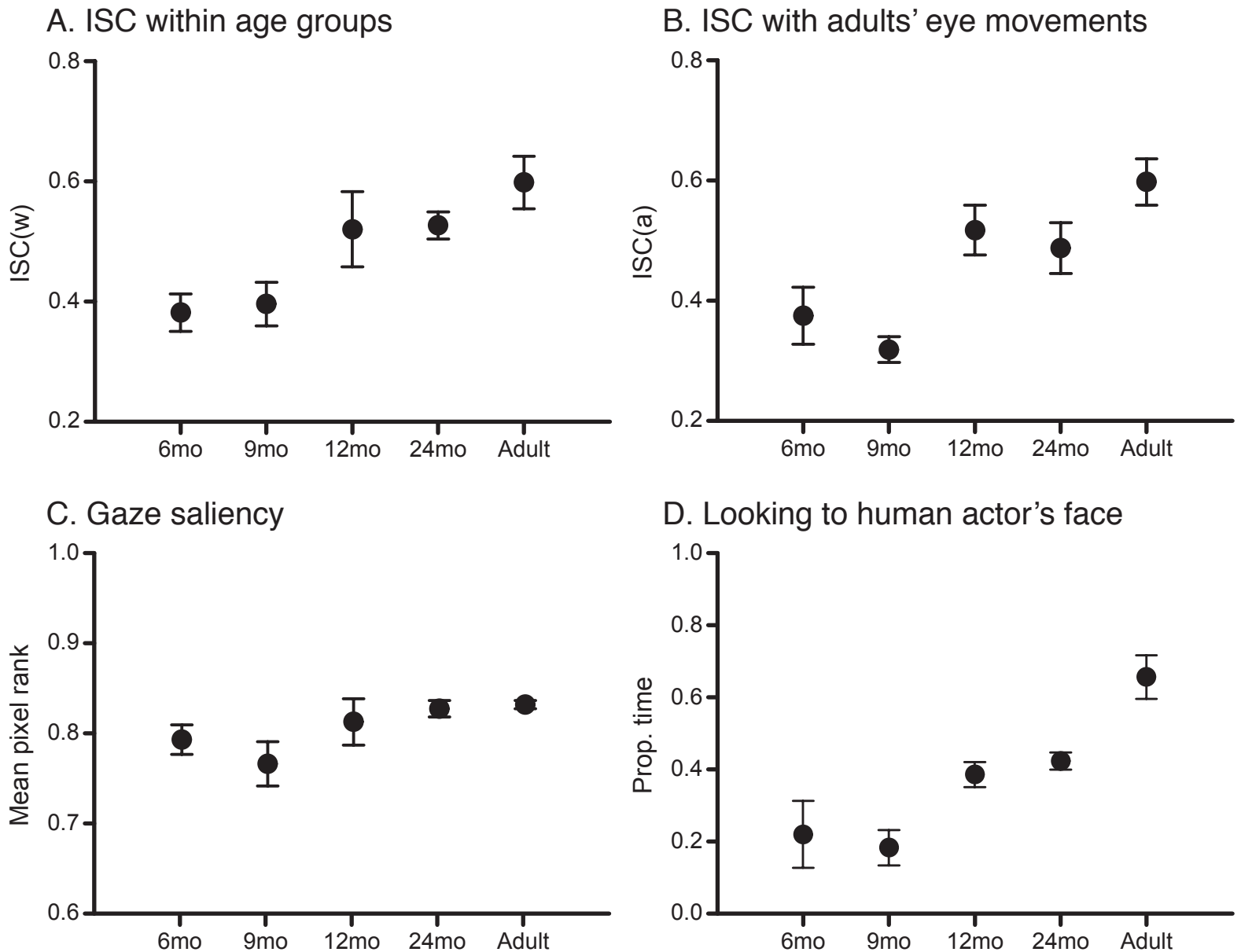


- ✚ 6-month-old gaze location
- ✚ Adult gaze location
- ✕ Most salient pixel

*Figure 1.* Example frames from the (A) “One Agent” and (B) “Multiple Agents” times of interest. Fixation locations are displayed for each individual 6-month-old (orange +’s) and adult (green +’s) observer. White ellipses indicate regions of interest for human and Muppet faces. The most salient pixel is marked by a blue cross.

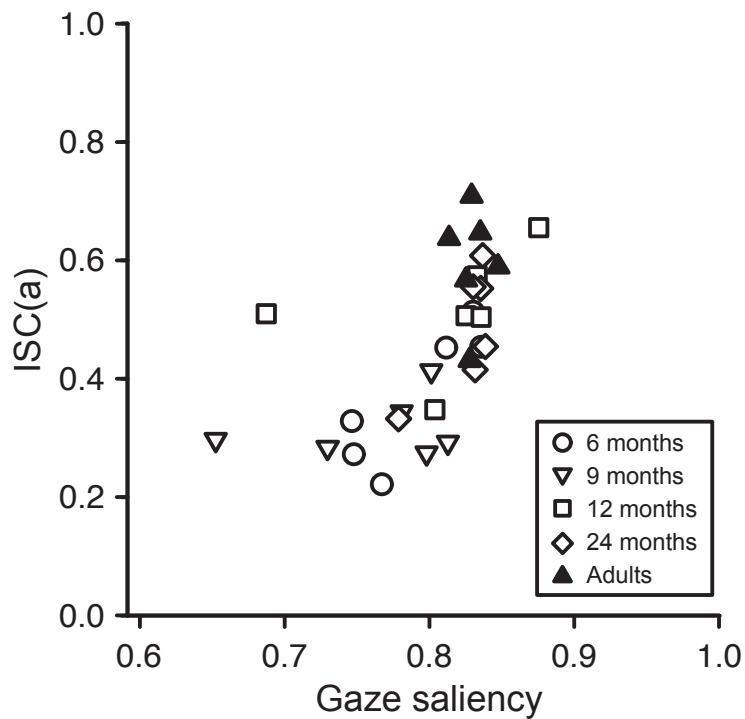


**Figure 2.** (A) Example horizontal eye movement time series for two adults, A1 and A2, resulting in a strong inter-subject correlation (ISC). (B) Averaging ISCs from each pair of observers within an age group (each line represents an ISC) yielded a measure of within-group correlation (ISCw). Thick lines highlight ISCs calculated for one observer, A1, paired with observers A2-A6. Dashed circles highlight the observers used in panel (A), A1 and A2. (C) Example horizontal ISC between a 6-month-old (B1) and an adult (A1) resulting in a weak correlation. (D) Averaging ISCs from pairs of observers in two different age groups yielded a measure of between-group correlation. Thick lines illustrate ISCs calculated for one observer, B1, paired with observers A1-A6 of a different age group. Dashed circles highlight the observers used in panel (C), B1 and A1.

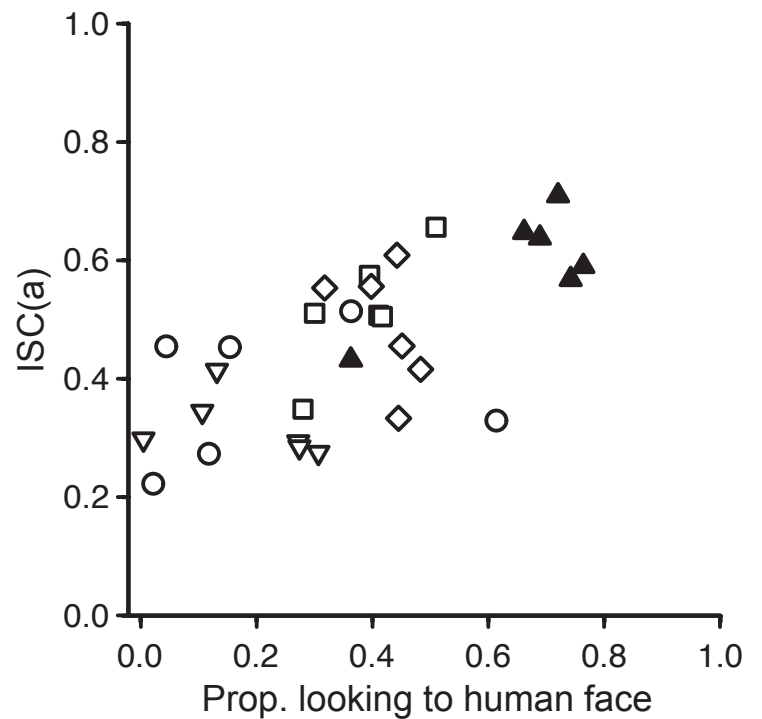


*Figure 3.* (A) Within-age eye movement consistency (ISCw), calculated across the entire stimulus video, and averaged across individuals in each age group. (B) Eye movement consistency between each age group and adults (ISCa), averaged across individuals in each age group. (C) Gaze saliency averaged across individuals in each age group. (D) Face looking averaged across individuals in each age group.

A. Gaze saliency predicts ISC(a)



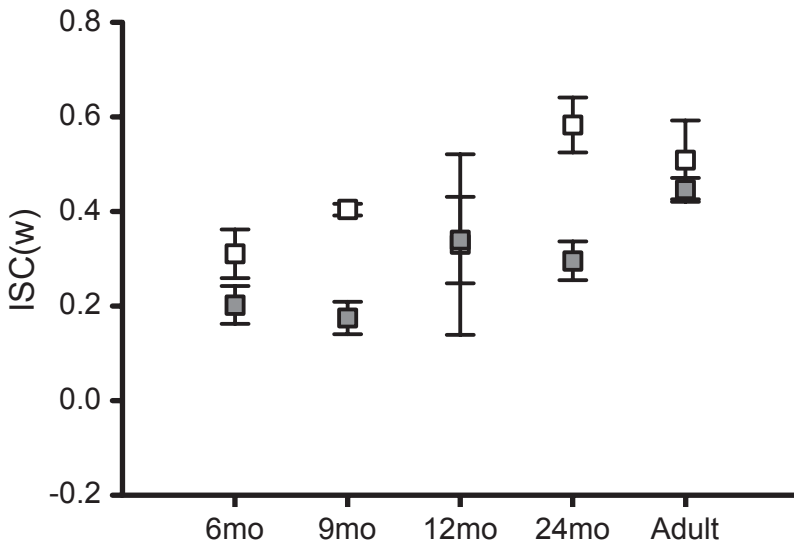
B. Looking at human face predicts ISC(a)



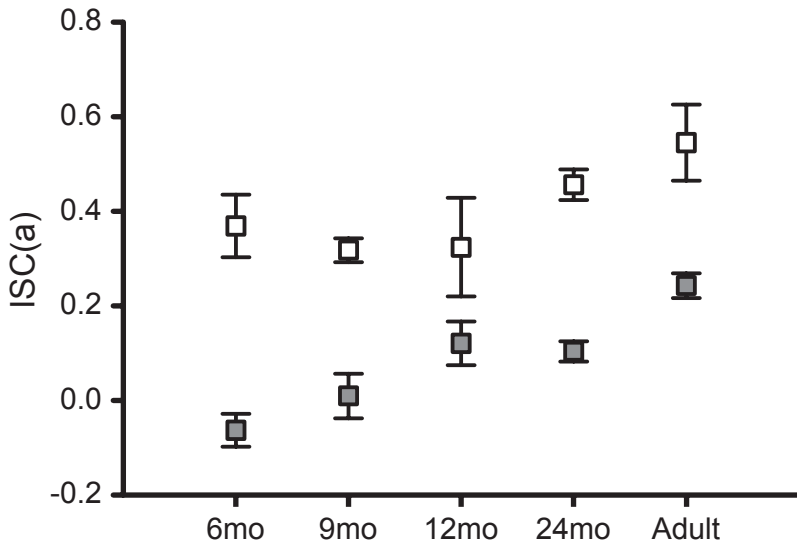
*Figure 4.* (A) Relation between gaze saliency and eye movement consistency with adults (ISCa), calculated across the entire stimulus video. (B) Relation between face looking and ISCa. Each symbol shows data from one infant (open symbols) or adult (closed symbols).



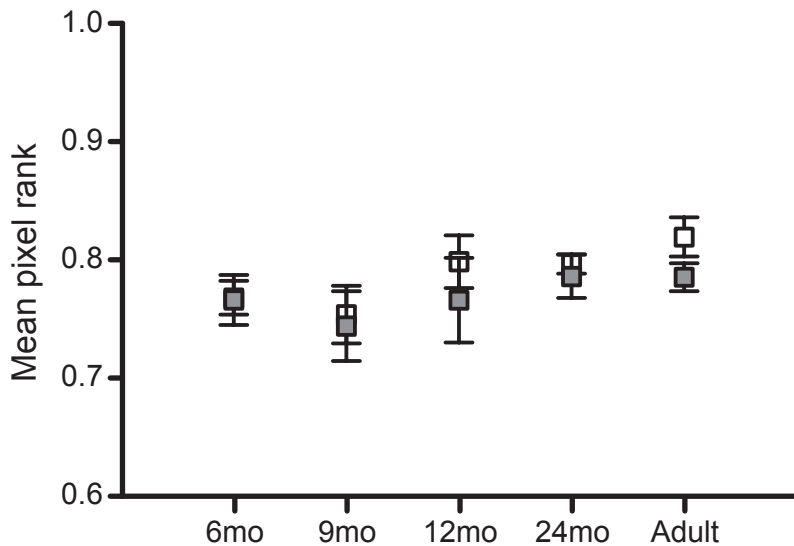
A. ISC within age groups



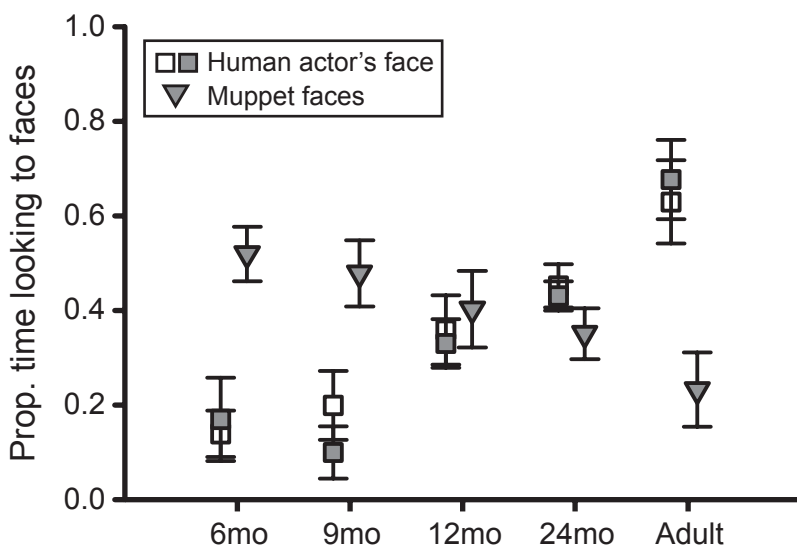
B. ISC with adults' eye movements



C. Gaze saliency



D. Looking to faces



□ One Agent  
■ Multiple Agents

Figure 5. (A) Within-age eye movement consistency (ISCw), calculated separately for the “One Agent” (white symbols) and “Multiple Agents” (gray symbols) times of interest, and averaged across individuals in each age group. (B) Eye movement consistency between each age group and adults (ISCa), averaged across individuals in each age group. (C) Gaze saliency averaged across individuals in each age group. (D) Face looking averaged across individuals in each age group. Gray triangles indicate the proportion of time looking at Muppet faces during the “Multiple Agents” time of interest.