



Published in final edited form as:

Dev Psychol. 2021 July ; 57(7): 1025–1041. doi:10.1037/dev0001020.

Developmental changes in natural scene viewing in infancy

Katherine I. Pomaranski^{1,2}, Taylor R. Hayes², Mee-Kyoung Kwon³, John M. Henderson^{1,2},
Lisa M. Oakes^{1,2}

¹Department of Psychology, University of California, Davis

²Center for Mind and Brain, University of California, Davis

³Psychology Department, University of Utah, Asia Campus

Abstract

We extend decades of research on infants' visual processing by examining their eye gaze during viewing of natural scenes. We examined the eye movements of a racially diverse group of 4- to 12-month-old infants ($N = 54$; 27 boys; 24 infants were White and not Hispanic, 30 infants were African American, Asian American, mixed race and/or Hispanic) as they viewed images selected from the MIT Saliency Benchmark Project. In general, across this age range infants' fixation distributions became more consistent and more adult-like, suggesting that infants' fixations in natural scenes become increasingly more systematic. Evaluation of infants' fixation patterns with saliency maps generated by different models of physical salience revealed that although over this age range there was an increase in the correlations between infants' fixations and saliency, the amount of variance accounted for by salience actually decreased. At the youngest age, the amount of variance accounted for by salience was very similar to the consistency between infants' fixations, suggesting that the systematicity in these youngest infants' fixations was explained by their attention to physically salient regions. By 12 months, in contrast, the consistency between infants was greater than the variance accounted for by salience, suggesting that the systematicity in older infants' fixations reflected more than their attention to physically salient regions. Together these results show that infants' fixations when viewing natural scenes becomes more systematic and predictable, and that predictability is due to their attention to features other than physical salience.

Keywords

infancy; eye movements; natural scene viewing; physical salience

For decades researchers have examined where and how long infants look at controlled experimental stimuli. This work has been revealing, and has shown that infants' eye movements reflect active planning (see Canfield & Kirkham, 2001 for review), sophisticated coordination of eye and head movements (von Hofsten & Rosander, 1997), and an interest in exploring novel or new information (Fantz, 1964). Moreover, the factors that control infants' looking to such stimuli changes over age (see Colombo, 2001 for review). For

example, Frank et al. (2009) found that while watching a Peanuts cartoon, younger infants looked more at areas of high physical salience, whereas older infants looked more at faces. Similarly, Kwon et al. (2016) found that when presented with images of 6 familiar items (e.g., a shoe, a sippy cup, a human face) presented in a circular array, 4-month-old infants looked first at physically salient items and 6- and 8-month-old infants looked first at the human face (which was never the most physically salient item) more than expected by chance (see also Gluckman & Johnson, 2013). Results like these have been used to confirm the conclusion that young infants' attention is more driven by physical salience than older infants' attention and that infants' attention becomes more driven by top-down forces (e.g., an interest in and recognition of the importance of the socially relevant face) over development.

The focus on controlled experimental stimuli, however, cannot provide as much understanding into infants' visual behavior when faced with more naturalistic, real-world contexts. That is, although the literature on infants' looking behavior has yielded important insights into the development of the visual system, an important question is whether these observations translate to their eye gaze when viewing more complex, naturalistic scenes.

Because natural scenes--e.g., photographs of everyday, real-world scenes--are more like the kind of visual stimuli people view in their everyday life, studying eye gaze when viewing such stimuli will provide more insight into typical visual behavior.

Decades of research has examined adults' eye movements as they view this type of natural scenes (Henderson, 2007, 2017; Henderson & Hollingworth, 1999; Rayner, 2009; Yarbus, 1967). This research has yielded both similarities and differences in how adults view more and less complex visual stimuli (Clifton et al., 2016). Adults' visual attention when viewing natural scenes is influenced by local physical salience (Itti, 2005; Koehler et al., 2014), local semantic properties including those related to objects (Henderson & Hayes, 2017, 2018; Stoll et al., 2015), global scene context (Malcolm & Henderson, 2010; Torralba et al., 2006), and viewing task (Castelhano et al., 2009; Hayhoe & Ballard, 2005; Land et al., 1999).

Recently, researchers have begun to focus on the developmental origins of these eye gaze patterns and have examined infants' eye movements as they view natural scenes (e.g., Mahdi et al., 2018; van Renswoude, Visser, et al., 2019). This work suggests that infants' fixations when viewing natural scenes develops over the first year. Specifically, across the first year infants' pattern of fixations become more adult-like (Helo et al., 2016; van Renswoude, Visser, et al., 2019), and their saccades become more precise (van Renswoude et al., 2016). However, this literature has not yielded a consistent picture of how the factors that determine fixations in natural scenes change over development. In part, this reflects the fact that no study to date has undertaken a comprehensive evaluation of the developmental changes in infants' gaze patterns as they viewed naturalistic scenes. Instead, previously published studies have tended to focus more narrowly on specific questions.

For example, two studies examined how infants' detected salient and non-salient faces in natural scenes. Amso et al. (2014) found no effect of salience on detection of faces in the first year. In this study, eye gaze was recorded as infants viewed natural, real-world scenes

(e.g., people sitting in an office space). In some of the scenes the human face present was the most physically salient region and in other scenes the human face present was not the most physically salient region. Although infants' gaze to the faces increased across the first year, they were no more likely to fixate salient faces than to fixate non-salient faces. In a similar study, Kelly et al. (2019) observed that infants between 3 and 12 months were sensitive to the presence of a face in natural scenes (both indoor and outdoor scenes). However, Kelly et al. found that infants were more sensitive to salient faces than to non-salient faces. This effect seemed to be more pronounced for the younger infants, a pattern consistent with the general developmental trajectory observed when infants are shown more controlled experimental stimuli. Several differences between these studies, including the stimulus set, likely contributed to the differences in findings. What is most important for the present discussion is that neither study examined the effect of salience comprehensively, but rather asked how the salience of a human face contributed to infants' visual behavior as they viewed natural scenes.

van Renswoude and colleagues have also examined infants' gaze behavior while free viewing real-world scenes. In one study, van Renswoude, van den Berg et al. (2019) measured the location of infants' first fixation to photographs of scenes. Before presenting the scenes, a fixation dot was presented to one side of the center. Nevertheless, infants' first fixations were more likely to the center of the screen, suggesting a strong center bias. This center bias interacted with physical salience, however. Infants were less likely to fixate the center if the periphery was more physically salient. In another study, van Renswoude, Visser et al. (2019) analyzed both the location and duration of infants' fixations during free viewing of photographs of real-world scenes. Overall infants had a center bias and were drawn to look at physically salient regions. In addition, over the first year, where infants' looked was increasingly predicted by where adults looked.

Although these studies provide an important step in understanding the development of how infants' eye gaze to natural scenes develops, there are several gaps that need to be addressed. A significant gap is that all of the studies discussed relied on a variation of one model of salience. Specifically, each of the studies described earlier estimated salience using a version of the Itti and Koch (e.g., Itti et al., 1998) model of saliency. The advantage of this model is that it is based on the primate neural architecture and allows to evaluate bottom-up salience based on stimulus features of intensity, color, and orientation. However, this is only one salience model, and it is possible that it is not the best model for explaining infants' eye gaze. Mahdi et al. (2018) did compare the effectiveness of different salience models at predicting infants' fixations of natural scenes. Because the sample size was small, further work is needed to determine how well different models account for infants' gaze patterns. To be clear, the focus in the infant literature has been on the role of bottom-up salience on explaining infants' gaze patterns, in contrast to other deep models that have used a combination of bottom-up features and more top-down features acquired through deep learning (e.g., Bruckert et al., 2019; Damiano et al., 2019; Rahman & Bruce, 2016). Thus, in the present study we focus on traditional saliency models that will inform us about whether infants' looking is driven by bottom-up stimulus factors as specified in those models.

In addition, although van Renswoude and colleagues (van Renswoude, van den Berg, et al., 2019; van Renswoude, Visser, et al., 2019) examined both center bias and salience in their studies, they did not systematically examine the effects of the center bias in the salience models on how well those models predicted infants' looking behavior. Here we adopted an approach used by Hayes and Henderson (2020) in which they compared the ability of a salience model and the center bias inherent in that salience model to predict adults' fixations.

Moreover, the work on infants' scene viewing has not taken into account the consistency of infants' looking behavior, and how that consistency changed over age. A common observation in the scene viewing literature is that adults are consistent in their fixation patterns, but that the level of consistency varies with features of the stimuli (Judd et al., 2011; Yun et al., 2013) (This has also been referred to as inter-observer visual congruency, e.g., Le Meur et al., 2011; Rahman & Bruce, 2016). That is, adults' eye gaze tends to land on the *same* regions of a scene. Developmentally, infants appear to become more consistent in their fixations, at least when viewing dynamic stimuli, such as a Peanuts movie or clips from Sesame Street (Franchak et al., 2016; Frank et al., 2012; Rider et al., 2018). This means that younger infants are more idiosyncratic in where their fixations are located, and over age infants increasingly fixate on similar locations of a stimulus. Thus, understanding the development of the consistency of infants' fixations on natural scenes will reveal if they become more systematic in how they view those scenes with age. In addition, because the consistency between observers provides an estimate of a noise ceiling, theoretically limiting how well a model can be expected to predict the behavior (e.g., Chen & Zelinsky, 2019), changes in consistency provide an important benchmark for interpreting the relation between eye gaze and other factors.

In the present work we extend the work on infants' eye movements during scene viewing using scenes from the MIT Saliency Benchmark (Judd et al., 2012), an online source of saliency model performance based on a sample of adults' eye gaze during scene viewing, and analytical methods similar to those used by Bylinskii et al. (2019). Note, there are other ways of evaluating patterns of eye gaze (e.g., Le Meur et al., 2017), but we selected this approach to allow for more direct comparisons between our developmental work and the literature on adult scene viewing, and will complement and extend the previous work in the literature on both infant and adult scene viewing.

Method

Participants

Our final sample included 54 healthy, full-term infants, ranging from 4 to 12 months (118 – 373 days, *Median* = 247.50 days; 27 boys), with no history of neurological or vision problems. Infants' ages were distributed evenly across this age range (see Figure 1). Because this was the first investigation of this kind with infants, it was not possible to conduct a power analysis to establish an appropriate sample size. We used what is known from the adult literature using analyses like those conducted here to identify a sample size that would be appropriate.

To achieve our final sample, we tested a total of 102 infants. We excluded the data from 48 of the infants (116–377 days, *Median* = 207 days) because of fussiness ($N = 4$; crying, refusing to look at the screen, turning toward the parent), failure to calibrate ($N = 18$), poor tracking of gaze despite adequate calibration (i.e., very low *track ratios*, as described in the section below, $N = 10$), failure to contribute at least 12 of the 24 trials despite successful calibration ($N = 14$, usually because the infant became fussy or generally uninterested in the stimuli), parental interference ($N = 1$), or experimenter error ($N = 1$).

It should be noted that we excluded a larger than typical number of infants from our analyses due to poor tracking or a failure to contribute sufficient data (24 infants). Because the data used in our analyses were the raw X and Y coordinates at each time point (rather than whether or not fixations fell into broadly defined AOIs), our conclusions required higher spatial precision than is often required in infant eye tracking studies. As a result, we eliminated a higher proportion of infants due to poor gaze tracking than is seen in many infant eye tracking studies. In addition, our analytic approach required a relatively large amount of data from each participant. That is, to be certain that any effect reflected patterns that were characteristics of infants *in general* (and not their responses to the features of a particular image), we needed data from the same infants looking at many different images. And, because we were interested in whether there was consistency in the looking activity across participants, it was necessary that we recorded infants' looking at the same set of images. As a result, we included only infants who had viewed more than half of the images.

None of the infants in our final sample were at risk for colorblindness based on family history (e.g., we excluded boys whose maternal grandfather was colorblind). Thirty-one infants were White, 4 were African American, 1 were Asian American, 12 were mixed race, 3 were reported as other, and 3 did not have race reported. Across these groups, 18 of the infants were reported to be Hispanic (7 White, 2 African American, 5 mixed race, 2 other, 2 no race reported). Of the mothers who provided information about their highest level of education ($N = 52$), all had graduated from high school and 28 of them had earned at least a bachelor's degree (the demographics of the final sample, as well as the infants who were excluded from the final analyses, can be found at (<https://osf.io/5j4ht/>)).

Names of infants were originally obtained from the State Office of Vital Records, and all parents with an address within a 30-mile radius of the lab were sent informational mailings. Parents who wished to volunteer for studies contacted us and were included in our database. When infants reached the appropriate age for this study, we contacted them about participating. All infants who participated (whether we included their data or not) received a small toy or T-shirt and a certificate in appreciation for their time.

To provide a baseline for comparisons to our infant sample, we collected data from 24 adults ($M = 21.73$ years, $SD = 2.72$, range 18.61 to 28.96 years, 6 men), recruited from students at University of California at Davis, who participated as part of a course. Ten participants reported as White, 6 reported as Asian American, 2 reported as native Hawaiian or Pacific Islander, 2 reported as other, and 4 chose not to report race. An additional two participants were tested but excluded from our adult sample because the monitor stand was not appropriately adjusted for participant height.

Apparatus

We measured eye movements using a SMI-RED N eye tracker, capturing eye gaze at a rate of 120 Hz. The eye tracker was attached to the bottom of a 22-in LCD monitor, that had a web-camera attached to the top to record the participants' head and body position throughout the duration of the experiment. The monitor was affixed to an ergo arm that allowed the experimenter to position it to optimally locate each infant's eyes in the center of the detection radius of the eye tracking system (Figure 2). A Dell laptop supplied by SMI was used to monitor the participant and run the experiment. A large white cloth screen was placed behind the SMI-RED M monitor to obstruct the infant's view of the additional equipment and the experimenter.

Adults were tested using the same equipment and testing space except that the monitor (with eye tracker attached) was attached to a traditional desk stand, and was adjusted, if necessary, by placing textbooks beneath the stand to achieve the appropriate height to locate the participant's eyes. Participants were seated (on chairs without wheels to prevent them from altering their position) so their eyes were approximately 60 cm from the monitor. A white poster board trifold was placed behind the monitor, blocking the view of the additional equipment and the experimenter sitting on the opposite side of the desk.

Stimuli

The stimulus images were 24 scene photographs selected from the MIT saliency benchmark study (Judd et al., 2012). The stimulus images completely filled the monitor (approximately 48 cm wide, 30 cm high, 1680 X 1050 resolution), and thus were approximately 46 X 28 degrees of visual angle at a viewing distance of 60 cm (and thus about 35 pixels per degree). This set of images varied in contrast, luminance, and colorfulness and was selected to have equal representation with and without people. Our final analyses were conducted on 22 of these images (see example images in Figure 3; the full set can be found in the supplementary materials at <https://osf.io/5j4ht/>). We excluded 2 images from our final analyses because fewer than 20% of the infants provided eye gaze data to those images.

As a sanity check, we compared the results of saliency toolboxes with the fixations of our adult sample. Our adults' fixations (using Pearson's Linear Correlation Coefficient) were similar to those reported in the MIT Saliency Benchmark study (see Table 1), giving us confidence that our results, in general, are consistent with those that would be obtained in other labs.

We also showed participants 10-second video clips taken from popular children's television shows (e.g., *Sesame Street*, *Baby Einstein*) periodically throughout the session. The purpose of the video clips was to provide infants with a break in between the pictures to maintain interest in the task in general.

Procedure.—The protocol was reviewed and approved by the University of California at Davis, IRB protocol “Understanding cognitive development in infancy: Attention and visual short-term memory” (protocol number 220219–50). We used standard procedures to obtain informed consent from the parent or legal guardian of each infant and from each

adult participant. Because we were interested in infants' natural free-viewing eye gaze, we instructed parents to interact with their infant as little as possible, remain quiet, and not to direct their infant's attention to the screen.

Infants and parents were escorted to a sound attenuated testing room. Infants sat on their parent's lap or in a highchair (with parent nearby), positioned so their eyes were approximately 60 cm from the monitor (the parent's chair lacked wheels to prevent subjects from altering their position after the initial placement). The experimenter located the infants' eyes using the SMI software iView. The position of the eye tracker was adjusted using the ergo arm until two steady reflections of the cornea and pupil were detected for both eyes. During this procedure, we played a cartoon video on the computer monitor to keep the infant focused on the screen.

Once the infant was positioned properly, a 5-point calibration procedure was initiated. We used stimuli expected to be visually interesting to infants (a looming circle for calibration and a ducky accompanied by a chirp during validation). The stimulus was presented at the center and the four corners of the display. Calibration progressed automatically; once the eye tracker detected gaze in the target location, the stimulus moved to the next location. The validation procedure began immediately after calibration; the chirping ducky was presented in the four corners of the screen to validate the accuracy of the calibration. Following validation, the experimenter received feedback about the quality of the calibration via an image containing a red dot next to each validation stimulus location (Figure 2c). The location of the dot represents the calculated fixation locations for the left and right eye during validation. If the dots had on average greater than 2° deviation from the validation stimulus' locations for either the X or Y position, we assumed there was an error during the calibration (e.g., the participant was not looking at the calibration stimuli, the eye tracker picked up a reflection from somewhere other than the participant's eyes). Thus, the experimenter would repeat calibration until the recorded infant's gaze was within approximately 2° of each of the four validation stimulus locations. Once calibration was complete, parents were instructed to wear felt-covered sunglasses so they would be unable to see the stimuli presented on the screen during the experimental trial to minimize bias. Then we presented the experimental trials.

The experimental trials were free-view trials; on each trial a single image was presented for 5 s. Trials were presented in blocks of four, with order randomized within each block. Each experimental trial was preceded by a flashing fixation crosshair presented at the center of the screen, accompanied by attention-grabbing sounds (i.e., bells, rattle). The fixation cross remained on until the SMI system detected a fixation within approximately 5° of the fixation crosshair for 200 ms; this triggered the start of an experimental trial during which one of the images was presented for five seconds. As is common in eye tracking studies with infants (Oakes & Ellis, 2013; Schlegelmilch & Wertz, 2019), the experimental trial was paired with classical music to encourage infants to attend the screen (the sound came from behind the monitor). Following each block of 4 trials, a short video clip was presented to maintain infants' interest and to reduce the chance of fussiness throughout the experiment (an example of trial sequences, with an infants' eye gaze superimposed, can be found at <https://osf.io/5j4ht/>). We used the same experimental procedure with the adult participants.

Data processing

The raw eye gaze data were processed in several steps before conducting our main analyses.

Track Ratio.—We utilized the SMI's BeGaze analysis software to extract the track ratio, a statistic that quantifies the stability of the track for each trial for each participant. The track ratio is simply the percentage of samples (i.e., 8.33 ms time periods) across the trial that produced non-zero gaze positions. Thus, it is a measure of *data robustness*, and if we were able to perfectly measure a participant's eye gaze (and they were looking at the stimulus for the full trial), that participant would receive a track ratio of 100, representing 0% track loss.

Fixation Characteristics.—We utilized the SMI's BeGaze analysis software to filter the eye movements into fixations, using the standard fixation parameters for low-speed (< 200 Hz) eye-tracking (as is common in this work, the first fixation was not included in the final analyses). Fixations were defined as any period of gaze that was at least 80 ms in duration, with maximum dispersion of 100 px. Using this definition, we filtered the stream of eye gaze positions for every participant to obtain the number of fixations the participants made during each trial, the duration of each fixation, and the X and Y coordinates of each fixation. These data points were used not only to create fixation density maps (see next paragraph), but also to characterize participants' fixations (e.g., duration, distance of eye movement between fixations).

Fixation Density Map.—We used the X and Y coordinates of each fixation to create fixation density maps for each image that a participant viewed. These fixation density maps are matrices representing the number of fixations made at each pixel location of the image (i.e., 1050×1680). They were created using the following steps: 1) generating a 1050×1680 matrix of zeros in MATLAB, 2) plotting the locations of each fixation in the matrix by adding 1 to the cells that corresponded to the fixation location, 3) applying a Gaussian low-pass filter with a circular boundary and a cutoff frequency of -6 dB to the fixation density map to account for eye tracker error (the effect of the filter is that locations near the fixated location will be weighted high, with the weights more reduced with greater distance from the fixated location). The specific parameters were based on scene research conducted with adults (see Henderson & Hayes, 2018) and the default settings of the MIT Saliency Benchmark code (https://github.com/cvzoya/saliency/blob/master/code_forMetrics/antonioGaussian.m). To simulate lower quality eye tracking characteristics of work with infants (e.g., Wass et al., 2013, 2014), we also applied the Gaussian filter using a cutoff frequency of -3 dB. This less conservative approach did not yield different results, so for the results reported here we adopted the default cutoff frequency parameters of -6 dB to allow more direct comparisons to the adult literature.

For each fixation density map generated using this procedure, we standardized the values using MATLAB's `mat2gray.m` function so that all cells in the matrix contained values ranging from 0 to 1. The end result is a matrix with higher values in cells corresponding to locations participants fixated more and lower values in cells corresponding to locations that participants fixated less. For example, in the fixation density maps presented in the bottom left quadrant of Figure 4 you can see that the majority of Participant A's fixations were

directed at the center of the image (i.e., more yellow toward the center), whereas Participant B's fixations were directed toward the crowd of people on the right side of the scene (i.e., more yellow toward the right side).

Map generation

Before evaluating participants' eye gaze, we created several maps to serve as references for interpreting patterns of looking (i.e., the infants' fixation density maps described in the previous sections). As described in the following paragraphs we created several maps based on infant fixations (leave-one-out fixation density maps), adult fixations (adult fixation density and adult center bias), and maps based on scene characteristics (image saliency and saliency center bias). The fixation based maps were generated using the parameters described earlier. In addition, we matched the density of different fixation density maps (for ease of comparison) using MATLAB's `imhistmatch` function.

Leave-One-Out Fixation Density Map.—We created fixation density maps for subsets of infants (based on age), leaving out the fixations from one infant in that subset. These leave-one-out maps will allow us to examine the consistency of infants' fixations (Le Meur et al., 2011), and will provide a theoretical upper limit for how well eye movements can be predicted (Chen & Zelinsky, 2019). Because we were particularly interested in how consistency changed with infant age, for each infant, we created an age-matched comparison group of the 13 infants who were closest in age to that infant. We then created a fixation density map for each of these comparison groups; these were the leave-one-out fixation density maps for the “left-out” infant. As a result, for each infant we created a leave-one-out density map for each image they viewed. We created leave-one-out density maps for subsets of similar aged infants, rather than creating such density maps with the group of infants as a whole, to allow us to evaluate age-related differences. We matched the density of each leave-one-out fixation density matrix to a reference matrix that contained all of the infants' fixations.

These leave-one-out density maps provided a way of comparing infants' fixations to the fixations of other infants of approximately the same age. In Figure 4, the fixation density maps for Participant A and Participant B are presented with their corresponding leave-one-out density maps. The leave-one-out map for Participant B demonstrates how when an individual fixates similar regions as their close-aged peers, a large positive Pearson correlation coefficient is obtained (.73 in the example). Notice that both Participant B and their peers (as indicated by the leave-one-out density map) focus in this scene on the crowd of faces present on the right side of the scene. In contrast, when an individual fixates unique regions compared to their peers (Participant A), a small or negative Pearson correlation coefficient is obtained (.00 in the example). Notice that the leave-one-out fixation density map for younger infants (Participant A's leave-one-out map) was more dispersed. Some infants in this age range looked at the crowd of people, and others spread their looks across the whole street scene.

Adult Fixation Density Map.—We created a single fixation density map for each image from all fixations produced by our adult sample. These adult fixation density maps provided

a standard for determining how “adult-like” infants fixation distributions are. We matched the density of the adult fixation density matrices to a reference matrix that contained all of the infants’ fixations.

Adult Center Bias.—We also created an adult center bias map by creating a single fixation density map that contained the X and Y coordinates of all of the fixations made by all of the adult participants across all of the images in the study. This map abstracted away image content and so reflected general spatial biases in viewing. We then matched the density of this map to a reference matrix that contained all of the infants’ fixations. The resulting center bias map was a 1050×1680 matrix with higher values in regions that adults’ typically fixated regardless of image content; it is presented in the bottom row of Figure 4, and it can be seen that across all scenes adults tended to fixate the center of images.

Image Saliency Maps.—For each stimulus image, we generated three saliency maps using three different models: 1) the Itti & Koch saliency model with Gaussian blur (IKB; 1998), 2) the Graph-Based Visual Saliency model (GBVS; Harel et al., 2006), and 3) the Attention By Information Maximization saliency model (AIM; Bruce & Tsotsos, 2007). We chose to use these saliency models for several reasons. First, they are commonly used in understanding infants’ fixation (Kadooka & Franchak, 2020; Kelly et al., 2019; Mahdi et al., 2018; Simpson et al., 2019). Thus, a systematic investigation of how well these models predict infants’ fixations will be useful for the interpretation of these other findings. Second, our goal was to understand the effect of low-level salience on infants’ fixation patterns. Although there are other approaches and other models of saliency, image saliency models let you isolate low-level features while deep saliency models contain a mix of low-, mid-, and high-level features which are difficult to pull apart and isolate. Thus, for these reasons in the present study we focused on these models of physical saliency. The IKB and the GBVS models use local differences in image features including color, edge orientation, and intensity to compute a saliency map (GBVS: Harel et al., 2006; IKB: Itti et al., 1998). The AIM model computes an image saliency map based on each scene region’s Shannon self-information (Bruce & Tsotsos, 2007).

IKB and GBVS saliency maps were generated using the Graph-Based Visual Saliency MATLAB toolbox with default IKB settings and default GBVS settings (Harel et al., 2006). The AIM saliency maps were generated using the AIM MATLAB toolbox with default settings and blur (Bruce & Tsotsos, 2007) (Figure 4).

Saliency Map Center Bias.—For each saliency model, we generated a single center bias map that reflected the unique scene independent spatial bias present in each saliency model. The model center bias maps were estimated using the same methods discussed in Hayes and Henderson (2020), but were resized to reflect the size of the images used in our study (1050×1680). We then matched the density of the saliency model center bias maps to a reference map that contained all of the infants’ fixations for a given image.

Results

We analyzed the data in several steps. First, to make sure that any age-related results we obtained were not the result of data quality differences, we examined the amount of track loss, the number of fixations, the duration of fixations, and saccade lengths. These metrics provide information about data quality, developmental changes in how information in the scenes is processed (e.g., if young infants are slower to process information, they will have longer individual fixations), control over attention and eye gaze (e.g., difficulty in disengagement may be reflected in the number and duration of fixations), as well as what information infants have access to in scenes (e.g., the number of fixations and length of saccades have implications for how much of the scene is examined).

Our primary analyses examined age-related changes in how well we can predict infants' gaze when viewing natural scenes. We asked how *consistent* infants' looking is -- that is, whether individuals' eye gaze tend to look at the same regions (and there is high consistency in fixation density maps) or whether individuals are idiosyncratic in how they look at scenes (i.e., each individual looks at different regions, yielding low consistency). We also asked whether infants' looking is predicted by adult gaze patterns.

To understand the factors that contribute to our observed consistency, we examined how eye gaze was influenced by physical saliency by comparing participants' patterns of fixations with saliency maps as generated by the three saliency models. Finally, we evaluated how infants' looking reflected the kinds of center biases that are reflected in these models, and that are observed in adults' fixations.

We evaluated age-related changes during infancy using simple linear regression to predict changes in our dependent variable as a function of age in days. To aid interpretation, we recoded age to be relative to the youngest infant participant in our sample (114 days) by subtracting 114 from the age in days for each infant. Therefore, the effect of age can be interpreted as the amount of change in the data per day beyond 4 months (approximately).

Data Quality and Fixation Characteristics

Our first analyses were aimed at determining whether there are significant age-related changes in the amount of data or the quality of the fixations. We analyzed the amount of data by evaluating track ratio, or the proportion of the trial that contained observable data. Track ratio may differ by age as a result of older infants moving more than younger infants, developmental changes in the pigment in the iris influencing the SMI system's ability to detect reflections in younger infants' eyes, or variations in pupil distance creating a problem for the SMI's binocular tracking system. Thus, track loss may reflect time-off-task or the inability of the eye tracker to maintain a stable track during periods of time when the infant is fixating the stimulus. Regardless of the source, because the amount of observed data may influence other measures of eye gaze, it is important to examine developmental differences in track loss.

We analyzed the average track ratio across all the trials that the participant contributed to the final analysis. In general track ratios were very high (Figure 5a); on average we recorded

gaze locations on over 80% of the eye tracking samples, $M = 83.97$, $SD = 8.38$, with only two infants demonstrating average track ratios less than 70%. Thus, not only were our stimuli effective at maintaining infants' attention, we also successfully tracked infants' eye gaze for most of the trial. A linear regression indicated that age explained 0% of the variance in track ratio, $R^2 < .01$, $F(1, 52) < 1.00$, $p = .995$, $\beta_{\text{age in days}} = -0.0001$. However, it should be noted that infants track ratios were on average lower than our baseline adult sample, $t(76) = 6.78$, $p < .001$, $d = 1.66$. All adult participants had track ratios greater than 90% ($M = 95.85$, $SD = 2.65$).

To evaluate fixation characteristics, we conducted separate analyses on the number of fixations, the average fixation duration, and the average saccade length. Collapsed across age, infants produced on average 8.13 fixations per image, $SD = 1.98$, with an average fixation durations of 341.91 ms, $SD = 111.72$, and average saccade length of 239.88 pixels, $SD = 49.49$ (approximately 48° , $SD = 13^\circ$, visual angle at a 60 cm viewing distance). Linear regressions on these fixation characteristics revealed that age did not account for a significant amount of variance in number of fixations, $R^2 = .01$, $F(1, 52) = 0.12$, $p = .73$, fixation durations, $R^2 = .01$, $F(1, 52) = 0.25$, $p = .63$, or saccade lengths, $R^2 < .01$, $F(1, 52) = 0.20$, $p = .66$. It should be noted that regressions conducted on each variable controlling for track quality did not yield different results. Thus, across the first year there does not appear to be a change in the characteristics of infants' fixations in naturalistic scenes (Figure 5). In addition, we conducted versions of our primary analyses controlling for these variables and found that our age effects did not reflect underlying differences in data quality.

Not surprisingly, the characteristics of adults' fixations were different from those of infants. Compared to infants, adults had more fixations ($M = 14.31$, $SD = 1.60$), $t(76) = 13.26$, $p < .001$, $d = 3.25$, marginally shorter fixation durations ($M = 301.46$, $SD = 41.99$), $t(76) = 1.72$, $p = .09$, $d = .42$, and longer saccade lengths ($M = 266.09$, $SD = 36.35$), $t(76) = 2.33$, $p = .02$, $d = .57$.

Infants' systematic eye movements toward natural scenes

Consistency of Eye Gaze.—Next we conducted a series of analyses that compared each participant's fixations to the fixations of the other participants within their age range. If participants' fixations fall on similar locations, each individual's fixations will be correlated with those of the others in their age range. However, if each participant exhibited an idiosyncratic pattern of viewing, their fixations would fall on different locations, and their fixations will not be correlated with others in their age range.

We evaluated eye gaze consistency by first correlating each individual participant's fixation density map with the corresponding leave-one-out density map (see *Data processing* section) of their close-aged peers for each image, and then calculating the average Pearson correlation coefficient across participants. The mean of the infants' average correlations (i.e., collapsed across images) revealed modest consistency, average correlation was $M = .42$, $SD = .10$. This is lower than is seen in adults. For example, our baseline adult sample had a mean correlation of $.66$ ($SD = .07$), which was significantly greater than the consistency seen in our infants, $t(76) = 10.55$, $p < .001$, $d = 2.59$.

We examined age-related changes in consistency by conducting a linear regression on infants' leave-one-out correlations with age as a predictor (see Figure 6). This regression revealed that age accounted for 39% of the variance, which was significant, $R^2 = .39$ $F(1, 52) = 33.06$, $p < .001$, $\beta_{\text{age in days}} = .0009$. Thus, overall older infants fixated more similar scene regions as their close-aged peers compared to younger infants; on average infants' leave-one-out Pearson correlation coefficients increased by approximately .03 per month. This suggests age-related changes in factors that drive infants to fixate similar regions within a scene. This means that the example illustrated in Figure 4 was generally true of the data as whole. Recall that on that image, older infants as a group were more likely to fixate the faces on the right side of the image, whereas younger infants' fixations were more dispersed. The remaining analyses are focused on determining what factors explain this increase in consistency.

Comparison to Adult Fixations.—To better understand the increase in consistency in eye movements across age we first asked whether infants fixated the scenes in the same way as do adults. In other words, are infants' fixation density maps like those of adults, and how does the correlation between infant and adult fixation density maps change with age? Using the fixation density map of our adult sample as a baseline, we asked whether infants' fixations were developing to become more “adultlike”, or whether with increasing age are infants' fixations better predicted by the pattern exhibited by the adults. We correlated individual participant's fixation density maps for each scene with a single adult fixation map that represented all of the adults' fixations directed toward that scene (see *Data processing* section). We then calculated the average Pearson correlation coefficient for each participant, collapsing across all the scenes. The mean average correlation for the group as a whole was .39 ($SD = .08$). Figure 7a shows the average correlations for each infant by age in days. It can be seen that the correlations between the infants fixation density map and the adult map increased over this 8 month range, suggesting that infants' fixations became more adult-like.

We tested this observation with a linear regression on correlation between infants' fixation density maps and the adult maps with age in days as our predictor. This regression accounted for 38% of the variance, $F(1, 52) = 31.91$, $p < .001$, $R^2 = .36$, $\beta_{\text{age in days}} = 0.0008$. Thus on average, starting at four months, the Pearson correlation coefficients between the infants' and adults' fixation maps increased by approximately .02 per month.

The leave-one-out correlations reported earlier, however, provide a context for understanding the relation between infants' fixation density maps and the adult maps. Given the natural variability in participants' eye movements, any model can only predict the patterns of infants' fixations as well as the group can predict one individual. Thus, the leave-one-out correlations provide a theoretical upper bound on other models' ability to predict infant fixations (Henderson & Hayes, 2018; Torralba et al., 2006). The increase in leave-one-out correlations across the eight month age range presented earlier shows that this upper bound increases as infants get older. Because older infants looked at the scenes in a more consistent pattern, their fixations should be easier to predict. Thus, we can ask how much the increase in the correlation between infant and adult fixation density maps reflects the increase in consistency among infant fixations

To address this question, we conducted a linear mixed effects (LME) model on the two correlations for each infant with map (leave-one-out versus adult map) and age as fixed effects, and subject as a random effect. The model revealed only a significant effect of age, $t(52) = 5.03$, $p < .001$, confirming the effect of the LME without the leave-one-out correlations. In fact, as can be seen in Figure 7b, the change over age for the relation between infant and adult fixation density maps and for the leave-one-out correlations were very similar, an observation that is consistent with the lack of an effect of map or interaction between age and map. This suggests that across the eight month age range the extent to which infant fixations are becoming more adultlike is similar to the extent to which infants' fixations are becoming increasingly consistent.

We also tested how adults' center bias predicted infants' fixations. Recall that although adults have a strong center bias (Clarke & Tatler, 2014; Tatler, 2007), there is some evidence that infants' center bias is weaker (van Renswoude, van den Berg, et al., 2019). The correlation between infants' fixation density map and the adult center bias map described earlier was on average .23 ($SD = .08$), clearly lower than the correlation between infants' fixations and the full adult fixation map. Thus, the correspondence between infants' looking pattern and adults looking pattern did not reflect an overall center bias. Linear regression on these correlations with age in days as a predictor did not account for a significant proportion of the variance, $R^2 = .02$, $F(1, 52) < 1$, indicating that the correlation between infants' fixation density maps and the adult center bias did not change with age.

Comparison to Saliency Models.—Next, we examined how well infants' eye gaze is explained by saliency models. Specifically, we correlated participants' fixation density maps for each scene with the corresponding IKB, GBVS, and AIM saliency maps (see *Data processing* section for details), then calculated a single score for each participant and saliency model by averaging the correlations (in the next section, we evaluated how well the center bias in each saliency map predicted infants' fixations).

The average correlations between each saliency model and fixation map for the infants (collapsed across age) and adults are presented in the top half of Table 1 (the correlations between the center bias in each map and infant fixations in the bottom half of Table 1 will be discussed in the next section). Several things are immediately obvious. First, the saliency models were better predictors of adult fixations than infant fixations; comparisons of the correlations between adults and infants revealed that the correlations for adults was higher for each of the three models. Second, for both infants and adults, the correlations were higher for the GBVS model than for the other two models. Finally, the correlations are generally lower than the leave-one-out or adult map correlations reported earlier.

We conducted a series of paired t-tests comparing the correlations for the different maps (to control for multiple comparisons, we use Bonferroni adjusted p value, $p < .0167$, as our criterion of significance). Collapsed across age, the GBVS model was better at predicting infants' fixations than IKB, $t(53) = 15.23$, $p < .001$, $d = 1.13$, or AIM, $t(53) = 15.40$, $p < .001$, $d = 1.79$. IKB was a better predictor of infants' fixations than was AIM, $t(53) = 7.73$, $p < .001$, $d = .69$. This ordering of model success mirrors what we observed with our baseline adult sample (see Table 1), as well as the MIT Saliency Benchmarks.

To examine how the prediction of the three saliency models changed with infant age, we conducted separate linear regressions for each saliency model (see Figure 8). The regression on the correlations with the IKB map accounted for 27% of the variance, $R^2 = .27$, $F(1, 52) = 19.71$, $p < .001$, $\beta_{\text{age in days}} = 0.0003$. The regression on the correlations with the GBVS map accounted for 18% of the variance, $R^2 = .18$, $F(1, 52) = 11.26$, $p = .002$, $\beta_{\text{age in days}} = 0.0003$. Finally, the regression on the correlations with the AIM map accounted for 26% of the variance, $R^2 = .26$, $F(1, 52) = 18.44$, $p < .001$, $\beta_{\text{age in days}} = 0.0003$. Thus, all of the models showed larger correlations with the infant fixation maps with increasing age.

We compared the performance of each saliency model to maximum performance (i.e., leave-one-out correlations) using LME as described for the analysis of the correlations with the adult map. The LME for each of the three saliency models revealed main effects of age, map (saliency map versus leave-one-out map), and significant interactions between age and map (see Table 2). The main effects of map reveal that none of the saliency models perfectly predicted the consistency in infants' looking; in each case, the leave-one-out correlations provided better prediction. This is clear in Figure 8d where there is a large difference between the leave-one-out correlations and the three saliency models

The significant interactions between map and age indicates that none of the saliency model predictions increased with age to the same extent as did the leave-one-out correlations. As can be seen in Figure 8d, the correlations between the fixation density maps and the saliency model are similar to the leave-one-out correlations for the youngest infants, but the leave-one-out correlations are greater than the correlations with the saliency maps for the oldest infants. Thus, although the prediction of the saliency models increased with infant age, saliency models actually predicted a smaller proportion of the predictable variation of infants' fixations with increasing age.

Why did the simple correlations (collapsed across age) show stronger correlations with the GBVS and infant fixations, and yet the regression analyses suggest that there are stronger changes with age for the correlations with the IKB model? At first this seems contradictory. However, inspection of Figure 8d reveals that the correlations between infant fixations and the GBVS map were higher than those between infant fixations and the IKB map at all ages, but the difference between the two sets of correlations was greatest at the youngest age. Moreover, because all the LMEs revealed main effects of age *and* interactions between age and map type, it is clear that regardless of the particular saliency model tested, with increased age infants' fixation patterns were more consistent with the models.

Comparison to Center Bias.—It is important to point out that the IKB and GBVS saliency models incorporate a strong center bias in their calculations of salience, and even the AIM saliency model contains some level of center bias. This is clear in Figure 4, which illustrates the center bias as included in each of the saliency models. Thus, it is possible that these models are good predictors of infants' fixations due to the high weights in the center of the maps rather than how the models calculate physical salience. Center bias has been widely observed in scene viewing in adults (Clarke & Tatler, 2014; Tatler, 2007), and in a recent paper by Hayes and Henderson (2020) scene-independent spatial biases, like center bias, explained significantly more variance in adults' fixation density than did full saliency

models. Much less is known about developmental changes in center bias. Both the analyses presented above and a recent study (van Renswoude, van den Berg, et al., 2019) suggest that infants have a weaker center bias than adults. Therefore, we next evaluated how well the center bias displayed by each of the saliency models predicted infant eye gaze.

For each image, we correlated each center bias map with the individual infant's fixation density maps for each scene, creating scores for each subject by averaging the Pearson correlation coefficients obtained for all the images the participant saw (see Table 1). What is immediately clear from the table is that the center biases did worse or about the same at predicting infants' fixations as did the maps generated by the adult fixations or by the full saliency models.

We tested this observation by comparing the correlations between the center bias maps from each saliency model with the full model counterparts (see Figure 10). Both the GBVS and IKB saliency models performed significantly better at predicting infant fixation compared to their center bias, $t(53) = 9.58, p < .001, d = 1.30$, and $t(53) = 3.19, p = .002, d = 0.43$, respectively. Thus, aspects other than the center bias contribute to the fit between these models and infant fixations. However, there was no significant difference between the saliency model and center bias model for AIM, $t(53) = -0.16, p = .878, d = 0.02$, indicating that for this model the center bias alone does as good a job at predicting infant gaze as does the full model. It is important to note that Hayes and Henderson (2020) found that the center bias models performed better than any of the saliency models with their adult sample. In our adult sample, in contrast, the IKB and AIM models and their center biases were not different in their prediction of adult fixations, but as was true for the infants' fixations the full GBVS model also predicted adults' fixations better than did the center bias in that model, $t(23) = 7.18, p < .001, d = 1.20$. Differences between the viewing tasks and the stimuli may contribute to the different finding. Nevertheless, these results suggest that the GBVS and IKB saliency models capture something about infants' eye gaze above and beyond general spatial biases.

Our linear regressions on the set of correlations between infant fixations and the center bias for each saliency model did not reveal an effect of age for any of the center biases. In summary, in contrast to the full saliency models, the prediction of these center biases of infants' fixations did not change with infant age.

Discussion

This study adds to the growing literature on what factors drive infant eye movements when viewing natural scenes. Specifically, we observed that between 4 and 12 months, infants' eye movements become increasingly more consistent and systematic, indicating that older infants look more at the same regions than do younger infants. In addition, across age infants' eye movements become increasingly adult-like, as evidenced by the correlations between infant and adult fixations. These results corroborate other work showing that infants' eye movements become more systematic with age (Helo et al., 2016; van Renswoude et al., 2016).

The primary evidence that infants become more consistent in their eye movements across the first year is our observation that the leave-one-out correlations were stronger for older infants than for younger infants. These leave-one-out correlations are commonly used in the literature to determine if subjects' fixation patterns are consistent (Judd et al., 2011; Torralba et al., 2006) or if there is relatively high inter-observer congruency (Le Meur et al., 2011; Rahman & Bruce, 2016). Higher leave-one out correlations indicate that older infants fixate similar regions within the scene as their same-aged peers, whereas younger infants' fixations are more idiosyncratic.

This can be seen in Figure 4. The oldest infants, who were around eleven to twelve months (illustrated by Participant B's leave-one-out fixation density map), focus in this scene on the crowd of faces present on the right side of the scene. In contrast, younger infants, who were around six to seven months (illustrated by Participant A's leave-one-out fixation density map), had more dispersed fixations. Some infants in this age range looked at the crowd of people, and others spread their looks across the whole street scene. One possibility is that this increase in consistency simply reflects increases in data quality with age. Wass and colleagues (Wass et al., 2013; Wass & Smith, 2014) demonstrated how eye tracking robustness can vary when looking across broad age ranges, and how differences in robustness can affect eye movement measurement. Specifically, factors such as participant mobility and iris pigmentation affect measures of fixation duration (Wass et al., 2013) and AOI statistics (Wass et al., 2014).

However, our findings of increased consistency likely do not reflect such factors. First, we did not find significant age differences in our measure of data quality. More importantly, in analyses conducted controlling for track ratio and number of fixations yielded the same results as those reported here. Thus, our observation that infants become more consistent with age is due to factors other than data quality, such as a common attentional strategy that directs them to look toward the same regions within the scene.

A second factor that has commonly been used to describe and understand developmental shifts in infants' eye movements is physical salience (Althaus & Mareschal, 2012; Frank et al., 2009; Kwon et al., 2016). In general, physical salience has been argued to have decreasing influence with age. Specifically, with age, factors other than salience (such as meaning or the presence of a face) are presumed to have a larger influence on infants' looking, decreasing the relative influence of salience. We found that saliency models actually become better at accounting for infants' eye gaze with increasing age, apparently conflicting with this previous research. Because physical salience and semantic content are often highly correlated in scenes (Hayes & Henderson, 2017), it is impossible to know whether this pattern reflects stronger control over eye gaze by physical salience, semantic content, or both. An important goal for future research is to explicitly test for the influence of both types of factors on developmental changes in infants' eye gaze.

But, the increase in the correspondence in salience and infants' fixation pattern with increasing age is best understood in the context of the change in overall *consistency* in infants' fixation. Specifically because the amount of consistency provides a noise ceiling (Chen & Zelinsky, 2019), the leave-one-out correlations provide an upper limit on how

well we can predict infants' fixations. Comparison of the consistency as measured by the leave-one-out correlations and the correlations with the saliency maps show that with age the consistency in eye movements increases to a greater degree than does the correlation between fixations and salience. That is, the leave-one out correlations provide a measure of how systematic groups of infants are, and therefore how well any model or approach can predict the group of infants' eye movements. Evaluating the correlations between the fixation maps and physical salience in the context of the leave-one out correlations revealed that saliency accounts for an increasingly smaller proportion of the consistency in infants' eye gaze. Thus, consistent with other findings that younger infants' eye gaze is more influenced by salience (Frank et al., 2009; Kwon et al., 2016), our results show that the effect of physical salience only contributes a small part in the increased consistency in fixations over the first year. Moreover, because infant eye movements become more adultlike within the first year, the increase in consistency reflects developing attentional strategies that are similar to adult eye movement control.

It is also important to remember that saliency models were designed to predict *adult* eye gaze. They were created taking into account the physical features and general spatial biases (e.g., center bias) that best match adults' preferences, rather than some unbiased coverage of the physical aspects of the scene. Thus, the increase in correspondence between infants' eye movements and saliency maps may reflect developmental changes in the visual system that correspond to these characteristics of saliency models. There are large developments in infants' color discrimination and visual acuity (Brown, 1990; Teller, 1998; Teller et al., 1986) within the first year. Therefore, infants likely are not as sensitive to the same visual features as are adults.

For example, consider the street scene presented in Figure 11. In the top row we show the saliency maps generated by the three saliency models we used here. There were differences, but all three models primarily chose the crowd of people on the right and the bright orange shirt of the boy in front as the most salient region of that scene. The bottom row of Figure 11 shows the saliency maps generated by the models when we used a version of this image manipulated to approximate infant vision; we reduced color saturation and lowered the contrast. It is clear that when this modified image is evaluated by the saliency models, the calculation of salience is shifted to include more central regions within the scene. We do not know how well this modified figure approximates how infants see the world, but this demonstration shows how what is identified as salient in a scene is dynamic and likely changes as a function of vision development. With increasing age and more adult-like visual abilities, infants will become more sensitive to the features used by the saliency models, and thus, their eye gaze will better match the predictions made by these models.

We found that among the three models tested, GBVS performed the best at explaining infant fixations — again similar to what has been observed with adults (Judd et al., 2012). Interestingly, this better fit did not simply reflect the strong center bias of GBVS. At least within the parameters of our eye tracking study (i.e., five second viewing time, infants ranging from 4– 12 months), the GBVS model's calculation of salience explained eye gaze above and beyond center bias. It is important to point out that such results tell us that the way GBVS models low-level physical features, such as color and orientation, are the best

match for how infants' fixate scenes. Thus, this model helps us to explain how such features contribute to infants' fixation patterns, and how that changes with age. These results do not tell us that GBVS is the best approach to predicting where infants look. Other models that include both bottom-up and top down factors (e.g., Bruckert et al., 2019; Damiano et al., 2019; Rahman & Bruce, 2016) likely will be better overall at predicting where infants' fixate. However, inclusion of GBVS will be useful if the goal is to understand how specific underlying factors drive infant eye gaze.

The discussion of our results thus far has focused on the effect of salience on fixations. However, we consider salience in the larger context of forces that influence visual attention. As discussed earlier, although the prediction of saliency models improves with age, the gap between saliency model performance and the theoretical maximum performance (i.e., leave-one-out correlations) increased across the first year. We therefore conclude that saliency is a good predictor of where the youngest infants in our sample look, but saliency alone cannot predict where older infants look even though the correlations between the saliency model and fixation density maps increased with age. Specifically, there was a great deal of variance left unexplained for the older infants (after accounting for the effect of salience). Therefore, factors other than salience must play a role in determining where infants look. This is consistent with other infant research showing that with increasing age other factors better predict eye gaze over salience, such as locations of faces (e.g., Franchak et al., 2016; Frank et al., 2009), and more informative regions of an object category (e.g., antlers on a deer; Althaus & Mareschal, 2012). Thus, the impact of salience on fixation does in fact increase across the first year, but as other more semantically relevant features have an increasing influence on fixation, the influence of salience is overshadowed.

The question that remains is what developmental processes account for the fact that infants' eye movements become increasingly consistent with age. Research has revealed increases in precision in saccades (van Renswoude et al., 2016) and bias to some physical properties (Amso et al., 2014). Other work has focused on top-down mechanisms, such as interest in faces (Franchak et al., 2016; Frank et al., 2009; Rider et al., 2018) and a "better comprehension of narrative content" (e.g., knowing who in the video is the main focus of that scene; Franchak et al., 2016). These proposals are consistent with the literature on adults' gaze during natural scene viewing demonstrating that adults' fixations are related the presence of social information (Birmingham et al., 2009) and local meaning (Henderson & Hayes, 2018), as well as recent findings that by 24 months scene context influences fixations (Helo et al., 2017). An important direction for future research is to understand how such factors contribute to infants' fixation and the increased consistency in fixation patterns with age.

Acknowledgments

This research and preparation of this manuscript were made possible by NIH grants R01EY022525 and R01EY030127 awarded to LMO and R01EY027792 awarded to JMH. KIP was supported on training grant EY015387. Research materials, statistical analysis, and data sets for all experiments are available from (<https://osf.io/5j4ht/>) and videos of subject sessions are available at Databrary (<https://nyu.databrary.org/volume/1131>). We thank the students and staff in the Infant Cognition Laboratory at the University of California, Davis, for their help with data collection.

References

- Althaus N, & Mareschal D (2012). Using saliency maps to separate competing processes in infant visual cognition. *Child Development*, 83(4), 1122–1128. 10.1111/j.1467-8624.2012.01766.x [PubMed: 22533474]
- 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(1), e85701. 10.1371/journal.pone.0085701 [PubMed: 24465653]
- Birmingham E, Bischof WF, & Kingstone A (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Research*, 49(24), 2992–3000. 10.1016/j.visres.2009.09.014 [PubMed: 19782100]
- Brown AM (1990). Development of visual sensitivity to light and color vision in human infants: A critical review. *Vision Research*, 30(8), 1159–1188. 10.1016/0042-6989(90)90173-I [PubMed: 2205977]
- Bruce N, & Tsotsos J (2007). Attention based on information maximization. *Journal of Vision*, 7(9), 950–950. 10.1167/7.9.950
- Bruckert A, Lam YH, Christie M, & Meur OL (2019). Deep Learning For Inter-Observer Congruency Prediction. *IEEE International Conference on Image Processing (ICIP)*, 3766–3770. 10.1109/ICIP.2019.8803596
- Bylinskii Z, Judd T, Oliva A, Torralba A, & Durand F (2019). What Do Different Evaluation Metrics Tell Us About Saliency Models? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(3), 740–757. 10.1109/TPAMI.2018.2815601 [PubMed: 29993800]
- Canfield RL, & Kirkham NZ (2001). Infant cortical development and the prospective control of saccadic eye movements. *Infancy*, 2, 197–211. 10.1207/S15327078IN0202_5
- Castelhano MS, Mack ML, & Henderson JM (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3), 1–15. 10.1167/9.3.6
- Chen Y, & Zelinsky GJ (2019). Is there a shape to the attention spotlight? Computing saliency over proto-objects predicts fixations during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 45(1), 139–154. 10.1037/xhp0000593 [PubMed: 30596438]
- Clarke ADF, & Tatler BW (2014). Deriving an appropriate baseline for describing fixation behaviour. *Vision Research*, 102, 41–51. 10.1016/j.visres.2014.06.016 [PubMed: 25080387]
- Clifton C Jr, Ferreira F, Henderson JM, Inhoff AW, Liversedge SP, Reichle ED, & Schotter ER (2016). Eye movements in reading and information processing: Keith Rayner's 40 year legacy. *Journal of Memory and Language*, 86, 1–19. <https://www.sciencedirect.com/science/article/pii/S0749596X15000960>
- Colombo J (2001). The development of visual attention in infancy. *Annual Review of Psychology*, 52, 337–367.
- Damiano C, Wilder J, & Walther DB (2019). Mid-level feature contributions to category-specific gaze guidance. *Attention, Perception & Psychophysics*, 81(1), 35–46. 10.3758/s13414-018-1594-8
- Fantz RL (1964). Visual experience in infants: Decreased attention familiar patterns relative to novel ones. *Science*, 146(Whole3644), 668–670. [PubMed: 14191712]
- Franchak JM, Heeger DJ, Hasson U, & Adolph KE (2016). Free viewing gaze behavior in infants and adults. *Infancy*, 21(3), 262–287. 10.1111/infa.12119 [PubMed: 27134573]
- Frank MC, Vul E, & Johnson SP (2009). Development of infants' attention to faces during the first year. *Cognitive Psychology*, 110(2), 160–170. 10.1016/j.cognition.2008.11.010
- Frank MC, Vul E, & Saxe R (2012). Measuring the development of social attention using free-viewing. *Infancy*, 17(4), 355–375. [PubMed: 32693486]
- Gluckman M, & Johnson SP (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology*, 4(AUG), 1–7. 10.3389/fpsyg.2013.00527 [PubMed: 23382719]
- Harel J, Koch C, & Perona P (2006). A saliency implementation in matlab [URL:http://www.Klab.Caltech.Edu/harel/share/gbvs.Php](http://www.Klab.Caltech.Edu/harel/share/gbvs.Php).
- Hayes TR, & Henderson JM (2017). Scan patterns during real-world scene viewing predict individual differences in cognitive capacity. *Journal of Vision*, 17(5), 23. 10.1167/17.5.23

- Hayes TR, & Henderson JM (2020). Center bias outperforms image salience but not semantics in accounting for attention during scene viewing. *Attention, Perception & Psychophysics*, 82, 985–994. 10.3758/s13414-019-01849-7
- Hayhoe M, & Ballard D (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4), 188–194. 10.1016/j.tics.2005.02.009 [PubMed: 15808501]
- Helo A, Rämä P, Pannasch S, & Meary D (2016). Eye movement patterns and visual attention during scene viewing in 3- to 12-month-olds. *Visual Neuroscience*, 33, E014. 10.1017/S0952523816000110 [PubMed: 28359348]
- Helo A, van Ommen S, Pannasch S, Danteny-Dordoigne L, & Rämä P (2017). Influence of semantic consistency and perceptual features on visual attention during scene viewing in toddlers. *Infant Behavior & Development*, 49, 248–266. 10.1016/J.INFBEH.2017.09.008 [PubMed: 29028583]
- Henderson JM (2007). Regarding scenes. *Current Directions in Psychological Science*, 16(4), 219–222. 10.1111/j.1467-8721.2007.00507.x
- Henderson JM (2017). Gaze control as prediction. *Trends in Cognitive Sciences*, 21(1), 15–23. 10.1016/j.tics.2016.11.003 [PubMed: 27931846]
- Henderson JM, & Hayes TR (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, 1(10), 743–747. 10.1038/s41562-017-0208-0
- Henderson JM, & Hayes TR (2018). Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *Journal of Vision*, 18(6), 10. 10.1167/18.6.10
- Henderson JM, & Hollingworth A (1999). High-level scene perception. *Annual Review of Psychology*, 50, 243–271. 10.1146/annurev.psych.50.1.243
- Itti L (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, 12(6), 1093–1123. 10.1080/13506280444000661
- Itti L, Koch C, & Niebur E (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 1254–1259. <https://www.computer.org/csdl/trans/tp/1998/11/i1254.pdf>
- Judd T, Durand F, & Torralba A (2011). Fixations on low-resolution images. *Journal of Vision*, 11(4), 1–20. 10.1167/11.4.14
- Judd T, Durand F, & Torralba A (2012). A Benchmark of Computational Models of Saliency to Predict Human Fixations. *Mit-Csail-Tr-2012*, 1, 1–7. <https://doi.org/1721.1/68590>
- Kadooka K, & Franchak JM (2020). Developmental changes in infants' and children's attention to faces and salient regions vary across and within video stimuli. *Developmental Psychology*, 56(11), 2065–2079. 10.1037/dev0001073 [PubMed: 32915049]
- Kelly DJ, Duarte S, Meary D, Bindemann M, & Pascalis O (2019). Infants rapidly detect human faces in complex naturalistic visual scenes. *Developmental Science*, e12829. 10.1111/desc.12829 [PubMed: 30896078]
- Koehler K, Guo F, Zhang S, & Eckstein MP (2014). What do saliency models predict? *Journal of Vision*, 14(3), 1–27. 10.1167/14.3.14
- Kwon M-K, Setoodehnia M, Baek J, Luck SJ, & Oakes LM (2016). The development of visual search in infancy: Attention to faces versus salience. *Developmental Psychology*, 52(4), 537–555. 10.1037/dev0000080 [PubMed: 26866728]
- Land M, Mennie N, & Rusted J (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11), 1311–1328. 10.1068/p2935 [PubMed: 10755142]
- Le Meur O, Baccino T, & Roumy A (2011). Prediction of the inter-observer visual congruency (IOVC) and application to image ranking Proceedings of the 19th ACM International Conference on Multimedia, 373–382. 10.1145/2072298.2072347
- Le Meur O, Coutrot A, Liu Z, Rama P, Le Roch A, & Helo A (2017). Visual attention saccadic models learn to emulate gaze patterns from childhood to adulthood. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, 26(10), 4777–4789. 10.1109/TIP.2017.2722238 [PubMed: 28682255]
- Mahdi A, Su M, Schlesinger M, & Qin J (2018). A Comparison Study of Saliency Models for Fixation Prediction on Infants and Adults. *IEEE Transactions on Cognitive and Developmental Systems*, 10(3), 485–498. 10.1109/TCDS.2017.2696439

- Malcolm GL, & Henderson JM (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, 10(2), 4.1–11. 10.1167/10.2.4
- Oakes L, Henderson JM, Pomaranski K, & Hayes TR (2021, 3 11). Developmental changes in how infants view naturalistic scenes 10.17605/OSF.IO/5J4HT
- Oakes LM, & Ellis AE (2013). An Eye-Tracking Investigation of Developmental Changes in Infants' Exploration of Upright and Inverted Human Faces. *Infancy*, 18(1), 134–148. 10.1111/j.1532-7078.2011.00107.x [PubMed: 23525142]
- Rahman S, & Bruce NDB (2016). Factors underlying inter-observer agreement in gaze patterns: predictive modelling and analysis Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, 155–162. 10.1145/2857491.2857495
- Rayner K (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. 10.1080/17470210902816461
- Rider AT, Coutrot A, Pellicano E, Dakin SC, & Mareschal I (2018). Semantic content outweighs low-level saliency in determining children's and adults' fixation of movies. *Journal of Experimental Child Psychology*, 166, 293–309. 10.1016/j.jecp.2017.09.002 [PubMed: 28972928]
- Schlegelmilch K, & Wertz AE (2019). The Effects of Calibration Target, Screen Location, and Movement Type on Infant Eye-Tracking Data Quality. *Infancy*, 24(4), 636–662. 10.1111/inf.12294 [PubMed: 32677249]
- Simpson EA, Maylott SE, Leonard K, Lazo RJ, & Jakobsen KV (2019). Face detection in infants and adults: Effects of orientation and color. *Journal of Experimental Child Psychology*, 186, 17–32. 10.1016/j.jecp.2019.05.001 [PubMed: 31185365]
- Stoll J, Thrun M, Nuthmann A, & Einhäuser W (2015). Overt attention in natural scenes: objects dominate features. *Vision Research*, 107, 36–48. 10.1016/j.visres.2014.11.006 [PubMed: 25478891]
- Tatler BW (2007). The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14), 4.1–17. 10.1167/7.14.4
- Teller DY (1998). Spatial and temporal aspects of infant color vision. *Vision Research*, 38(21), 3275–3282. [PubMed: 9893838]
- Teller DY, McDonald MA, Preston K, Sebris SL, & Dobson V (1986). Assessment of visual acuity in infants and children: the acuity card procedure. *Developmental Medicine and Child Neurology*, 28(6), 779–789. 10.1111/j.1469-8749.1986.tb03932.x [PubMed: 3817317]
- Torralba A, Oliva A, Castelano MS, & Henderson JM (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, 113(4), 766–786. 10.1037/0033-295X.113.4.766 [PubMed: 17014302]
- van Renswoude DR, Johnson SP, Raijmakers MEJ, & Visser I (2016). Do infants have the horizontal bias? *Infant Behavior & Development*, 44, 38–48. 10.1016/j.infbeh.2016.05.005 [PubMed: 27281348]
- van Renswoude DR, van den Berg L, Raijmakers MEJ, & Visser I (2019). Infants' center bias in free viewing of real-world scenes. *Vision Research*, 154, 44–53. 10.1016/j.visres.2018.10.003 [PubMed: 30385390]
- van Renswoude DR, Visser I, Raijmakers MEJ, Tsang T, & Johnson SP (2019). Real-world scene perception in infants: What factors guide attention allocation? *Infancy*, 24, 693–717. 10.1111/inf.12308 [PubMed: 32677279]
- von Hofsten C, & Rosander K (1997). Development of smooth pursuit tracking in young infants. *Vision Research*, 37(13), 1799–1810. [PubMed: 9274766]
- Wass SV, Forssman L, & Leppänen J (2014). Robustness and Precision: How Data Quality May Influence Key Dependent Variables in Infant Eye-Tracker Analyses. *Infancy*, 19(5), 427–460. 10.1111/inf.12055
- Wass SV, & Smith TJ (2014). Individual differences in infant oculomotor behavior during the viewing of complex naturalistic scenes. *Infancy*, 19(4), 352–384. 10.1111/inf.12049 [PubMed: 25635173]
- Wass SV, Smith TJ, & Johnson MH (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, 45, 229–250. 10.3758/s13428-012-0245-6 [PubMed: 22956360]

- Yarbus AL (1967). Eye Movements During Perception of Complex Objects. In Yarbus AL (Ed.), *Eye Movements and Vision* (pp. 171–211). Springer US. 10.1007/978-1-4899-5379-7_8
- Yun K, Peng Y, Samaras D, Zelinsky GJ, & Berg TL (2013). Exploring the role of gaze behavior and object detection in scene understanding. *Frontiers in Psychology*, 4, 917. 10.3389/fpsyg.2013.00917 [PubMed: 24367348]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

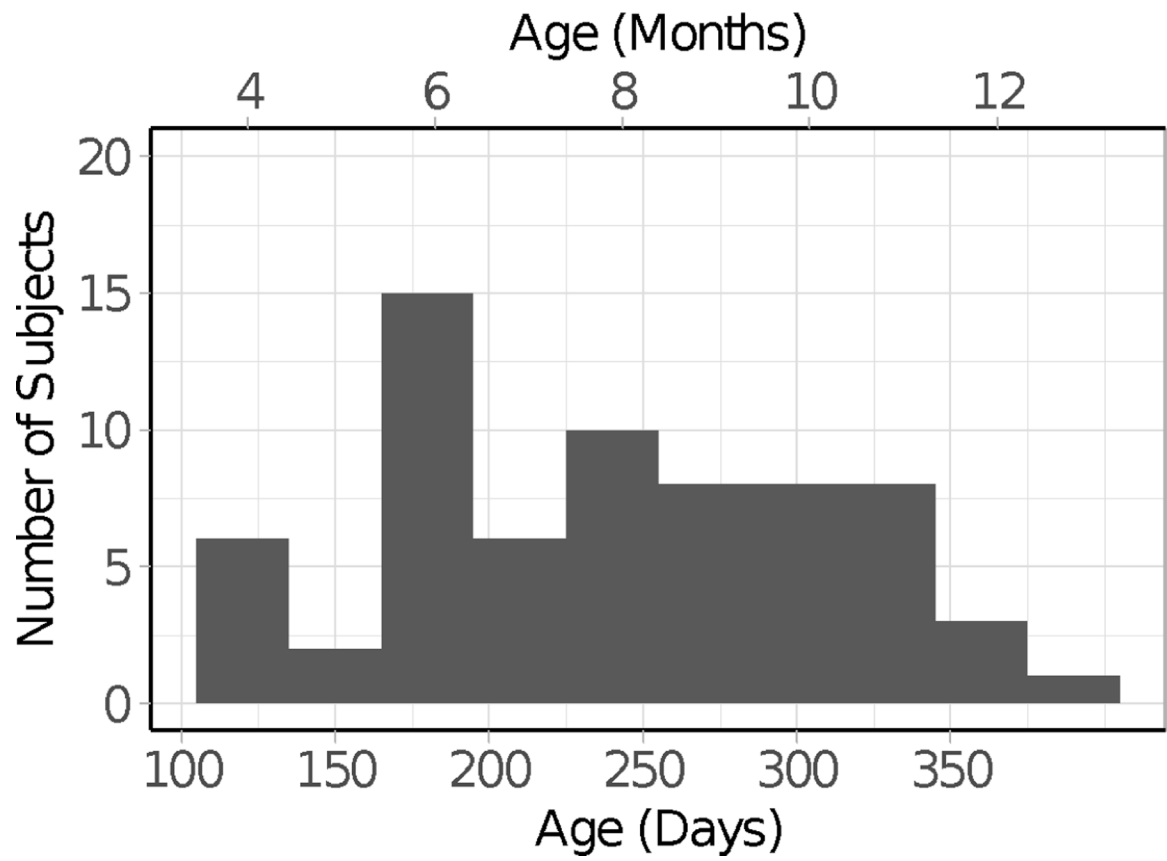


Figure 1.
Age distribution of the infants included in the final sample.

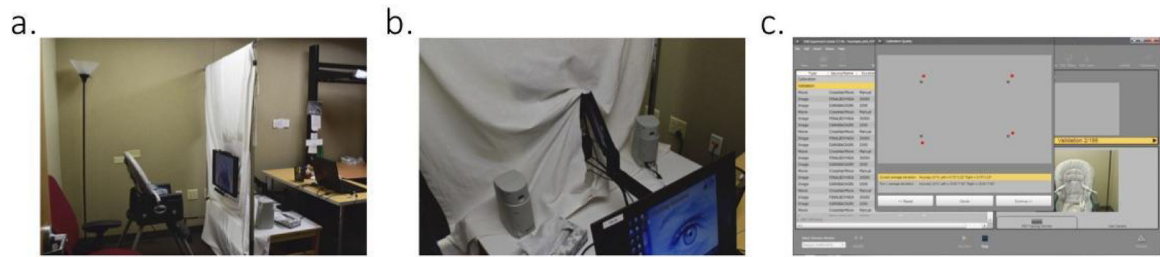


Figure 2.

(a) SMI-RED N set up. (b) Ergo arm used to optimally position the monitor for infant eye gaze. (c) Calibration quality feedback presented by SMI's Experiment Center software. The gray dots represent the location of the validation stimuli on the screen. The red dots represent the measured location of the infant's eye gaze during stimulus presentation. The illustrated feedback represents typical validation results seen with infant populations. Specifically, this infant had an average deviation of 0.75° in the horizontal and 1.32° in the vertical direction.

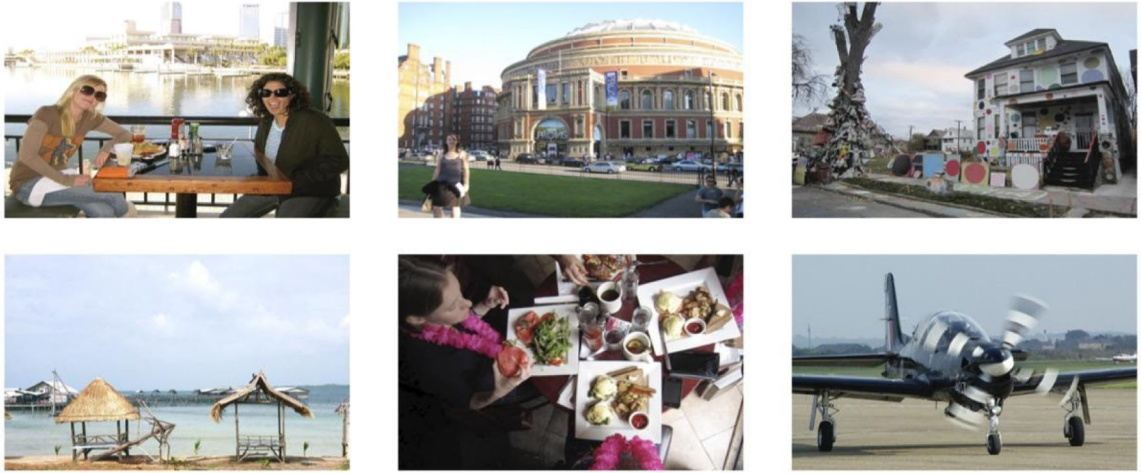


Figure 3. Example of scene images taken from the MIT300 Benchmark (Judd et al., 2012) and used as stimuli in the present study.

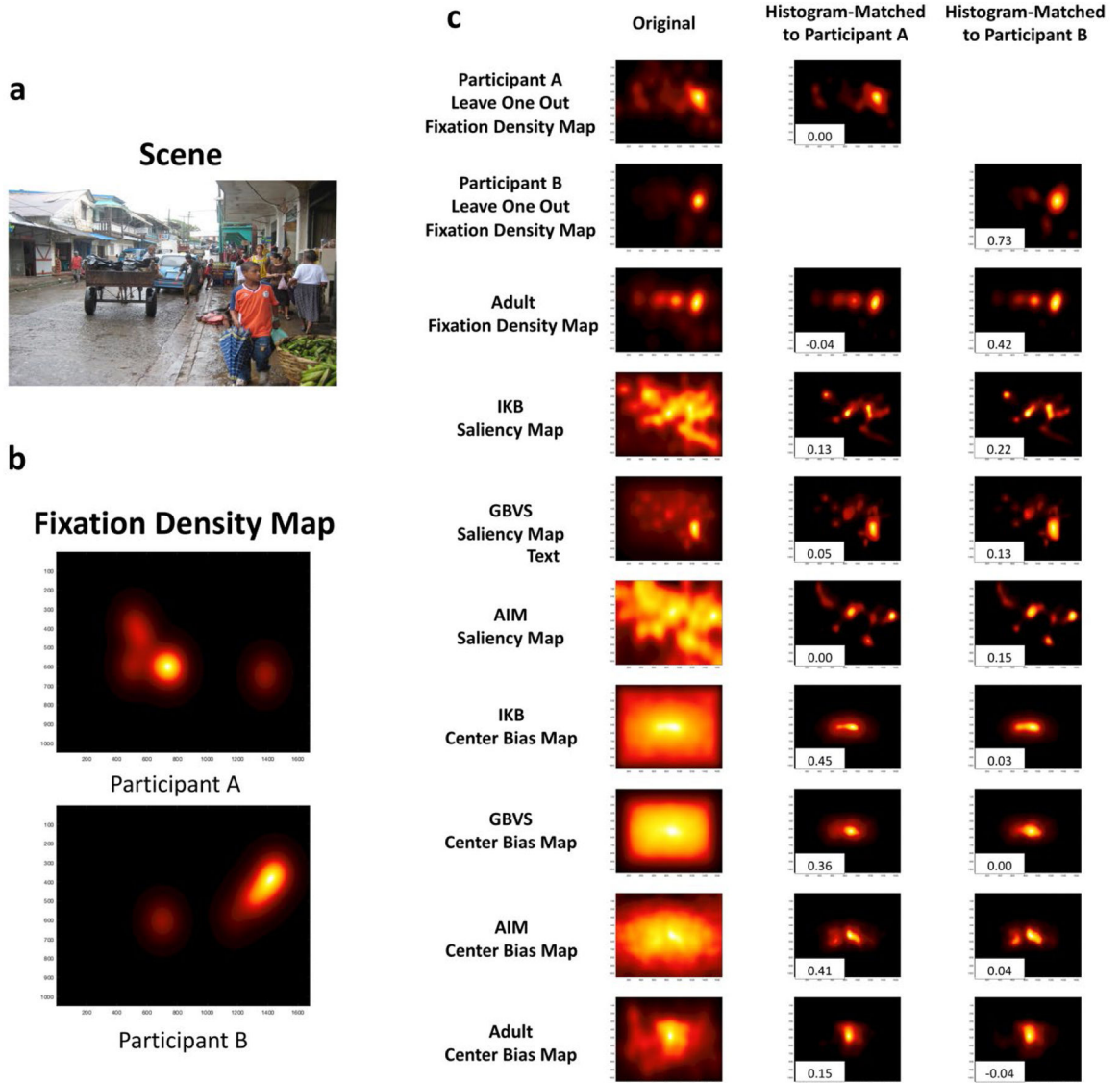


Figure 4. Example of fixation density maps to a specific scene and the corresponding maps generated. a. The scene the participants were viewing. b. The fixation density map for each of two participants (higher values are indicated by yellow; lower values are indicated by black); Participant A is a 6.5-month-old infant, with 8 fixations to the scene, and Participant B is a 11.5-month-old participant who directed 15 fixations to the scene. c. The maps that serve as reference for the infants’ fixation patterns (right column, from top to bottom: the participants’ leave-one-out fixation maps, adult fixations, saliency models, the center bias in the saliency models, and the adult center bias), and the histogram density-matched versions of those maps for each participant. The numbers in the bottom left of each of these maps are the Pearson’s linear correlation coefficient between the infant’s fixation density map and the reference map.

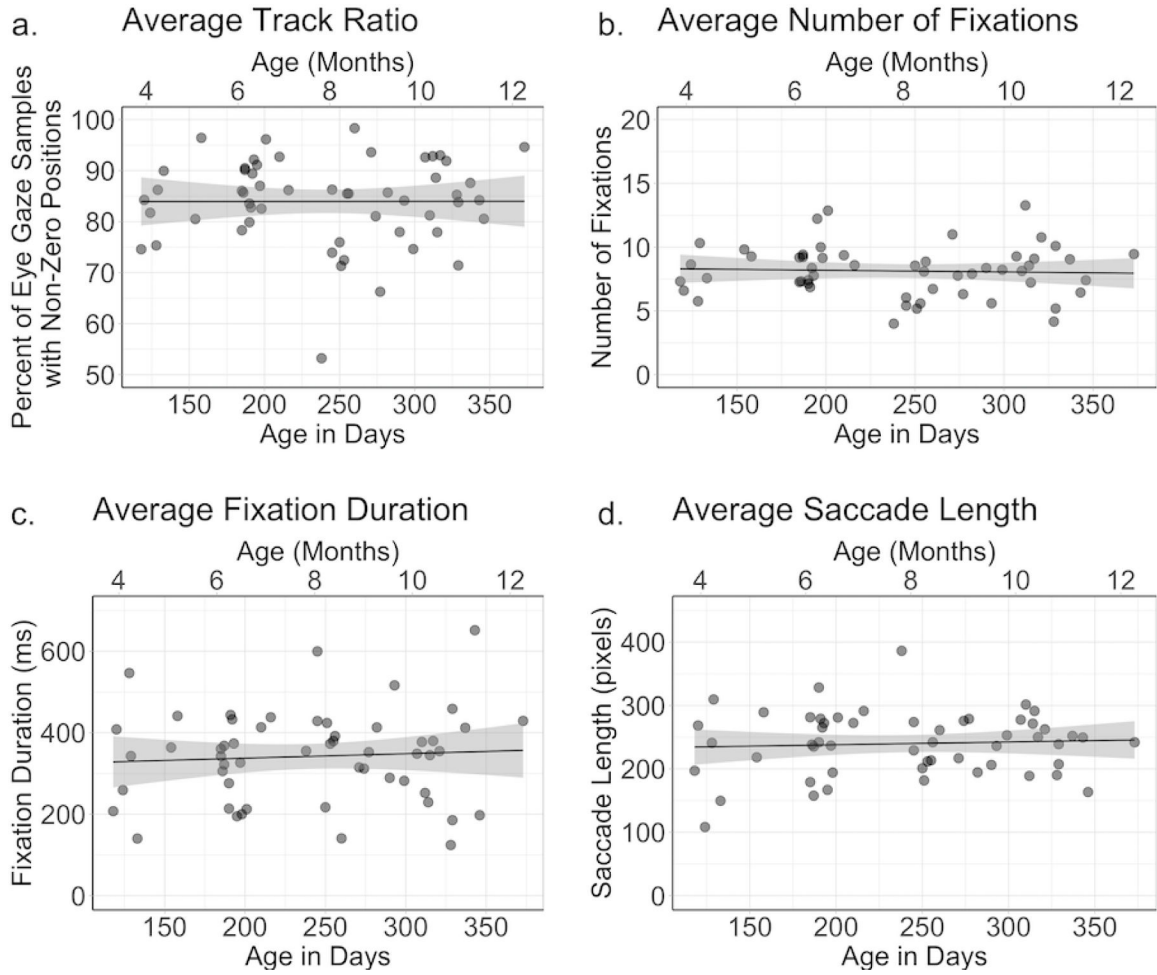


Figure 5. Measures of infant quality presented by infant age (age in months is along the top of each figure, and corresponding age in days is along the bottom of each figure). Each infant’s score (averaged across the scenes by the infant) on the four measures is indicated by a single dot in each graph. The four graphs depict the (a) average track ratio for each infant participant, (b) average number of fixations per scene generated by each infant participant, (c) average fixation duration produced by each infant participant, and (d) average saccade length produced by each infant participant.

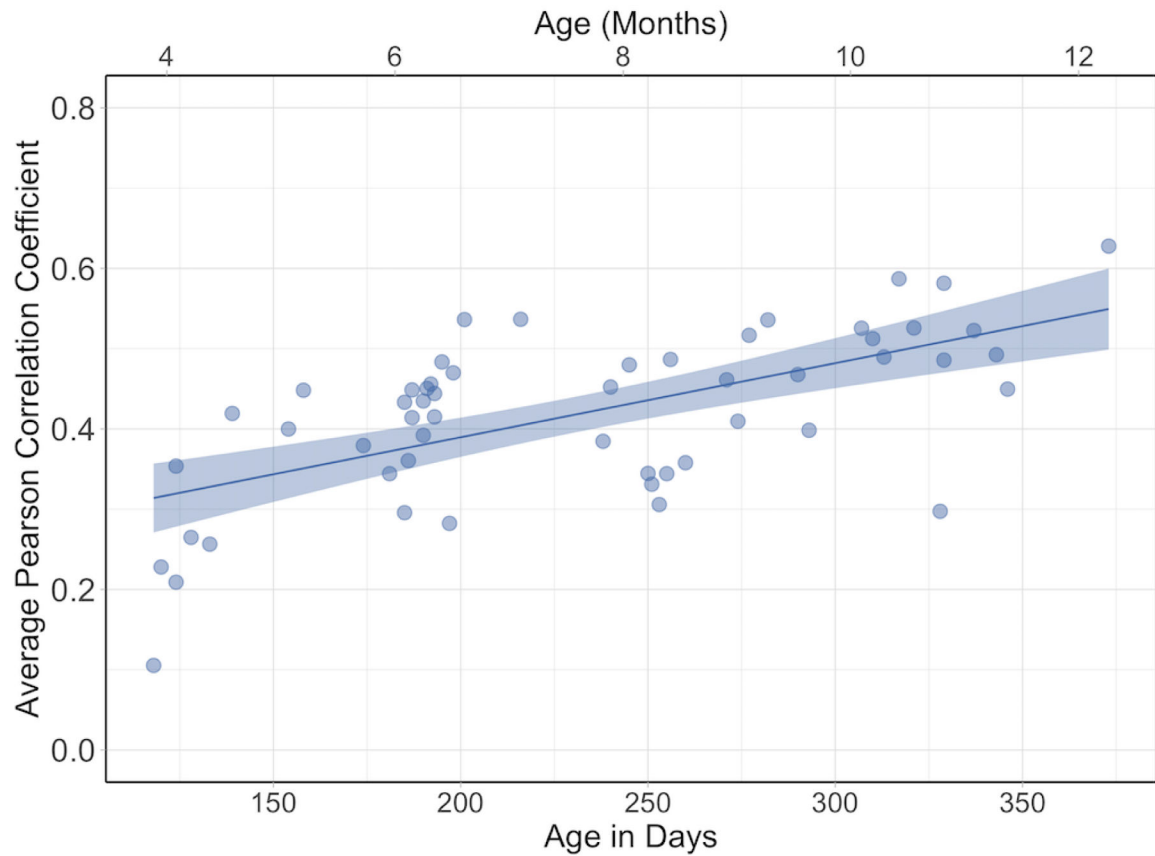


Figure 6.

Consistency of infants' eye gaze as measured by Leave-One-Out correlations. Each dot represents the correlation with an individual participant's fixation density map with the fixation density map of the 13 nearest aged infants (averaged across the scenes viewed by the infant); higher correlations represent more consistency between the infants. The line represents the effect of age across the eight month age span, and the shaded area reflects standard error.

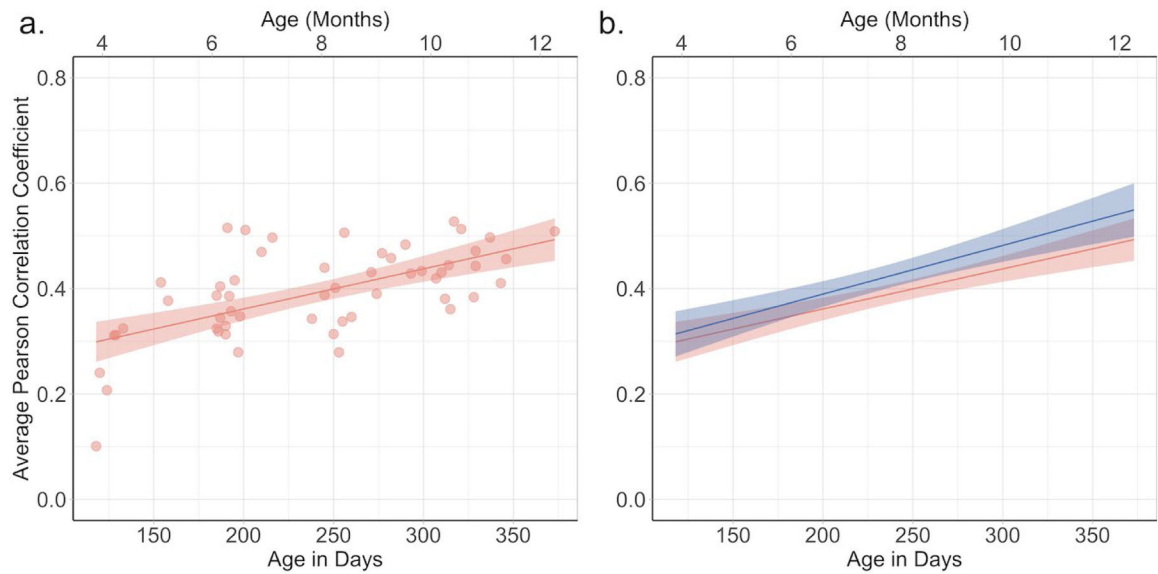


Figure 7.

The correlation between infants' fixation density maps and the adult fixation density map (a) and the comparison of the effect of age for the adult fixation density map (red) and the leave-one-out fixation density map (blue) (b). Each dot in graph (a) represents the average of an individual infant's correlation with the adult map for each of the images viewed by the infant. The values on the x-axis are age in days (age in month are indicated on the top of each graph). In each graph the line represents the effect of age across the eight month age span and the shaded area reflects standard error.

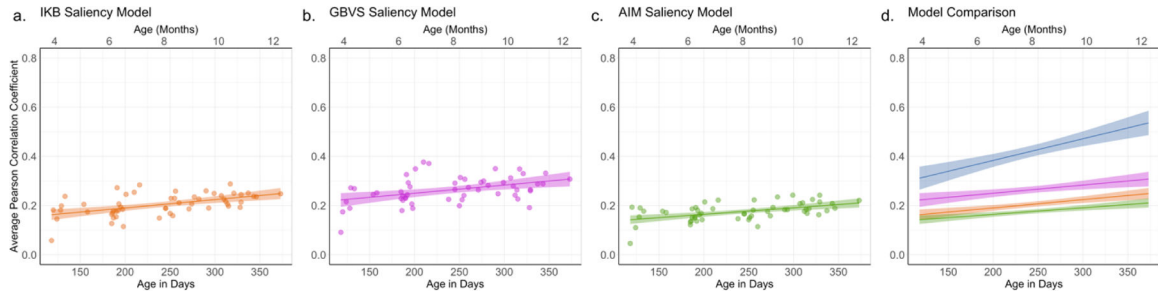


Figure 8.

Average correlation between infants' fixation density maps and the (a) IKB saliency model, (b) GBVS saliency model, and (c) AIM saliency model. Each dot represents the correlation between an individual infant's fixation density map and the map generated by the saliency model (averaged across the scenes viewed by the infant). The x-axis represents age in days (age in months is indicated at the top of each graph), and the line represents the effect of age across the eight month age span. The shaded area reflects standard error. Panel (d) shows a direct comparison of effect of age on the three saliency models and the theoretical maximum correlation as indicated by the leave-one-out fixation density map correlations (blue).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

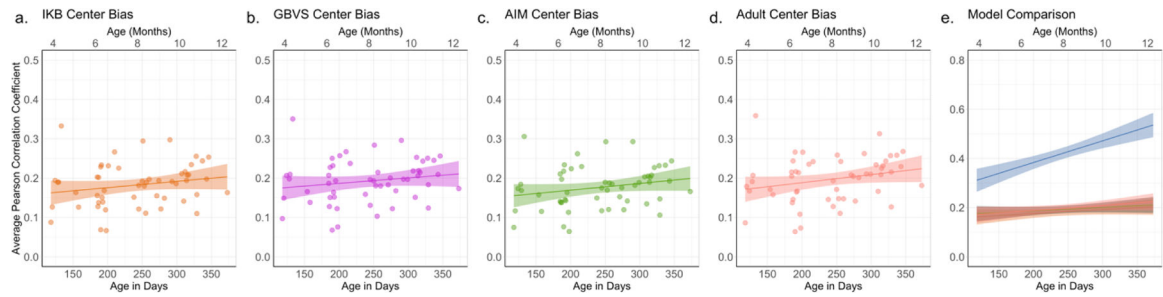


Figure 9:

Average correlation between infants' fixation density maps and the (a) IKB center bias, (b) GBVS center bias, (c) AIM center bias, and (d) adult center bias. Each dot represents the correlation for each infant's fixation density map and the saliency model (averaged across the scenes viewed by the infant). The line represents the effect of age across the eight month age span. The shaded area reflects standard error. Panel (e) shows a comparison of the center bias models and the leave-one-out fixation density map (blue).

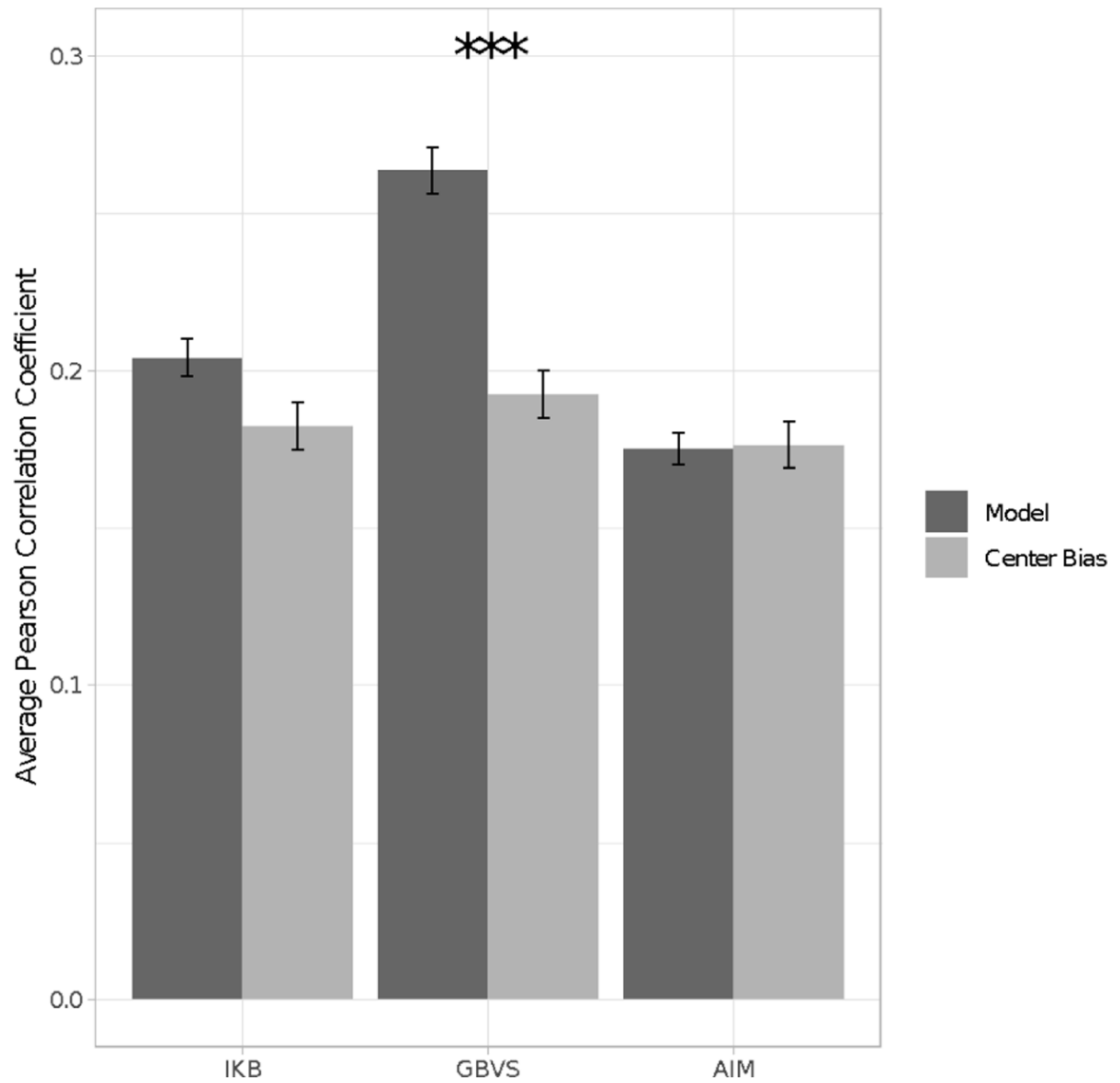


Figure 10: The average correlation between the fixation density maps and the saliency models and between the fixation density maps and the center bias for each saliency model averaged across infant age and scene. Error bars represent \pm standard error.

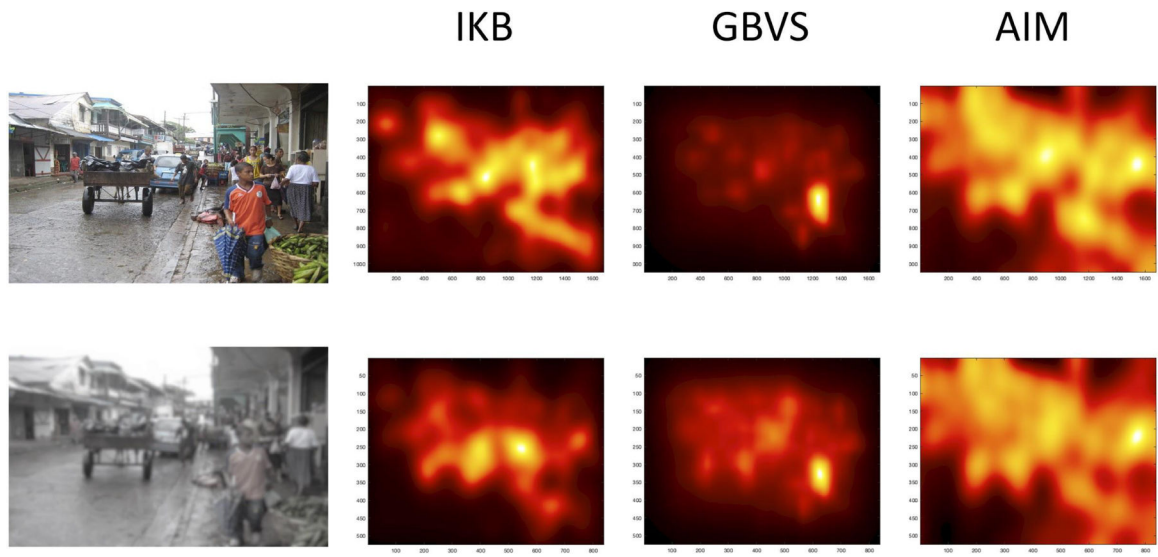


Figure 11.
Two versions of a scene and the saliency maps generated by each model for each version. The top row represents the scene as presumably seen by an adult. The bottom row approximates how the scene might be seen by infants, with reduced color saturation and lowered contrast.

Table 1.

Mean Pearson correlations (*SD*) between infant and adult fixation maps and maps generated by three saliency models and the center biases in the saliency models.

<i>Map</i>	Infant	Adult	t-test
IKB	.20 (.04)	.31 (.03)	$t(76) = 9.92, p < .001, d = 2.43$
GBVS	.26 (.05)	.37 (.04)	$t(76) = 8.07, p < .001, d = 1.98$
AIM	.18 (.04)	.25 (.03)	$t(76) = 9.31, p < .001, d = 2.28$
<i>Center Bias maps</i>			
IKB center bias	.18 (.06)	.28 (.06)	$t(76) = 7.07, p < .001, d = 1.73$
GBVS center bias	.19 (.06)	.29 (.07)	$t(76) = 6.70, p < .001, d = 1.64$
AIM center bias	.18 (.05)	.27 (.06)	$t(76) = 7.14, p < .001, d = 1.75$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Results of the LME analyses evaluating the correlations with fixation density maps and the salience maps as generated by three salience models comparing salience and leave-one-out correlations as a function of age.

Variables in the model	Effect of Age	Effect of Map	Age X Map interaction
Age, leave-one-out, IKB	$\beta = 0.0003, t(52) = 2.61, p = .01$	$\beta = 0.15, t(52) = 7.129, p < .001$	$\beta = 0.0005, t(52) = 3.66, p < .001$
Age, leave-one-out, GBVS	$\beta = 0.0003, t(52) = 2.43, p = .02$	$\beta = 0.09, t(52) = 4.77, p < .001$	$\beta = 0.0005, t(52) = 4.15, p < .001$
Age, leave-one-out, AIM	$\beta = 0.0003, t(52) = 2.13, p = .04$	$\beta = 0.17, t(52) = 8.47, p < .001$	$\beta = 0.0006, t(52) = 4.33, p < .001$