HHS Public Access

Author manuscript

*Infancy*. Author manuscript; available in PMC 2025 March 01.

Published in final edited form as:

*Infancy*. 2024 ; 29(2): 284–298. doi:10.1111/infa.12582.

# The role of local meaning in infants' fixations of natural scenes

**Lisa M. Oakes**[1,2,*], **Taylor R. Hayes**[2], **Shannon M. Klotz**[1,2], **Katherine I. Pomaranski**[1,2], **John M. Henderson**[1,2]

[1]Department of Psychology, University of California, Davis

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

## Abstract

As infants view visual scenes every day, they must shift their eye gaze and visual attention from location to location, sampling information to process and learn. Like adults, infants' gaze when viewing natural scenes (i.e., photographs of everyday scenes) is influenced by the physical features of the scene image and a general bias to look more centrally in a scene. However, it is unknown how infants' gaze while viewing such scenes is influenced by the semantic content of the scenes. Here, we tested the relative influence of *local meaning*, controlling for physical salience and center bias, on the eye gaze of 4- to 12-month-old infants ($N = 92$) as they viewed natural scenes. Overall, infants were more likely to fixate scene regions rated as higher in meaning, indicating that, like adults, the semantic content, or local meaning, of scenes influences where they look. More importantly, the effect of meaning on infant attention increased with age, providing the first evidence for an age-related increase in the impact of local meaning on infants' eye movements while viewing natural scenes.

The visual world presents infants with an overwhelming amount of information. To learn about the world, infants must select some information to focus on, shifting their attention from location to location to access new information. As they go about their daily routine, infants view both familiar scenes, like their own bedroom and kitchen, and novel scenes, like new parks or coffee shops, all that contain both new and older information. How do infants prioritize which information to attend to? What features of the world determine their shifts of attention as they view complex scenes? Decades of research have answered such questions about infants' attention to and perception of well-constrained experimental stimuli (Colombo, 2001; Johnson, 1990); much less is known about how attentional allocation of more complex, natural scenes develops over infancy.

A large literature has revealed the factors that influence how adults fixate natural scenes (Henderson, 2003). For decades, researchers have been exploring the factors that contribute to how adults distribute their attention when viewing natural scenes, such as photographs. When viewing such natural scenes, adult viewers tend to fixate physically salient regions (Itti & Koch, 2001) and the center of scenes (Hayes & Henderson, 2019a; Tatler, 2007). Similarly, studies examining infants' looking at naturalistic scenes have also revealed that where infants from Western Cultures and middle-class families look is influenced

*Corresponding author lmoakes@ucdavis.edu.

by physical salience (Pomaranski et al., 2021; van Renswoude, Visser, et al., 2019) and proximity to the center (van Renswoude, van den Berg, et al., 2019).

However, adults' fixations are also determined by the semantic content of scenes (Henderson, 2020). For example, adults' fixations are influenced by the presence of items that are incongruous with the scene, such as a printer in a kitchen scene (Vo & Henderson, 2009). Adult's fixation in such scenes is also influenced by *local meaning*, or the spatial distribution of meaning (Hayes & Henderson, 2019b; Henderson & Hayes, 2017). Scenes as a whole can be characterized in terms of their gist or semantic content. In addition, just as relatively small regions of scenes can be examined for physical salience (e.g., contrasts in brightness, color, orientation), they also can be evaluated for whether or not they convey any meaningful information (e.g., a table leg, part of a face, or a cup) (see Figure 1). Moreover, like physical features of images, this kind of meaning is unevenly distributed across scenes (Henderson et al., 2019). Consider the image in Figure 1a. Regions that contain the microwave, a cabinet door handle, or a refrigerator magnet are relatively high in meaning or information content. Regions that are in the middle of the curtain or a cabinet are low in meaning or information content.

To understand how the distribution of adults' fixations (i.e., *where* adults look) was related to the distribution of local meaning, Henderson and Hayes (Hayes & Henderson, 2019b; Henderson & Hayes, 2017) constructed *meaning maps* from ratings of the meaning or informativeness of regions of the scenes. To construct these maps, Henderson and Hayes presented adult raters with small patches of scenes and asked them to rate the informativeness of those patches (see Figure 1). As illustrated in Figure 1, patches judged as very high in meaning typically contained a recognizable object, and patches very low in meaning typically contained a relatively uniform patch of texture or color (e.g., a patch from the wall or window). Between these two extremes exists a rich continuum of patches of varying degrees of meaning (for an example of ratings produced by a typical rater see https://osf.io/yt2dk/). The adult judgments were then combined to generate maps that reflect the extent to which a given location looks like it contains something meaningful. In several studies, these maps better predicted adults' fixations during scene viewing than physical salience (see Henderson, 2020 for a review). In other words, where adults move their eyes from fixation to fixation is more strongly guided by where there is more meaning in a scene (controlling for salience) than where there is more salient information (controlling for meaning).

These meaning maps reflect *local meaning* and not the overall gist or conceptual representation of the scene. In fact, local meaning influences adults' fixations regardless of the context or the task, including when adults are engaged in free viewing of scenes (i.e., they are not given any explicit instructions or task) (Henderson, 2020; Peacock et al., 2019b). Moreover, the effectiveness of ratings of local meaning at predicting adults' fixation does not depend on understanding the scene as a whole. Peacock et al (2019) had adults' rate the informativeness of patches presented in isolation or of patches presented along with the scene context (Peacock et al., 2019b). Both sets of ratings similarly predicted a different group of adults' fixations in those scenes, indicating that the effectiveness of ratings of local meaning are not dependent on understanding scene gist. In addition to local meaning, adults'

viewing of scenes is influenced by gist and general conceptual information (e.g., if it is a bedroom or a beach) as well as their goals (e.g., looking for a blender versus a pair of shoes in a kitchen scene), although local meaning influences their fixations regardless of the context or the task (Henderson, 2020; Peacock et al., 2019a, 2019b, 2023). The point is that for adults, although fixations of natural scenes are strongly influenced by local meaning, their eye movements as they view natural scenes reflect their recognition of meaningful or informative regions of scenes at many levels.

The literature on the development of visual attention would lead to the prediction that infants' use of local meaning to guide their fixations would increase with age. A large body of research, primarily using highly controlled experimental stimuli, suggests a developmental trajectory from mainly bottom-up to increasingly top-down control of attention over the first postnatal year (Colombo, 2001; Johnson, 1990). A few studies using more complex scenes suggests a similar trajectory. Pomaranski et al. (2021) found that as infants from middle-class families in North America viewed natural scenes (e.g., photographs of kitchens, buildings), where they looked was less influenced by physical salience with increasing age (Pomaranski et al., 2021). In addition, Kiat et al. (2021) used a convolutional neural network (CNN) model of the ventral object processing pathway to examine how infants' eye movements during natural scene viewing are related to activity in different types of visual processing. In general, lower layers of the CNN (corresponding to lower visual areas) better predicted the eye gaze during natural scene viewing of younger infants and higher layers of the CNN (corresponding to higher visual areas) better predicted the eye gaze during natural scene viewing of older infants. Finally, between the first and second birthday, children's looking at natural scenes apparently is influenced by the overall scene gist and the presence of inconsistent objects, at least when those objects are salient (Duh & Wang, 2014; Helo et al., 2017). Thus, as has been argued for infants' looking to simple stimuli, fixation patterns within natural scenes appear to reflect more sophisticated and abstract representations in older infants relative to younger infants.

Interestingly, studies using complex stimuli with faces suggest a less clear developmental pattern. Using static images, Amso et al. (2014) and Kelly et al. (2019) found that infants detected faces in natural scenes regardless of whether or not those faces were the most salient region in the image. Moreover, they found little developmental change over the first year in infants' prioritization of social stimuli (faces) and physical salience. Other studies have examined infants' looking at faces versus physical salience in dynamic stimuli. Frank et al. (2009) found that when viewing an animated Peanuts video, younger infants tended to fixate the physically salient regions and infants 6 months and older tended to fixate regions that contained a face. Franchak and Kadooka (2022) found that when watching a video, 6- to 24-month-old infants' looked more at salient than non-salient faces. Thus, when scenes contain faces, infants' attention is complexly determined by a combination of physical salience and meaningful regions (i.e., where the faces are located). In part because of these findings, we used only scenes without people or faces in the present study.

Despite this interest in how high level features, and socially relevant features in particular, influence young children's visual attention when viewing complex stimuli, little is known about how the semantic properties of scene elements impact infant attention when viewing

natural scenes. One possibility is that semantic information is built up over years of experience with different environments and scenes and has relatively little influence on the control of attention during infancy. That is, despite evidence that high-level cortical structures are functional early in infancy (Cusack et al., 2016), the representations of objects in high-level visual cortex develop over a period of years (Deen et al., 2017). As a result, attention may be controlled by low-level stimulus properties throughout infancy, and their gaze may be relatively unrelated to the semantic content of natural scenes. The findings reviewed earlier are consistent with such a conclusion, with most reported findings showing an increasing influence of semantic features, such as the presence of a socially relevant element or the scene gist, across infant age.

Alternatively, because meaning maps reflect which *locations* are more or less meaningful, even infants' fixations may be related to local meaning. Specifically, infants acquire substantial expertise with the properties of natural objects as they explore their visual worlds and they may use this expertise to guide their attention to locations that appear to be higher in informativeness. That is, even if infants do not yet have labels for common objects and have only limited knowledge of how these objects function, high-level factors, including the kind of information reflected in meaning maps, may influence where infants look. As infants' visual attention increasingly is controlled by higher-level cortical areas, they may increasingly be able to shift their gaze to locations that they recognize as potentially more meaningful. The point is that local meaning may guide infants' fixations even if their viewing does not reflect all the same conceptual processes that contribute to adults' fixations.

The present study asked how one aspect of semantic content contributes to infants' attention as they view natural scenes. Specifically, we examined the effect of local meaning, as defined by Henderson and Hayes (2019; 2017) on infants' eye gaze during natural scene viewing. We asked how infants' fixations are predicted by the meaning maps generated by Henderson and Hayes (2019; 2017) from adult judgments of the informativeness of scene patches. To be clear, we do not expect that *meaning* is the same for infants as for adults. Adults' extensive experience with and additional knowledge about the elements in the scene certainly influences both their meaning ratings and their eye gaze. Because meaning maps represent *where* the meaningful elements of a scene are located, but not *what* each element means, infants' gaze patterns may be predicted by the meaning maps constructed from adult ratings of the scene patches even if they do not have an adult-like understanding of "meaning" in these regions. Local meaning may guide their visual attention, perhaps because they detect objects in the meaningful regions or some other property that coincides with adult ratings of meaningfulness. Of course, other aspects of meaning may also influence infants' attention in these scenes, but the focus of this study was to establish whether infants were more likely to look at scene regions rated as more meaningful by adults, regardless of whether or not they processed what was meaningful at those regions. On the basis of prior work showing that infants' eye gaze while viewing scenes becomes more adultlike over the first year (Pomaranski et al., 2021), we predicted that over the first year, infants' gaze patterns would become more strongly associated with meaning maps.

We tested this prediction by examining the eye gaze of infants between 4 and 12 months of age as they viewed the natural scenes used by Henderson and Hayes (2017). We adopted the approach used by Hayes and Henderson (2021) and compared the properties of fixated locations and randomly selected non-fixated locations. Specifically, we asked whether fixated locations differed from non-fixated locations in terms of local meaning, physical salience, and proximity to the center of the image. We also asked if the effects of these factors varied with infant age.

## Method

### Participants

Our final sample included 92 healthy, full-term infants between 4 and 12 months of age (121–379 days, *Mdn* = 244 days, see supplemental Figure 1 at https://osf.io/h2sq8/ for distribution; 40 boys and 52 girls), with no history of neurological or vision problems and who were not at risk for colorblindness based on family history. Infants were tested between 08/23/2017 and 08/26/2019.

Infant names were originally obtained from the State of California office of vital records, and parents were sent mailings about our research program. Parents who were interested in volunteering contacted us by phone, email, webform, or returning a postage paid card, providing us with their contact information. When infants reached the appropriate age for our study, we contacted the families and scheduled an appointment, if they were interested. Parents were not paid for their time, but infants were given a book or small toy and parents were given a certificate with their child's picture.

Originally, we used G*Power and established a sample size of 35 to 40 as sufficient to provide 80% power for the effects observed by Pomaranski et al. (2021). However, our anticipated analyses were more complicated than those reported by Pomaranski et al., so we identified a target sample size of 64 infants. We continued testing until we had enough data to ensure an adequate sample size, even if only 50% of the infants were included in our final analysis, as it is not uncommon in studies of this type for 40% to 50% of the infants to be excluded from the analyses (Pomaranski et al., 2021; van Renswoude, Visser, et al., 2019). Our multilevel modeling approach allowed us to include infants even if they completed only a small number of trials, yielding a final sample of 92 infants.

Infants were reported as White ($N$ = 59), African American or Black ($N$ = 2), Asian or Asian American ($N$ = 11), multiracial ($N$ = 17), or no race reported ($N$ = 3). Across these groups, 23 of the infants were reported to be Hispanic (15 White). Only one mother in our sample had less than a high school diploma, and 67 mothers had earned at least a 4-year degree. Infants received a small toy or t-shirt and a certificate in appreciation for their time.

An additional 30 infants were tested but not included in the final analyses because they did not provide useable eye tracking data (e.g., too fussy to participate, failure to calibrate, poor quality of data, equipment or experimenter error, $n$ = 24), parental interference (e.g., talking to the infant, $n$ = 2, ), or the infant was ineligible to participate (e.g., being premature or having a family history of colorblindness, $n$ = 4, ).

### Stimuli

The stimuli were 48 digitized color photographs of indoor and outdoor scenes (e.g., office space, living rooms, parks) from Henderson and Hayes (2017). Images were approximately 40 cm wide and 30 cm high (1400 × 1050 resolution, 37˚ wide by 28˚ high visual angle at a viewing distance of 60 cm) when presented on our 48 cm wide by 30 cm high (1680 × 1050 resolution) monitor, and varied in contrast, luminance, and colorfulness. None of the images used included humans or faces.

### Apparatus

Eye movements were measured using an SMI REDn eye tracker at a rate of 120 Hz, affixed to the bottom of a 22-in LCD monitor, which was fastened on an adjustable mount. A web camera attached to the top of the monitor recorded the infants' head and body movements. Stimuli were presented on the monitor using SMI's Experiment Center. A large white curtain hung behind the monitor to obscure the infant's view of the experimenter and equipment used to run the study.

### Procedure

The study was conducted following guidelines laid down in the Declaration of Helsinki, and approved by the Institutional Review Board of the University of California, Davis. Parents provided written informed consent before data collection began.

Infants were seated on their parents' lap or in a highchair, with the parent nearby. Parents were provided with felt-covered glasses to obstruct their view of the stimuli during the session and reduce bias. A video played on the monitor while the experimenter adjusted the monitor to obtain the clearest view of the infants' eye and the best track possible. The session began with a standard 5-point calibration procedure, in which an animated character first appeared in the top left corner and then moved to the other calibration points (top right, bottom left, bottom right, and center) (see Pomaranski et al., 2021). The calibration was immediately followed by a validation procedure, in which the experimenter received feedback about the quality of the calibration. If the calibration was poor (e.g., average systematic error > 1º horizontal, > 1.5º vertical, or at least one of four validation fixations appeared in an obvious outlier position), the procedure was repeated. In our final sample (excluding one infant whose validation data was corrupted and thus not recoverable), the average horizontal deviation from the intended target position was 0.94º, while the average vertical deviation was 0.88º.

Then, the experimental trials were presented. Before each trial, a fixation cross flashed in the center of the screen accompanied by attention-grabbing sounds (i.e., bells, rattle). We used the trigger AOI procedure that is a feature of Experimenter Center, the software provided by SMI to control the experiment and present stimuli. When the SMI system detected that 200 ms of gaze had accumulated within 5° of the fixation cross, the trial was initiated. During each trial, a single scene image was presented on the monitor for 5 s accompanied by classical music (to help infants maintain interest). During the trial, infants were free to look at the image however they wanted.

Trials were presented in blocks of four, and each block contained the same four stimuli (i.e., all infants saw the same four stimuli within trials 1–4). The stimuli chosen for each block were randomly selected from the overall set, and the order of the scenes within blocks randomized for each infant. We used the same set of stimuli in each block to increase the number of infants who saw some of the stimuli (i.e., virtually all infants saw the first block of stimuli, so those images were seen by the largest number of infants), which may be important for future analyses that might be conducted on this dataset. Short video clips were presented between blocks to maintain infants' interest and reduce the chance of fussiness (an example of trial sequences, with an infants' eye gaze superimposed, can be found at https://osf.io/h2sq8/).

## Data Processing

The data stream was filtered into fixations using standard parameters for low-speed eye tracking (< 200 Hz) in BeGaze, the software provided by SMI to process eye tracking data. Fixations were defined as any period of gaze within a dispersion of 100 pixels for at least 80 ms. Each infant's fixation on each trial was stored with the fixation index (i.e., which fixation it was in a trial), fixation duration in ms, and fixation X and Y coordinates. We used these data to calculate the total looking on each trial (i.e., the sum of the duration of all the fixations in the trial), the number of fixations on each trial, and the average duration of each fixation.

As is typical in the literature, we excluded the infants' first recorded fixations during each trial (i.e., to each image) from further analysis. We began with 2849 trials across our 92 included infants. The track ratio (i.e., the percentage of 8.33 ms time samples in a trial that produced non-zero gaze positions) for 429 of these trials was less than 25%; these trials were excluded from the final analyses. Our final analyses included 21053 fixations.

## Map generation

To derive values for meaning, physical salience, and center proximity for each fixated location and for corresponding non-fixated locations (see Analysis Approach section), we used several maps for each scene (e.g., Figure 2). We used the *Meaning Maps* generated by Henderson and Hayes (2017, 2018) as a representation of the spatial distribution of high-level semantic scene features. Henderson and Hayes created the meaning maps by decomposing each scene into a dense array of overlapping circular patches at a fine spatial scale (300 patches with a diameter of 87 pixels) and coarse spatial scale (108 patches with a diameter of 207 pixels) (Figure 1a). Adult raters then viewed random scene patches (without scene context) and indicated on a 6-point Likert scale how informative or recognizable each patch was. Patches that were uniform in color or texture were rated as low in meaning (Figure 1d, leftmost patches), and patches that had identifiable objects or parts of objects were rated as high in meaning (Figure 1d, rightmost patches). These ratings were used to generate a meaning map (Figure 2b) for each scene; this map provides a meaning value at each location (defined by x and y coordinates) of the scene.

We used the Graph-based visual saliency model (GBVS) with blur with default settings (Harel et al., 2007) to generate for each image an *Image Saliency Map*, reflecting low-level

image feature saliency (Figure 2c). This map reflects the predicted fixation density for each scene based on low-level, pre-semantic image features. Thus, this map allowed us to determine a saliency value at each location.

Finally, we used a *Center Proximity Map* to explicitly control the general bias for observers to look more centrally than peripherally in scenes, independent of the underlying scene content (Hayes & Henderson, 2019a; Tatler, 2007). This map represented the inverted Euclidean distance from the center pixel of the scene to all other pixels in the scene image (Figure 2d) and served as a global representation of how far each fixated location in the scene image was from the scene center. As with the other maps, this map provided a center proximity score at each location.

## Analysis approach

Our primary question was how the spatial distributions of meaning and physical salience contributed to where infants looked, or their overt attention. As in previous studies (van Renswoude et al. 2019; Pomaranski et al. 2021), our primary measure was the location of each fixation. Modeling the relationship between scene features and overt attention requires comparing the features of locations where each subject looked to the features of locations where they did not look in each scene (Nuthmann et al., 2017). Therefore, for each fixation a participant made in a specific scene, we computed the mean predictor value (i.e., meaning, saliency, center proximity) by taking the average over a 3° window around each fixation in each map for that scene. To represent scene features that were not associated with overt attention for that participant and that scene, we randomly sampled an equal number of scene locations where that individual participant did not look in that specific scene and computed the mean predictor values at those locations for comparisons.

The non-fixated locations were selected for each trial for each infant by randomly sampling scene locations where the infant did not look to the scene on that trial, with the constraint that the 3° window for each non-fixated location could not overlap with any of the 3° windows of the fixated locations (Hayes & Henderson, 2021). This reflects the logical constraint that for a given scene viewed by a given infant a scene region cannot be both fixated and not-fixated. This process yielded for each scene for each infant a set of fixated locations and a corresponding set of non-fixated locations.

The results of this process can be seen in Figure 2a. In this figure, the green dots represent where one of our infants fixated this scene and the blue dots represent non-fixated locations (one for each fixated location). Using the maps described earlier, we computed for each fixated and non-fixated location the mean meaning (Figure 2b), GBVS (Figure 2c), and center proximity (Figure 2c) value by taking the average over a 3° window around each location. This procedure provided the meaning, image salience, and center proximity values that were and were not associated with attention for each individual scene viewed by each individual infant.

We then applied a general linear mixed-effects (GLME) model to our data using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2019). All predictors were standardized to have mean 0 and standard deviation of 1. We used the *glmer* function with a binomial

distribution, logit link function, and the default optimizer (bobyqa and Nelder Mead). We chose a mixed-effects modeling approach because it does not require aggregating the eye movement data at the subject or scene-level; instead, both subject and scene are explicitly modeled as random effects. Additionally, the GLME approach allowed us to control for the role of center bias by including the distance from the screen center (Figure 2d) as both a fixed effect and as an interaction term with all the other model terms. We used a GLME logit model to investigate which factors were predictive of whether a scene region was attended or not.

Whether a region was fixated (1) or not fixated (0) served as the dependent variable, and meaning, GBVS image saliency, center proximity, and age were the main fixed effects. The model included several interactions. Because previous work had revealed that meaning and salience are correlated (Henderson et al., 2019) and that center proximity is a strong factor in scene viewing and interacts with other variables (Hayes & Henderson, 2019a; Tatler, 2007), we included in the model 2-way interactions between each of the maps (meaning x GBVS image saliency, meaning x center proximity, GBVS image saliency x center proximity) as fixed effects. Finally, to test our predictions that the effect of meaning would vary as a function of age, we included 2-way interactions between age and each of the maps (e.g., age x meaning, age x GBVS image salience, age x center proximity). We included scene and subject as full random effects. This model was used to understand how infant attention in scenes is associated with meaning, physical saliency, and center proximity, with a focus on how these associations differ as a function of infant age.

## Results

All de-identified data and data analysis scripts are available at https://osf.io/h2sq8/. To gauge infants' interest in the stimuli, we first examined general features of their looking behavior. Overall, infants were engaged in this task. They contributed an average of 29.04 trials ($SD$ = 12.31, range 8–48), looked an average of 2135.58 ms on each trial ($SD$ = 767.43, range = 506.28–3473.45), had on average 8.17 fixations per trial ($SD$ = 1.77, range = 4.53–12.09), and those fixations lasted on average 324.60 ms ($SD$ = 123.73, range = 127.11–771.17). Data quality was generally high; average track ratio was 67.40% ($SD$ = 11.98, range = 42.36–92.19). To determine whether infants' interest varied with age, we conducted pearsons correlations between infant age (in days) and measures of infant interest. Age was not significantly related to the duration of individual fixations, $r(90) = .06$, $p = .54$, or track ratio, $r(90) = .20$, $p = .06$, but was was significantly correlated with the number of trials completed, $r(90) = -.23$, $p = 0.03$, the average number of fixations per trial, $r(90) = .30$, $p = .004$, and total duration of looking at scenes, $r(90) = .25$, $p = .03$. Younger infants completed more trials than older infants; however, older infants attended more to the trials they were presented.

Our primary analyses were those that examined the factors that contributed to where infants looked. We evaluated the location of infants' fixations, and not the duration, for several reasons. First, research with adults has examined the effects of meaning and other variables on where adults look (Henderson et al., 2019). Second, previous studies on the factors that influence infants' looking at natural scenes have also evaluated the location of fixations (van

Renswoude et al. 2019; Pomaranski et al. 2021). Finally, although factors such as physical salience and center bias contribute to where infants' look, they seem to be not related to how long infants look (van Renswoude et al. 2019).

We tested our main question by fitting a model comparing fixated and non-fixated locations, as described in the Analysis Approach section. The results of the model are in https://osf.io/h2sq8/. The red horizontal line indicates non-significance (an estimated effect of $z = 0$); significant effects are either to the left ($z < 0$) or right ($z > 0$) of the red line.

The model revealed a significant effect of meaning (see Table 1; Figure 4a, red line), and a significant positive meaning by infant age interaction, illustrated visually using discrete groupings of age in Figure 4b. Overall, infants were more likely to fixate regions that were high in meaning than regions lower in meaning, but the interaction further reflects that this effect of meaning varied with infant age. As can be seen in Figure 4b, with increasing age, infants become both less likely to look at low meaning scene regions and more likely to look at high meaning regions. The three lines in that figure represent infants who are 2 SD above the mean in age (dark red line), infants who are at the mean age (intermediate red line), and infants who are 2 SD below the mean in age (pink line). The shaded bands illustrate the 95% confidence intervals for the estimated likelihood of a fixation. It can be seen that the confidence intervals for the oldest and youngest infants do not overlap at relatively high and relatively low meaning levels. Together, these results demonstrate for the first time that infant attention is associated with local scene meaning and that this relationship strengthens during infant development.

In addition, the model revealed the expected effects of GBVS image saliency and center proximity. These effects reflect that infants were more likely to look at more salient scene regions than less salient regions (Figure 4a, blue line) and were more likely to fixate central than peripheral scene regions (Figure 4a, green line). The image saliency effect was less than half the size of the meaning effect (Table 1 and Figure 4a), while the center proximity effect was comparable to the meaning effect size (Table 1 and Figure 4a). In addition, there was a significant GVBS by center proximity interaction, indicating a greater effect of salience farther from the center than locations near the center. However, we did not observe a significant GBVS by infant age interaction or a center proximity by infant age interaction. The lack of these interactions suggests that, when controlling for the effect of meaning on where infants looked, the associations between infant attention, image saliency, and center proximity remained relatively static during this developmental period from 4 to 12 months of age.

## Discussion

These data show for the first time that infants' eye gaze as they view natural scenes is influenced by local meaning (i.e., the extent to which a given patch is rated as meaningful by adults). Specifically, infants were more likely to fixate regions higher in meaning than regions lower in meaning. Our results are like those previously observed with adults, who also preferentially fixate meaningful regions (Henderson & Hayes, 2017). Thus, where infants look at natural scenes is related to the local meaning. In addition, although infants'

fixations were also related to saliency, the effect of meaning was much larger than the effect of saliency, indicating that meaning had more of an impact on where infants looked. This mirrors what has been observed for adults' eye gaze during natural scene viewing (Henderson et al., 2019).

Moreover, the influence of meaning on infants' fixations increased between 4 and 12 months of age. This is the general pattern predicted by previous research, in which infants' visual attention transitions from being controlled by stimulus factors to being controlled by higher-level processes (Colombo, 2001; Frank et al., 2009, 2014). Evidence indicates that very early in infancy, eye movements are controlled by the superior colliculus, with cortical control over eye movements increasing across the first year (Amso & Scerif, 2015; Colombo, 2001; Johnson, 1990). This understanding of the development of eye gaze control predicts that across development, high-level factors, such as semantic content, should have an increasingly greater influence on where infants look, as we observed here. The present findings are consistent with previous results suggesting that across the first year infants' gaze during natural scene viewing becomes increasingly adult like (Helo et al., 2016; Pomaranski et al., 2021). The results reported here also provide further support for the conclusion from the computational work by Kiat et al. (2021) that with increasing age infants' fixation of natural scenes is driven more by abstract, high-level properties of the scene.

Low-level factors also contributed to infants' eye gaze. We observed that infants were more likely to fixate highly salient regions and regions nearer to the center of the images, as has been found in other studies of infants' natural scene viewing (Pomaranski et al., 2021; van Renswoude, van den Berg, et al., 2019; van Renswoude, Visser, et al., 2019). In contrast to Pomaranski et al. (2021), however, we did not observe a significant interaction between age and physical salience. Specifically, not only did Pomaranski et al. observe that infants' fixation patterns became more adult-like, they also observed that over the first year salience accounted for less of the variance in infants' fixation patterns. In the sample reported here, the effect of salience did not appear to vary with infant age. However, physical salience and local meaning are correlated (Henderson et al., 2019). It is therefore possible that because Pomaranski et al. (2021) measured salience but not meaning, the increased effect of salience in the previous study actually was an increased effect of meaning on infants' looking. In the present analyses, in which the effect of salience was examined while controlling for meaning, only the increased effect of meaning with age was revealed.

It should be pointed out just because infants' gaze was predicted by local meaning as established by adult raters, this does not mean that the infants' understanding of that meaning was the same as adults. Indeed, the present results should not be taken as evidence that infants' *understanding* of the meaning in high meaning regions determines where they look. Given that the scenes used here were adult-focused, and sometimes contained content that was likely to be unfamiliar to infants (i.e., images of offices and laboratories), it seems likely that the scenes and meaning maps used here actually underestimate the effect of meaning on infants' gaze, and that meaning would have an even stronger influence over infants' eye gaze in scenes with more familiar content. However, because the adult-generated meaning maps did explain infants' eye gaze in the present context, there appears

to be some overlap between what adults judge as meaningful and what features draw infants' attention, presumably because they both reflect developmental timepoints in the same general system.

Finally, it must be acknowledged that as in many studies of infants' visual attention, we used static images rather than dynamic scenes. However, we know that infants' attention is different for static and dynamic stimuli. For example, infants attend for longer durations to dynamic stimuli (Shaddy & Colombo, 2004) and individual fixations are longer for dynamic compared to static stimuli (Wass & Smith, 2014). In a study of individual differences in infants' fixations, Wass and Smith (2014) found both similarities and differences in infants' fixations to dynamic and static stimuli. Importantly, studies using either static and dynamic stimuli have revealed a transition between 4 and 6 months in the influence of physical salience on where infants look. Franchak and Kadooka (2022) found age-related changes across infancy and early childhood in sensitivity to stimulus features when viewing dynamic stimuli. Thus, although it is likely that infants' eye movements differ when viewing static and dynamic stimuli, work with static images, like that presented here, adds to our overall understanding of the development of infants' visual attention, particularly in the context of the literature on adults' attention to such scenes. Nevertheless, an important goal for future research is to develop tools and procedures for addressing these questions with infants' viewing of dynamic stimuli.

In summary, we show for the first time that infants' fixations of natural scenes, like adults, are related to local meaning. As has been observed for adults, the spatial distribution of semantic features better predicts infants' fixations than does physical salience. Moreover, when controlling for the effects of physical salience and center proximity, the effect of meaning on infants' fixations increases with age. Together, these findings add to our understanding of the development of infants' visual attention when viewing natural scenes.

## Acknowledgments

We thank Steve Luck for helpful comments and the students and staff in the Infant Cognition Laboratory at the University of California, Davis, for their help with data collection. This research and preparation of this manuscript were made possible by NIH grants R01EY030127 awarded to LMO and R01EY027792 awarded to JMH. The authors declare no conflicts of interest with regard to the funding source for this study.

## Data availability statement:

All de-identified data and analysis scripts for this paper are openly available in the open science framework at DOI 10.17605/OSF.IO/H2SQ8. Videos of the gaze replays of the experimental sessions are available to authorized users in the Databrary repository at https://nyu.databrary.org/volume/1154.

## 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(1), e85701. [PubMed: 24465653]

Amso D, & Scerif G. (2015). The attentive brain: insights from developmental cognitive neuroscience. Nature Reviews. Neuroscience, 16(10), 606–619. [PubMed: 26383703]

Bates D, Mächler M, Bolker B, & Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, Articles, 67(1), 1–48.

Colombo J. (2001). The development of visual attention in infancy. Annual Review of Psychology, 52, 337–367.

Cusack R, Ball G, Smyser CD, & Dehaene-Lambertz G. (2016). A neural window on the emergence of cognition. Annals of the New York Academy of Sciences, 1369(1), 7–23. [PubMed: 27164193]

Deen B, Richardson H, Dilks DD, Takahashi A, Keil B, Wald LL, Kanwisher N, & Saxe R. (2017). Organization of high-level visual cortex in human infants. Nature Communications, 8, 13995.

Duh S, & Wang SH (2014). Infants detect changes in everyday scenes: The role of scene gist. Cognitive Psychology, 72(C), 142–161. [PubMed: 24751990]

Franchak JM, & Kadooka K. (2022). Age differences in orienting to faces in dynamic scenes depend on face centering, not visual saliency. Infancy, 27(6), 1032–1051. [PubMed: 35932474]

Frank MC, Amso D, & Johnson SP (2014). Visual search and attention to faces during early infancy. Journal of Experimental Child Psychology, 118, 13–26. [PubMed: 24211654]

Frank MC, Vul E, & Johnson SP (2009). Development of infants' attention to faces during the first year. Cognitive Psychology, 110(2), 160–170.

Harel J, Koch C, & Perona P. (2007). Graph-Based Visual Saliency. In Schölkopf B, Platt J, & Hofmann T. (Eds.), Advances in Neural Information Processing Systems 19 (NIPS 2006) (pp. 545–552). MIT Press.

Hayes TR, & Henderson JM (2019a). Center bias outperforms image salience but not semantics in accounting for attention during scene viewing. Attention, Perception & Psychophysics. 10.3758/s13414-019-01849-7

Hayes TR, & Henderson JM (2019b). Scene semantics involuntarily guide attention during visual search. Psychonomic Bulletin & Review, 26(5), 1683–1689.

Hayes TR, & Henderson JM (2021). Looking for Semantic Similarity: What a Vector-Space Model of Semantics Can Tell Us About Attention in Real-World Scenes. Psychological Science, 956797621994768.

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. [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.

Henderson JM (2003). Human gaze control during real-world scene perception. Trends in Cognitive Sciences, 7(11), 498–504. [PubMed: 14585447]

Henderson JM (2020). Meaning and attention in scenes. In Psychology of Learning and Motivation (pp. 95–117). 10.1016/bs.plm.2020.08.002

Henderson JM, & Hayes TR (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. Nature Human Behaviour, 1(10), 743–747.

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.

Henderson JM, Hayes TR, Peacock CE, & Rehrig G. (2019). Meaning and Attentional Guidance in Scenes: A Review of the Meaning Map Approach. Vision Research, 3(2), 19.

Itti L, & Koch C. (2001). Computational modelling of visual attention. Nature Reviews. Neuroscience, 2(3), 194–203. [PubMed: 11256080]

Johnson MH (1990). Cortical Maturation and the Development of Visual Attention in Early Infancy. Journal of Cognitive Neuroscience, 2(2), 81–95. [PubMed: 23972019]

Kelly DJ, Duarte S, Meary D, Bindemann M, & Pascalis O. (2019). Infants rapidly detect human faces in complex naturalistic visual scenes. Developmental Science, e12829.

Kiat JE, Luck SJ, Beckner AG, Hayes TR, Pomaranski KI, Henderson JM, & Oakes LM (2021). Linking patterns of infant eye movements to a neural network model of the ventral stream using representational similarity analysis. In Developmental Science. 10.1111/desc.13155

Nuthmann A, Einhäuser W, & Schütz I. (2017). How Well Can Saliency Models Predict Fixation Selection in Scenes Beyond Central Bias? A New Approach to Model Evaluation Using

Generalized Linear Mixed Models. Frontiers in Human Neuroscience, 11, 491. [PubMed: 29163092]

Peacock CE, Hayes TR, & Henderson JM (2019a). Meaning guides attention during scene viewing, even when it is irrelevant. Attention, Perception & Psychophysics, 81(1), 20–34.

Peacock CE, Hayes TR, & Henderson JM (2019b). The role of meaning in attentional guidance during free viewing of real-world scenes. Acta Psychologica, 198, 102889. [PubMed: 31302302]

Peacock CE, Singh P, Hayes TR, Rehrig G, & Henderson JM (2023). Searching for meaning: Local scene semantics guide attention during natural visual search in scenes. Quarterly Journal of Experimental Psychology, 76(3), 632–648.

Pomaranski KI, Hayes TR, Kwon M-K, Henderson JM, & Oakes LM (2021). Developmental changes in natural scene viewing in infancy. Developmental Psychology, 57(7), 1025–1041. [PubMed: 34435820]

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Shaddy DJ, & Colombo J. (2004). Developmental changes in infant attention to dynamic and static stimuli. Infancy, 5, 355–365.

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.

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. [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? In Infancy. 10.1111/infa.12308

Vo MLH, & Henderson JM (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. In Journal of Vision (Vol. 9, Issue 3, pp. 24–24). 10.1167/9.3.24

Wass SV, & Smith TJ (2014). Individual differences in infant oculomotor behavior during the viewing of complex naturalistic scenes. Infancy, 19(4), 352–384. [PubMed: 25635173]

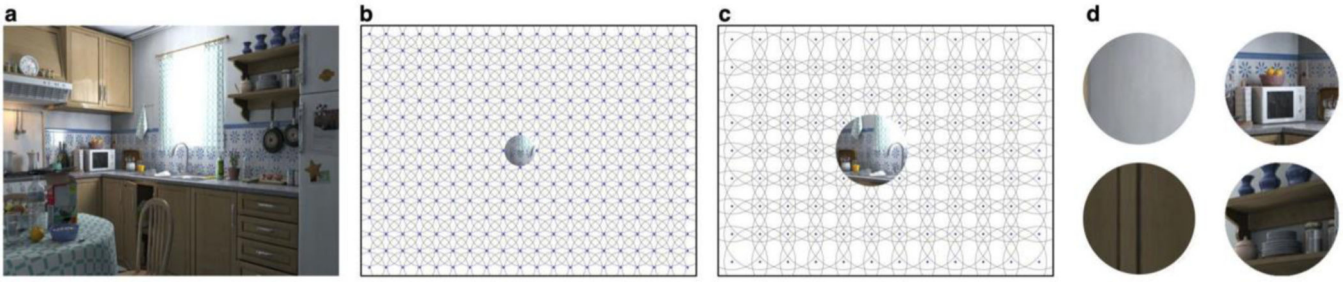


**Figure 1.**

An illustration of how patches from scenes are generated from a scene (a). The scene is divided into small (b) and medium (c) patch grids. Each scene was divided into 300 small patches with a diameter of 87 pixels to represent a fine spatial scale, and into 108 patches with a diameter of 207 pixels to represent a coarse spatial scale. The resulting patches (d) were rated by adults as having low meaning (left) or high meaning (right). The meaning ratings for each patch are then used to create a *meaning map* for the image as a whole, indicating which regions are higher and lower in local meaning. Figure adapted from Henderson, J. M., & Hayes, T. R. (2018). Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *Journal of Vision*, *18*(6), 10.

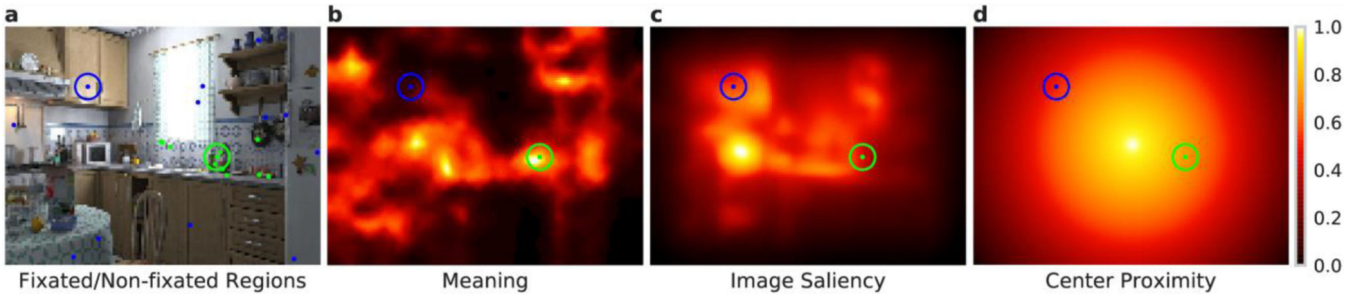


**Figure 2.**

An example scene with the fixated (green dots) regions for a single subject. The blue dots indicate randomly sampled non-fixated regions that represent where the infant did not look (a). Together, these locations provide an account of which scene regions did and did not capture this infant's attention. For each fixated and non-fixated location, a 3˚ window (the green and blue circles around the dots) was used to compute an average meaning (b), GBVS image saliency (c), and center proximity (d) map value.

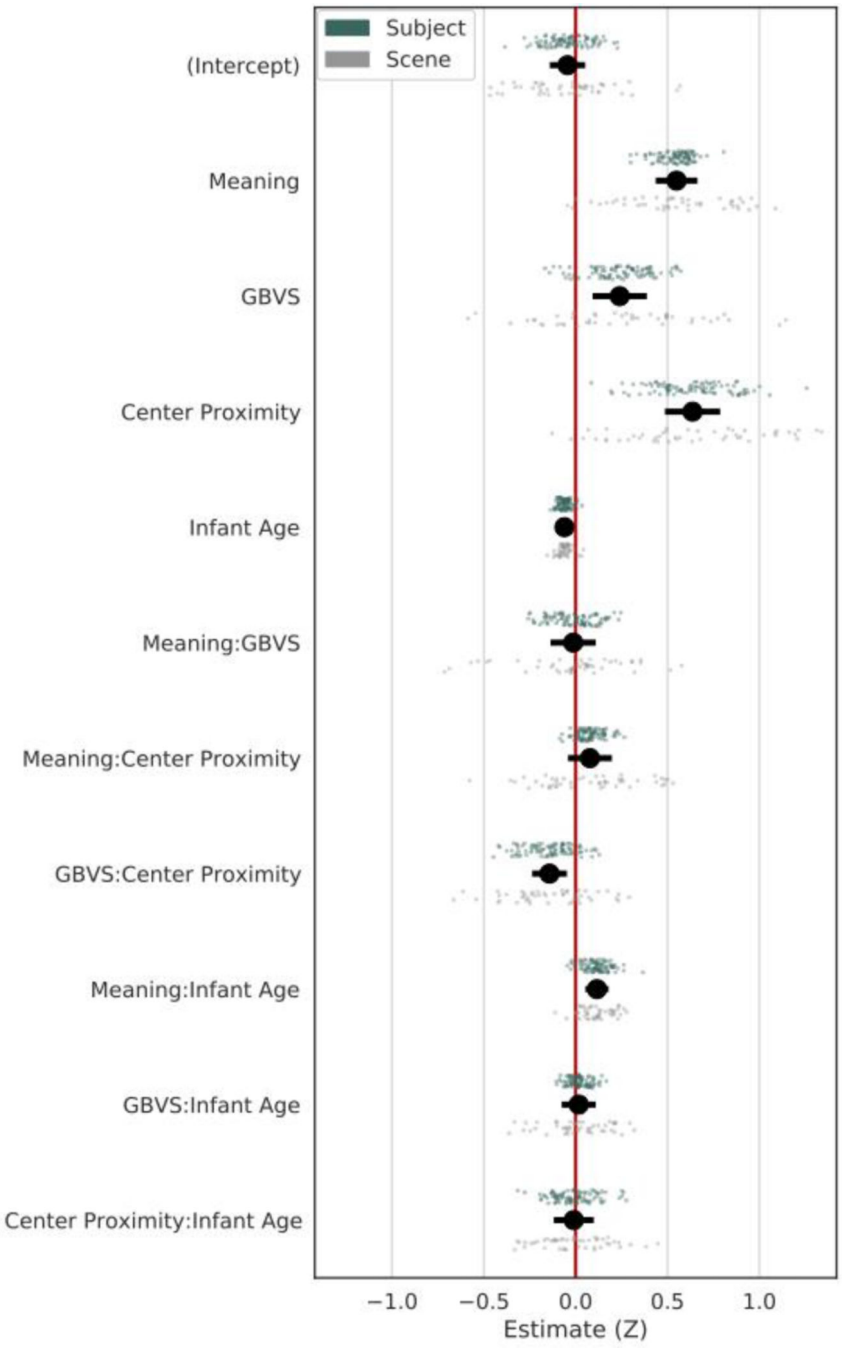


**Figure 3. Full general linear mixed-effects model results.**

Whether a scene region was fixated or not served as the dependent variable while the meaning, GBVS image salience, center proximity, infant age, and their interactions were included as fixed effects. The black dots with lines show the fixed effect estimates and their 95% confidence intervals. Subject (green dots) and scene (gray dots) were both accounted for in the model as random effects (intercept); negative z-scores indicate lower values, a zero z-score indicates average value, and positive z-scores indicate higher values of the predictor (See Table 1 for significance levels). Error bars represent 95% confidence intervals.

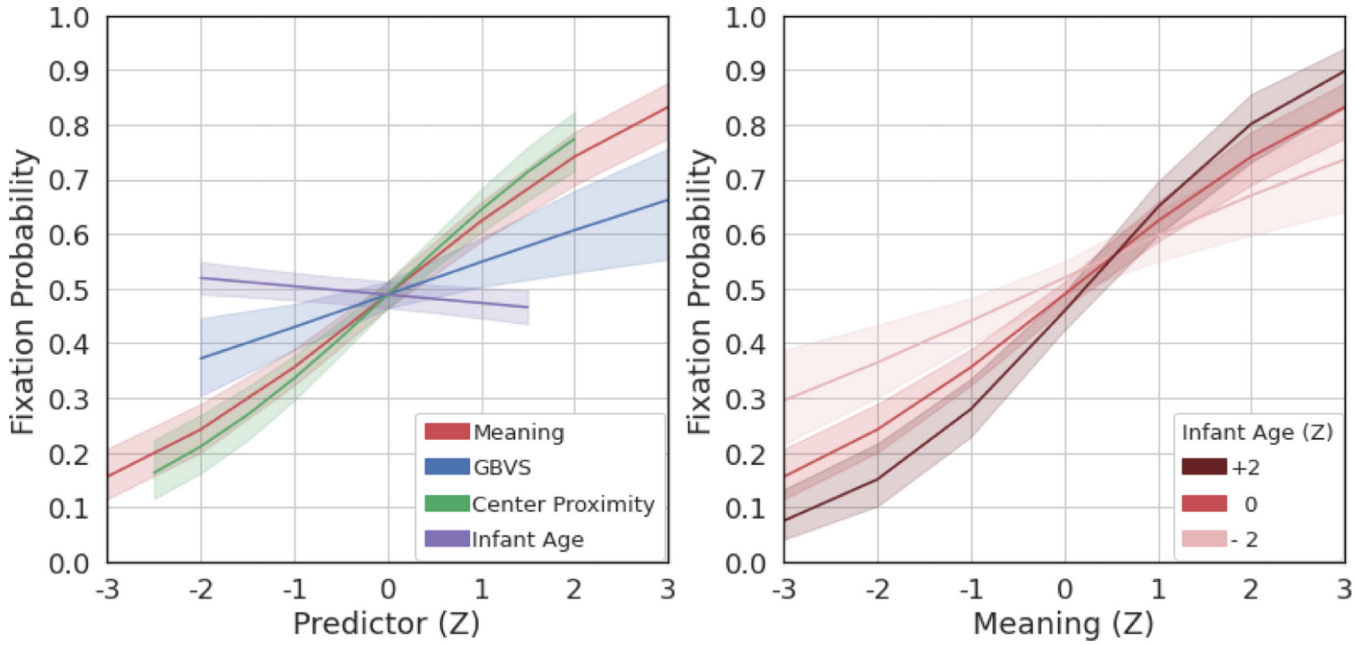


**Figure 4. Line plots of significant effects of the model.**

Line plots of the model estimated effects for each predictor (a) and the meaning by age interaction (b) shown as a function of fixation probability. In each panel, the x-axis reflects predictor values as standard deviations from the mean (where zero reflects the mean) and the y-axis reflects the model estimated probability a scene region with that value will be fixated. In panel b, the 3 different lines reflect different infant ages as standard deviations from the mean where zero reflects the mean age, −2 standard deviations reflects younger infants, and +2 standard deviations reflects older infants. All error bands reflect 95% confidence intervals.

**Table 1.**

Results from the General Linear Mixed-Effects Model

| | Fixed effects | | | | | Random Effects (SD) | |
|---|---|---|---|---|---|---|---|
| **Predictor** | **β** | **95% CI** | ***SE*** | ***z*-statistic** | ***p*** | **Subject** | **Scene** |
| Intercept | −.004 | [−.141, .052] | .049 | −.900 | .368 | .151 | .270 |
| Meaning | .549 | [.436, .663] | .058 | 9.466 | <.001 | .121 | .329 |
| GBVS | .240 | [.092, .388] | .075 | 3.187 | .001 | .214 | .435 |
| Center Proximity | .636 | [.486, .786] | .077 | 8.305 | <.001 | .267 | .422 |
| Infant Age | −.061 | [−.104, −.019] | .022 | –2.817 | .005 | .048 | .046 |
| Meaning x GBVS | −.014 | [−.137, .110] | .063 | −0.216 | .829 | .168 | .334 |
| Meaning x Center Proximity | .078 | [−.042, .198] | .061 | 1.277 | .202 | .124 | .321 |
| GBVS x Center Proximity | −.142 | [−.237, −.048] | .048 | −2.948 | .003 | .172 | .258 |
| Meaning x Infant Age | .116 | [.053, .179] | .032 | 3.603 | <.001 | .096 | .104 |
| GBVS x Infant Age | .017 | [−.076, .110] | .047 | 0.356 | .722 | .084 | .202 |
| Center Proximity x Infant Age | −.020 | [−.119, .098] | .055 | −.186 | .852 | .164 | .216 |

Note: CI = confidence interval.