DOI: 10.1111/infa.12308

RESEARCH ARTICLE

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

Real-world scene perception in infants: What factors guide attention allocation?

Daan R. van Renswoude[1,2] | **Ingmar Visser**[1,2,3] | **Maartje E. J. Raijmakers**[1,2,3,4] | **Tawny Tsang**[5] | **Scott P. Johnson**[5]

[1]Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

[2]Research Priority Area YIELD, Amsterdam, The Netherlands

[3]Amsterdam Brian and Cognition Center, Amsterdam, The Netherlands

[4]Educational Studies & Learn!, Free University, Amsterdam, The Netherlands

[5]Department of Psychology, University of California, Los Angeles, California

Correspondence
Daan R. van Renswoude, Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands.
Email: D.R.vanRenswoude@uva.nl

Funding information
NIH Blueprint for Neuroscience Research, Grant/Award Number: R01-HD73535

Abstract

The foci of visual attention were modeled as a function of perceptual salience, adult fixation locations, and attentional control mechanisms (measured in separate tasks) in infants ($N = 45$, 3- to 15-month-olds) as they viewed static real-world scenes. After controlling for the center bias, the results showed that low-level perceptual salience predicts where infants look. In addition, high-level factors also played a role: Infants fixated parts of the scenes frequently fixated by adults and this effect was stronger for older than younger infants. In line with this finding, infant fixation durations were longer on regions more frequently fixated by adults, implying longer time taken to process the available information. Fixation durations decreased with age, and this decline interacted with orienting skills such that fixation durations decreased faster with age for infants with high orienting skills, relative to infants with low orienting skills. There was a further interaction between fixation durations and selective attention abilities: Infants with low selective attention skills showed a decrease in fixation durations with age, whereas infants with higher selective attention skills showed a slight increase in fixation durations with age. These findings imply that infants' visual processing of static real-world stimuli develops in accord with attentional control.

1 | INTRODUCTION

Visual attention helps infants to select useful information from their environments to learn about the world. Targeted visual exploration may depend on developments in attentional mechanisms (Colombo, 2001) and processing of stimulus-specific properties, such as semantic content, social cues, or perceptual salience (Frank, Vul, & Johnson, 2009). Traditional analyses of gaze patterns often use experimenter-defined "areas of interest" (AOIs), to assess patterns of visual attention to specific areas of the stimulus, reflecting a priori hypotheses about infants' attention to these properties (Gredebäck, Johnson, & von Hofsten, 2009). Free-viewing methods, in which observers view unconstrained scenes as their scanning behavior is recorded, have been used in the past to quantify development of infants' attention to faces in complex static (Amso, Haas, & Markant, 2014; Di Giorgio, Turati, Altoè, & Simion, 2012; Gliga, Elsabbagh, Andravizou, & Johnson, 2009; Gluckman & Johnson, 2013; Kwon, Setoodehnia, Baek, Luck, & Oakes, 2016) and dynamic scenes (Franchak, Heeger, Hasson, & Adolph, 2016; Frank et al., 2009; Frank, Vul, & Saxe, 2012), and to quantify development of fixation durations in typically developing infants (Reynolds, Zhang, & Guy, 2013; Wass & Smith, 2014) and infants at elevated risk for later-developing autism (Wass et al., 2015). These studies have made vital contributions to our knowledge of infant social attention, for example, evidence for a bias to orient toward and attend to faces (Leppänen, 2016). Interpretation of infant visual attention during unconstrained free-viewing tasks, therefore, represents an important theoretical opportunity for understanding perceptual development. However, little is known about infants' spontaneous gaze patterns when viewing static real-world scenes, in particular how gaze patterns develop in accord with the emergence of attentional control.

Visual attention is often studied using eye movements. There is a strong link between the two, because their neural systems overlap (Amso & Scerif, 2015; Nobre, Gitelman, Dias, & Mesulam, 2000). Selecting information is important, because the selected information can be processed and used for higher order cognition, such as learning, thinking, and memorizing. Attending toward useful information is far from trivial, because the environment is often complex with many different objects cluttered together. The goal of the present study is to identify how infants' attention when viewing complex static real-world scenes develops into adult like viewing behavior.

In adults, four main factors are known to guide visual attention during static scene viewing. First, scene characteristics are known to guide eye movements in a low-level way, via perceptual salience (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998), and in a high-level way, via objects (Nuthmann & Henderson, 2010) and expectations (i.e., where things are located [Torralba, Oliva, Castelhano, & Henderson, 2006] and what things look like [Kanan, Tong, Zhang, & Cottrell, 2009]). Second, individual differences in attentional control are known to play a role affecting, for instance, fixation durations (Castelhano & Henderson, 2008; Henderson & Luke, 2014). Third, task demands (e.g., memorizing or searching) influence eye movements (Buswell, 1935; Castelhano, Mack, & Henderson, 2009; Yarbus, 1967). Fourth, adults have a strong general tendency to fixate at the center of a scene, known as the *central bias* (Clarke & Tatler, 2014; Tatler, 2007), and make most saccades in the horizontal directions, known as the *horizontal bias* (Foulsham & Kingstone, 2010; Foulsham, Kingstone, & Underwood, 2008). In analyzing the contribution of cognitive and perceptual factors, it is important to account for these general biases as they are known to explain a lot of the variance during adults' static scene viewing (Tatler & Vincent, 2008, 2009). The present study examines how scene characteristics and attentional control influence infant gaze behavior during free viewing (FV) after controlling for general biases.

1.1 | Scene characteristics

There are two types of scene characteristics that are extensively studied in adults: perceptual salience (i.e., low-level features such as edge content, color, and contrast) and semantic relevance (i.e., objects). Low-level scene characteristics can be expressed as a perceptual salience map of the image (Itti et al., 1998). The "salience-view" of visual attention (Itti & Koch, 2000; Itti et al., 1998), in contrast to the "object-view" discussed below, states that visual attention is driven toward salient regions based on these low-level properties. Perceptual salience is known to predict fixation locations above chance level in adults (Borji, Sihite, & Itti, 2013), but the "salience-view" has been challenged by the "object-view" as several studies show that attention allocation can as well be explained by looking at objects or semantically relevant characteristics of a scene (Einhäuser, Spain, & Perona, 2008; Nuthmann & Henderson, 2010; Stoll, Thrun, Nuthmann, & Einhäuser, 2015).

There is ample evidence for the "object-view" in adults. Perceptual salience has little predictive value for fixation locations above and beyond object locations (Einhäuser et al., 2008, although see Borji et al. (2013)). That is, objects are often perceptually salient and the good performance of perceptual salience as a predictor of fixation locations can be explained by this overlap between perceptual salience and objects. Moreover, fixations are known to fall on the center of objects (Foulsham & Kingstone, 2013; Nuthmann & Henderson, 2010; Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014), which is difficult to explain from a saliency point of view, as saliency models typically highlight object contours and edges. In addition, it is known that salience has little predictive power when task demands are important (Tatler, Hayhoe, Land, & Ballard, 2011). In a recent paper, Henderson and Hayes (2018) showed that meaning is a better predictor of where people look than salience. This is in line with earlier work showing the importance of meaningful factors such as text (Wang & Pomplun, 2012) and semantics (Nyström & Holmqvist, 2008). Taken together, the role of salience in adult scene viewing is limited and adult eye movements are more likely to be driven by meaning.

The fact that meaningful information seems to be more important than perceptual salience in adults is a reason to expect that infant viewing deviates from adult viewing of complex scenes. The development of infant attention is characterized as a shift from relying on largely exogenous features (e.g., perceptual salience) toward relying more on endogenous features (e.g., knowledge) (Johnson, 1990, 2008). However, the findings in the scene-viewing literature are mixed with regard to the role salience plays. For instance, Frank et al. (2009) found that salience was a better predictor of attention in younger infants (3-month-olds) than faces, whereas in older (6- and 9-month-olds) infants, faces were a better predictor than salience. Yet, overall the predictive value of perceptual salience actually increased with age, which is in line with what others report as well (Amso et al., 2014; Franchak et al., 2016). Recent work examining different saliency-based models of visual attention also suggests that salience is a better predictor for adults' eye movements than it is for infants' eye movements (Mahdi, Su, Schlesinger, & Qin, 2017). Although infant visual attentional theory and adult scene-viewing studies would predict that salience becomes less important as infants get older, the experimental results actually point in the other direction. These conflicting findings may be explained by the fact that objects are perceptually salient (Elazary & Itti, 2008). The reported increase in perceptual salience with age may actually be an increase in attention toward meaningful objects.

To potentially guide infants' eye movements, objects must not only be meaningful, but infants must at least have sufficient perceptual and cognitive abilities to select objects in scenes, such as figure–ground segregation and perception of partly occluded objects as wholes (e.g., object unity). Infants do have these abilities; they are born with the ability to achieve figure–ground segregation (Slater et al., 1990) and object unity starts to develop between birth and two months and is robust at four months (Johnson, Slemmer, & Amso, 2004). A categorical learning study by Althaus and Mareschal (2012) also provides some evidence that semantically meaningful information can play a role in guiding infant attention. Older infants (12-month-olds) learned to attend to semantic information (e.g., antlers in animal stimuli, diagnostic of category boundaries), whereas younger infants (4-month-olds) did not learn to use this information.

1.2 | Variation in attentional control

Subtle differences in attentional control occur as a function of task demands; for instance, when adults are performing a search task, fixation durations are shorter than during free-viewing (Castelhano et al., 2009). And when searching for a specific item, for instance, a coffee mug, fixations are more likely to fall on tables and cupboards than on floors and walls (Henderson, 2017; Torralba et al., 2006). Also when instructions do not play a role, differences in attentional control can be identified both within and between individuals. Unema, Pannasch, Joos, and Velichkovsky (2005) showed that individual gaze patterns can be characterized by an ambient mode that allows extraction of global information from the scene and a focal mode during which local information is processed. These different processing modes are characterized by different eye movements; the ambient mode is characterized by short fixation durations and saccades of long amplitude, whereas the focal mode has longer fixation durations and shorter saccade amplitudes. Between individuals, there are also differences in fixation durations (Henderson & Luke, 2014) and saccade amplitude (Castelhano & Henderson, 2008). Risko, Anderson, Lanthier, and Kingstone (2012) even found a link between gaze patterns and personality. Individuals that scored higher on the personality trait "openness" were more likely to scan a greater proportion of the scenes than individuals with lower openness scores. How do these individual differences in attentional control develop?

Visual attention of infants develops rapidly (Bronson, 1990, 1994; Hunnius & Geuze, 2004): 1- to 2-month-old infants often have long fixations, known as sticky fixations. Already at the age of 3 to 4 months, fixation durations decrease (<500 ms) and infants gain more endogenous control over their eye movements. This decrease in fixation durations continues until adolescence (Luna, Velanova, & Geier, 2008) and is also reported in scene-viewing studies with infants (Helo, Rämä, Pannasch, & Meary, 2016; Wass & Smith, 2014). Individual differences in fixation durations are linked to attentional and behavioral control in childhood (Papageorgiou et al., 2014). The voluntarily control over eye movements develops hand in hand with attentional mechanisms (Colombo, 2001). In scene perception, important attentional mechanisms include *spatial orienting* (the ability to select an object or salient region in a scene) and *selective attention* (the ability to direct visual attention by the inhibition of irrelevant information).

Spatial orienting can be measured using the classic visual search (VS) paradigm that requires selecting a target that differs on one dimension (e.g., orientation or color) from an array of identical distracters (Treisman & Gelade, 1980). Experiments using this paradigm can successfully be conducted with young infants (Dannemiller, 1998, 2000) and have a good test--retest reliability (Hessels, Hooge, & Kemner, 2016). Evidence for development in efficient visual selection comes from an increase in the tendency to detect the target across trials with age. Frank, Amso, and Johnson (2014) found that orienting abilities (measured using the VS task) predicted attention toward faces in 3-month-old infants (but not older infants).

Selective attention is often assessed in a *spatial negative priming* (SNP) paradigm. This task consists of two phases: *prime* and *probe*. In the prime phase, two stimuli are briefly presented, a small animated target and a static "distracter" considerably less salient than the target. The probe phase commences after a short delay. The target again appears either at the location previously occupied by the distracter or at a new location (the distracter is absent during the probe phase). Evidence for efficient attentional inhibition comes from lengthened eye movement latencies toward the target on trials when the target is located at the location previously occupied by the distracter versus trials when the target is at a location *not* previously occupied by the distracter.

In the current study, we include measures of orienting abilities and selective attention in our analyses of infant attention allocation to examine its predictive value in scene viewing. These measures are unlikely to directly elicit responses toward specific regions in a scene. Instead, these attentional mechanisms may *interact* with other predictive factors. For instance, infants with better orienting skills may be more likely to attend to salience information.

1.3 | General biases

During free viewing, adults tend to focus more on the center than on peripheral regions of scenes (central bias, Clarke & Tatler, 2014) and make most saccades in relatively horizontal directions, (horizontal bias, Foulsham et al., 2008). These biases exist irrespective of the positional placement of pre-trial fixation markers (Foulsham et al., 2008; Tatler, Baddeley, & Gilchrist, 2005), presentation in landscape, square, or portrait format (Foulsham, Teszka, & Kingstone, 2011), or scene content (Tatler, 2007). These biases have been suggested to be acquired strategies for maximizing information uptake (Tatler, 2007). They are important in modeling adult eye movements (Foulsham & Kingstone, 2012), and they predict fixation locations above chance level (Tatler & Vincent, 2009). Recently, it was shown that infants also demonstrate the horizontal bias (van Renswoude, Johnson, Raijmakers, & Visser, 2016) and the central bias (van Renswoude et al., 2019). This indicates that infants, similar to adults, are more likely to fixate central locations and locations along the horizon. These central and horizontal locations are also more likely to be salient and contain meaningful information (Tatler, 2007). Therefore, we controlled for the influence of these general biases by adding them as covariate in our analysis.

1.4 | Current study

The aim of the current study is to understand how scene characteristics and attentional mechanisms influence static real-world scene perception in infants. We incorporated these factors in a single analysis to assess the unique contribution of each. These factors are often studied in isolation, but it is important to assess them simultaneously because there is much overlap between these different factors. There is a general bias toward the center of a scene, which is also often the location of meaningful objects (Tatler, 2007). Furthermore, objects are often perceptually salient (Einhäuser et al., 2008). The influence of all these factors may change with age (Frank et al., 2009) due to the development of attentional mechanisms (Colombo, 2001). Combining all these factors into one model allows us to account for correlations between the factors and study their unique effects on visual attention in explaining differences between infant and adult viewing during development. To model the foci of visual attention as a function of scene characteristics and differences in attentional control while accounting for the center bias, we adopted an approach first described by Nuthmann and Einhäuser (2015, see Figure 1 for a schematic representation).

A general linear mixed model (GLMM) was used to assess the influence of scene characteristics and their interactions with age and attentional control mechanisms, while accounting for the influence of the center bias. Infants freely viewed (FV) a set of natural visual scenes, which consisted of full-colored photographs. For analysis purposes only, all scenes were overlaid with a 6-by-8 grid (see the leftmost scene in Figure 1) to define the variables. The dependent variable *fixation location* was operationalized as to whether each grid location was fixated (1) or not (0) during the presentation of

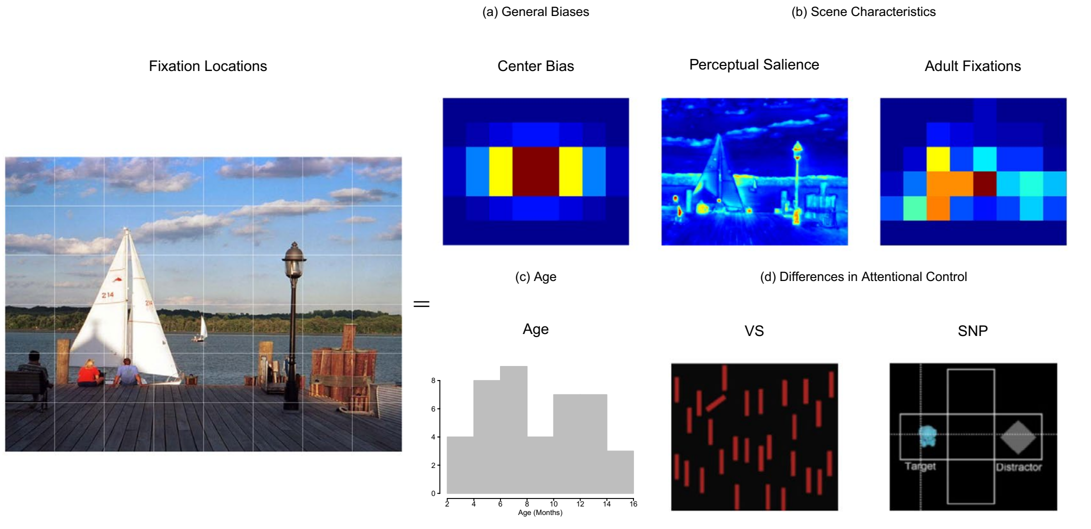

FIGURE 1 Schematic representations of the relationship between fixation locations and factors that guide eye movements. Fixation locations are a function of (a) general biases, such as the central and horizontal bias; (b) scene characteristics, such as perceptual salience and adult fixation locations; (c) age; and (d) differences in attentional control, such as orienting skills and selective attention. General biases and scene characteristics can have a direct influence on attention allocation, whereas age, orienting skills, and selective attention can influence attention in an indirect way

the scene. To define the *center bias* variable, we generated a salience map that highlighted the center and horizontal regions of the scenes (see the upper left panel in Figure 1). *Perceptual salience* was operationalized as the maximal salience measure in every patch of the 6-by-8 grid, based on the Itti and Koch (2000) algorithm (see the upper middle panel in Figure 1). We recorded eye movements of adults (who bring top-down knowledge of real-world scenes to the task) and calculated the total number of adult fixations within each patch of the grid for every scene (see the upper right panel in Figure 1). Differences in performance as a function of age were accommodated by using a wide age range of infants (see the lower left panel in Figure 1). Attentional mechanisms were assessed using the VS (lower middle panel in Figure 1) and SNP (lower right panel in Figure 1) tasks, respectively.

Although the GLM approach allows us to identify the unique contribution the factors that are important in the location of attention, it does not incorporate the duration of attention. Because fixation durations also provide important information, a linear mixed model was fitted with fixation durations as the dependent variable and center bias, perceptual salience, adult fixations locations, age, VS, SNP, and the interactions with age as independent variables.

2 | METHODS

2.1 | Participants

2.1.1 | Infants

Infant participants were recruited from birth records provided by the state. A parent or caregiver accompanied infant participants at all times and was asked not to talk or direct the infant's attention during stimulus presentation. Forty-five infants 3–15 months of age (M age = 8.6 months, $SD = 3.4$;

17 male; see Figure 2) completed all aspects of the experimental protocol, providing eye-tracking data for the VS, spatial negative priming, and free-viewing (FV) tasks. An additional 27 infants were observed but excluded from the final sample due to reasons compromising the data quality, such as poor calibration (n = 16), excessive head or body motion (n = 8), or inattention due to fussiness or sleepiness (n = 3). All infant participants were full -term with no known developmental difficulties.

2.1.2 | Adults

Forty-seven adults (M age = 21.50 years, SD = 4.55; 20 male) provided eye-tracking data for the FV task. Participants were recruited through the undergraduate subject pool and were given course credit for participating. All participants had corrected-to-normal vision and were not color-blind. The data of the adults were used as predictor of the infant data. The present study was conducted according to the Declaration of Helsinki guidelines, with written informed consent obtained from adult participants and a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Institutional Review Board at the University of California, Los Angeles as part of the project "Brain Mechanisms of Visual Development."

2.2 | Stimuli and presentation

2.2.1 | Free viewing

Twenty-eight full-color photographs were selected from the LabelMe database (Russell, Torralba, Murphy, & Freeman, 2008). The photographs depicted natural scenes with and without human artifacts, cityscapes including people with indiscernible faces, or rooms with household items. All photographs but one were presented in landscape orientation (1,024 pixels by 768 pixels); the other photograph was presented in portrait orientation (576 pixels by 768 pixels). The photographs were centered within the monitor against a black background. Each photograph was presented for 4 s, and an attention-grabbing centering stimulus appeared between trials.

2.2.2 | Visual search

The stimuli and procedure for the VS task were similar to those used previously by Amso and Johnson (2006) and Frank et al. (2014). Twenty-seven red vertical distracter rods (1.9 × 7.0 cm each,

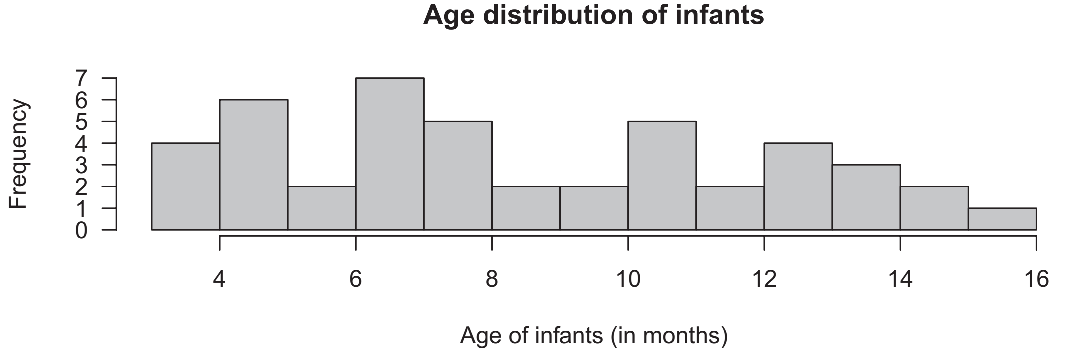

FIGURE 2 Histogram of the age distribution of infants

0.9° × 3.3°) and one red target rod were presented against a black background. There were two types of target rods—a stationary rod oriented 30°, 60°, or 90° from vertical, or a moving rod that traversed through 6.5 cm (3.1°) at 1, 1.5, or 2 Hz. A total of 48 trials were presented in random order for each infant. Trials had a maximum length of 4 s, but terminated if the target rod was fixated for a cumulative total of 100 ms within a 30-pixel radius.

2.2.3 | Spatial negative priming

The stimuli and procedure for the SNP task were similar to those employed by Amso and Johnson (2005, 2008). Each of the 48 trials began with a white cross-shaped grid against a black background with four possible target locations and consisted of a prime and a probe. Primes and probes lasted 2000 ms each and were separated by an inter-stimulus interval of 550 ms. In the prime trials, a target and a distracter (gray diamond) appeared simultaneously in two different locations. Targets were randomly selected from a library of small, dynamic animations with accompanying sounds. This was followed by either an ignored repetition (IR) or a control probe trial. The IR and control trials differed in the location of the target. In the IR trial, the target appeared in the location previously occupied by the distracter; in the control trial, the target appeared in a new location. Target and distracter locations were randomized across trials.

2.3 | Experimental procedure and apparatus

Eye movements were recorded using a remote-optics corneal reflection eye-tracker (SR EyeLink 1000). The eye-tracking system has an average tracking accuracy of 0.5° of visual angle and collected data at 500 Hz. Visual stimuli were presented on a 22-inch ViewSonic VX2268xm monitor in full color. All participants were individually tested and seated approximately 60 cm from the monitor. After the participant was properly situated with respect to the eye-tracking equipment, lights were dimmed and black curtains were drawn such that only the stimuli presented on the computer monitor could be seen. The experimenter controlled stimulus presentation and eye-tracking equipment from an adjoining room but could monitor the participant with video feed from an eye-tracking camera and a second camera capturing a full-body video of the seated participant. Prior to beginning experimental procedures, each participant's point of gaze was calibrated using a 5-point calibration scheme. The experiment began once calibration criteria had been reached. The mean calibration error was 0.89° of visual angle, and this error did not correlate significantly with the age of the infants ($r = -0.12$, $p = 0.435$). The infant experimental protocol proceeded in a fixed order (VS, SNP, and FV) due to the need to use more engaging tasks throughout the testing session (Frank et al., 2014). The adults were only presented the images in the FV task.

2.4 | Fixation identification

The identification of fixations from infant eye-tracking data is far from trivial. Infants' raw gaze signals are known to be noisy (Wass, Smith, & Johnson, 2013) and correlated with outcome measures such as age (Wass, Forssman, & Leppänen, 2014). In order to limit the influence of noise, fixations were identified using an R (R Core Team, 2017) software package *gazepath* (van Renswoude et al., 2018) that parses raw eye-tracking data into fixations and saccades. This R package is designed to identify fixations in noisy infant data while accounting for individual differences in data quality. To further increase the reliability of the data, fixations had a minimum duration of 100 ms and a maximum variance of 0.6° of visual angle. Only the fixation data are reported in this paper, and the saccade data from the FV task were published in van Renswoude et al. (2016).

2.5 | Dependent variables

2.5.1 | Fixation location

To create the dependent variable of fixation locations, all scenes were overlaid with a 6 × 8 grid (each patch 128 × 128 pixels, approximately 4° × 4° of visual angle)[1]. For every participant (45) on every trial (28), every patch (48) of the grid was either fixated (1), when one or more fixations were identified within that patch, or not fixated (0), when no fixations were identified within that patch. Eye-tracking data pertaining to the first fixation were excluded from data analyses to remove fixation artifacts from the inter-trial centering stimulus. An illustration of this procedure can be seen in Figure 3a. This figure shows two gaze patterns of different infants in the panel at the left and their corresponding fixation maps in the right panel.

2.5.2 | Fixation duration

Throughout this paper, the term fixation duration refers to the period between two saccadic eye movements. First fixations of each trial were excluded from data analyses to remove fixation artifacts from the inter-trial centering stimulus. Fixation durations showed a typical skewed-to-the-right distribution, which is common with fixation durations (Helo, Pannasch, Sirri, & Rämä, 2014; Velichkovsky et al., 2000). To meet the assumptions of the linear mixed model, fixations were log-transformed to assure a normal distribution of the residuals.

2.6 | Independent variables

2.6.1 | General biases

The central bias included in the analysis used a saliency map that had the highest values in the middle, decreasing toward the sides. In order to also incorporate the horizontal bias, the values decreased more in the vertical than in the horizontal direction. More specifically, we followed the recommendations of Clarke and Tatler (2014) and used a multivariate Gaussian probability density function with mean zero and covariance structure [σ^2;0;0; $\sigma^2\nu$], where σ is 0.17 and ν is 0.45. The mean values of this Gaussian probability density function within each grid cell were used in the model.

2.6.2 | Perceptual salience

Visual saliency maps were generated using the algorithm described by Itti and Koch (2000). This algorithm was chosen as it is known to perform reasonably well in predicting infant fixations. Mahdi et al. (2017) compared several saliency algorithms predicting infant fixations; the Itti and Koch algorithm came out among the best on several measures. In addition, the algorithm is biologically plausible: The algorithm aims to mimic tendency of the visual system, already in place in infants, to select what stands out as salient. Here, we used the Python implementation from https://github.com/akisato-/pySaliencyMap. The generated saliency maps provide a saliency measure for every pixel. To produce these saliency maps on the 6 × 8 grid, the maximal saliency

[1]For the scene in portrait format, the same grid was used making each patch 128 × 64 pixels, approximately 4° × 2° of visual angle.

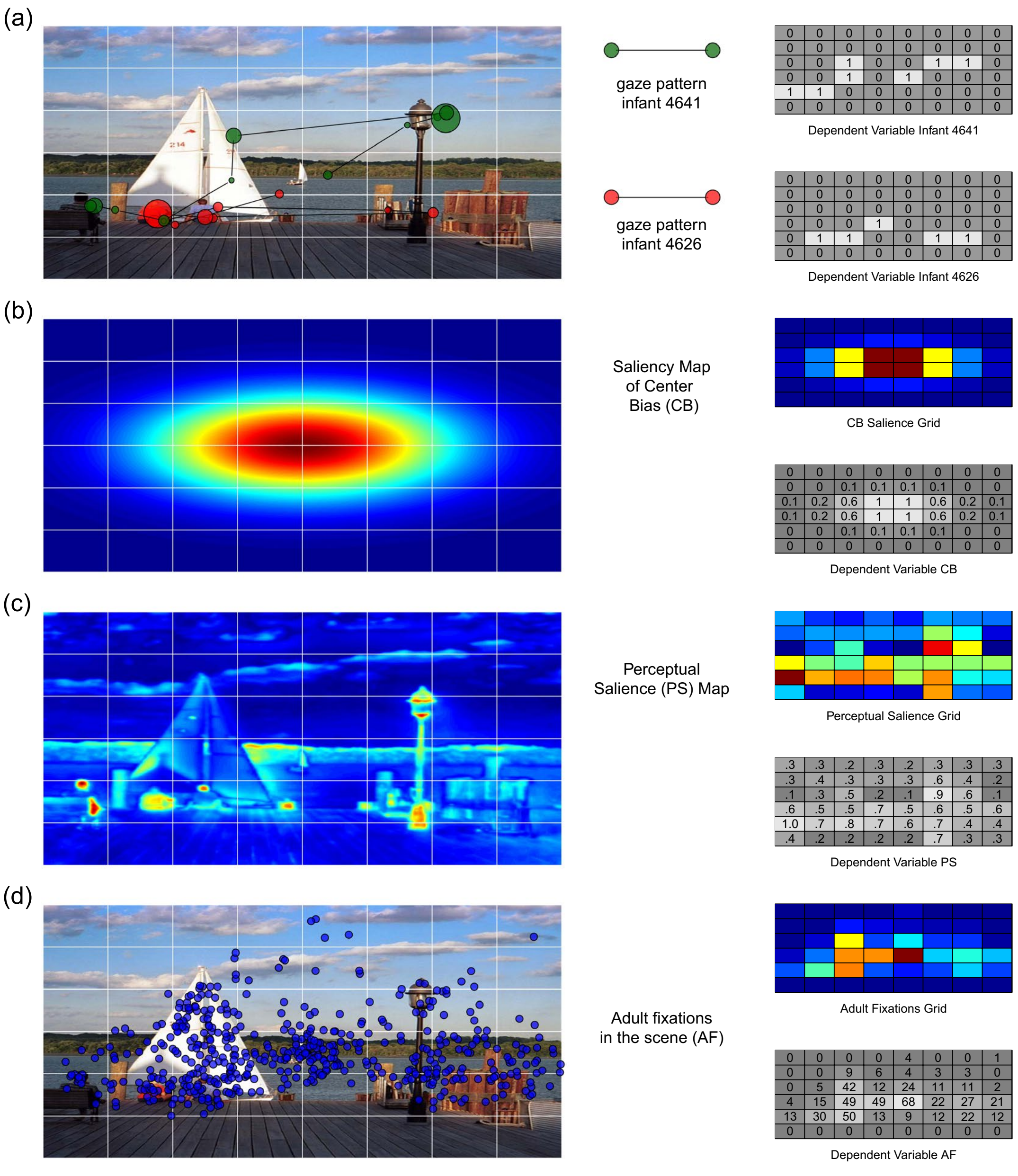

FIGURE 3 Illustration of the method to measure the different variables. The left part of panel a shows two gaze patterns of two infants, and the right part shows the corresponding dependent variable on the 6 × 8 grid. Panel b shows the center bias input and the corresponding 6 × 8 grid. Panel c shows an example of the perceptual salience map of a scene and its corresponding values on the 6 × 8 grid. Panel d shows all adult fixations on a scene and the corresponding values on the 6 × 8 grid. In the analysis, all variables are scaled to have mean 0 and standard deviation 1

values of every patch were calculated. Values ranged from 0 (very low perceptual salience) to 1 (very high perceptual salience). Figure 3c provides an example of the procedure. On the left, the saliency map is shown, and on the right, the corresponding grid saliency map and numerical values per patch are shown.

2.6.3 | Adult fixations

For every patch on the 6 × 8 grid overlaid on every image, the total number of fixations made by adults was calculated. The most fixated patch was fixated 119 times, while the least fixated patch received no fixations. Figure 3d provides an example of the procedure. On the left, all adult fixations over the scene are shown, and on the right, the corresponding grid saliency map and numerical values per patch are shown.

2.6.4 | Visual search

The measure of orienting abilities was obtained via the VS task. For every participant, the percentage of trials in which a target rod was fixated was calculated and served as the orienting measure. For the purposes of our study, we computed a composite accuracy measure for moving and static targets regardless of speed or orientation. Trials on which there were no data were omitted from the analysis. Participants contributed an average of 11.08 out of 24 trials ($SD = 4.57$ trials) for the static condition and an average of 19.88 out of 24 trials ($SD = 5.17$ trials) for the moving target condition. The mean accuracy score was 0.71 ($SD = 0.18$, range = 0.38–1), which shows that overall, infants performed reasonably well with enough variation between individuals.

2.6.5 | Spatial negative priming

The measure of selective attention was obtained via the SNP task. For every participant, the mean difference between the IR and control trials in latency to fixate on the target during the probe phase of each trial was calculated and served as the selective attention measure. Data from individual trials were considered invalid if the infant did not fixate on the target during the prime presentation, or initiated a saccade prior to target presentation in probes (e.g., 167 ms before the stimulus appeared). A positive value indicates inhibition of the attended location, whereas a negative value indicates a location facilitation effect, which is analogous to an immature response to SNP (Amso & Johnson, 2008). The mean response time was −35.28 ms ($SD = 167.34$, range = −607.81–372.09) and did not significantly deviate from 0 ($t\ (44) = -1.41$, $p = 0.16$). This implies there was no overall inhibition or facilitation effect, but looking at the range and standard deviation, there is a lot of variation between individuals.

2.7 | Model specifications

A general linear mixed model and a linear mixed model were used to analyze the data of fixation locations and durations, respectively. Mixed models provide some advantages over traditional ANOVA: First, mixed models can account for individual characteristics. No grouping over scenes or participants is required as both sources of variance (scenes and participants) are accounted for as random effects. Second, by simultaneously including all factors in a single analysis, the contribution of each factor can be made explicit while controlling for the influence of other factors.

Both analyses were conducted using the *lme4* package in R (R Core Team, 2017). All independent variables were scaled to have means of zero and standard deviations of one. We report the beta values, standardized errors, z-values (for fixation locations), and t-values (for fixation durations) to determine whether estimates differ from zero and thus have a significant impact.

2.7.1 | Fixation locations

A GLMM with fixation location during FV as the dependent variable and the center bias, perceptual salience, adult fixations, age, VS, and SNP as independent variables (as well as the interactions between age, VS, and SNP and the other variables) was fitted to the data. The *glmer* function of the *lme4* package was used with the *logit* link function. This resulted in a model with 15 predictor terms (six main effects and nine interaction effects). All predictors were included as fixed effects, and scene and participant were included as random effects. For scenes, the full random-effects structure was fitted; for participants, only the random-effects structure for the center bias, perceptual salience, and adult fixation locations was fitted. This was done in order to have a model that included most of the random-effects structure and was still able to converge.

2.7.2 | Fixation durations

A linear mixed model with fixation durations as dependent variable and the center bias, perceptual salience, adult fixations, age, VS, SNP, and the interactions between age, VS, and SNP as independent variables was fitted using the *lmer* function of the *lme4* package. As in the location analysis, we aimed to fit the full random-effects structure for scenes and for participants the random-effects structure for the center bias, perceptual salience, and adult fixation locations. However, this model turned out to be too complex to converge. In order to be able to fit the fixation duration model, we set the covariance structure between the random effects at zero. This resulted in a model that did converge.

3 | RESULTS

In total, the 45 infants made 6478 fixations on an average of 26.16 out of 28 scenes ($SD = 3.13$) per person. For all infants (45) and scenes (28), fixation locations were mapped onto the 6×8 grid, except for the scenes that had no fixations (83 scenes, 7%). This resulted in a total of 56,496 observations of locations that were fixated (1) or not (0). Over participants and scenes, 8 percent of the patches were fixated, which corresponds to 4–5 fixations per infant per trial.

3.1 | Fixation locations

A GLMM was fitted to the data, with fixation locations as dependent variable and center bias (CB), perceptual salience (PS), and adult fixations (AF), and their interactions with age, VS, and SNP as independent variables. Table 1 shows the correlations between these variables, and Table 2 shows the fixed and random effects. All three main factors (CB, PS, and AF) and the intercept have both fixed and random effects for scenes and participants. The other variables only have random effects for scenes, but not for participants (see the Methods section). The fixed effect can be interpreted as regression coefficients in ordinary regression. For instance, the negative beta for the intercept implies that most patches were unlikely to be fixated. This is intuitive as there are only 4–5 fixations per trial and 64 patches. Positive beta values indicate that patches were more likely to be fixated. For instance, the positive beta value of PS indicates that more perceptual salient patches were more likely to be fixated than less perceptual salient patches. The random effects reflect the variation between individuals and scenes. For every individual and every scene, the effects of the predictor variables can differ. For instance, fixation locations in some scenes may be better predicted by PS than fixation locations in other scenes. The same applies for individuals: Some individuals may, for instance, have a stronger

TABLE 1 The left part shows the correlations of the independent variables on the level of the location: center bias (CB), perceptual salience (PS), and adult fixations (AF). The right part shows the correlations of the independent variables on the level of the individual: age, visual search (VS), and spatial negative priming (SNP)

	Location (N = 1344)				Individual (N = 45)		
	CB	**PS**	**AF**		**Age**	**VS**	**SNP**
CB	–			Age	–		
PS	0.21**	–		VS	0.42*	–	
AF	0.72**	0.32**	–	SNP	0.22	0.24	–

*$p < 0.01$.
**$p < 0.001$.

center bias than other individuals. To model these two sources of variance, regression coefficients are estimated for each scene and all individuals. Every random effect is the standard deviation of these regression coefficients.

3.1.1 | Main effects

The fixed effects in Table 2 show that the three main factors (CB, PS, and AF) are able to predict fixation locations. Looking at the columns with the beta values and z-values, it can be seen that CB has the largest effect and PS and AF contribute about equally. To get a better understanding of these effects, Figure 4 shows a graphical representation. The three plots of the main effects in Figure 4 show the probability of fixations as a function of CB, PS, and AF. All predictors have a positive influence meaning that central locations, salient locations, and adult fixation locations are more likely to be fixated than less central, less salient, and locations less frequently fixated by adults. The shaded region corresponds to the 95% confidence intervals, and the plotted effects are estimated using the R package effects (Fox, 2003).

It is important to note that the effects visualized in Figure 4 all make a unique contribution to the model. This is important to consider since these three predictors do correlate with each other (see Table 1). In the GLMM analysis, the relationships between the factors are taken into account, and therefore, it can be concluded that both perceptual salience and adult fixation can predict infant fixation location above and beyond the central bias.

3.1.2 | Interaction effects

Apart from the main effects, the analysis also showed two significant interaction effects. The most prominent one is the interaction between AF and age, given its relatively large beta value and z-value. The left panel of Figure 5 shows the relationship between AF and fixation probability for six different age-groups. Note that the age-groups are only for illustrative purposes; in the actual analysis, age is a continuous variable. The figure clearly shows that older infants have a stronger relationship between AF and fixation probability than younger infants, as can be seen from the steeper slopes. In other words, adult fixation locations are a better predictor for older than for younger infant fixation locations.

The right panel of Figure 5 shows the interactions between CB and VS. This interaction indicates that infants that do well on the search task have a higher probability to fixate the center regions than infants that perform worse on the search task. In other words, the ability to efficiently select information in a visual search task transfers to efficient looking behavior in the free-viewing task.

TABLE 2 Beta estimates, standard errors (*SE*) and z-values for the center bias (CB), perceptual salience (PS), adult fixations (AF), age, VS, and SNP, and interactions of CB, PS, and AF with age, VS, and SNP

	Fixed effect				Random effects, *SD*	
	Beta	*SE*	*z*-value	Sig.	By scene	By participant
Intercept	−2.760	0.048	−57.389	***	0.163	0.205
CB	0.337	0.045	7.456	***	0.205	0.084
PS	0.295	0.047	6.301	***	0.222	0.058
AF	0.319	0.050	6.324	***	0.227	0.109
age	0.047	0.043	1.101		0.066	
VS	−0.058	0.043	−1.360		0.073	
SNP	−0.022	0.040	−0.564		0.066	
CB × age	−0.038	0.031	−1.205		0.100	
CB × VS	0.058	0.028	2.076	*	0.068	
CB × SNP	−0.031	0.026	−1.198		0.063	
PS × age	0.030	0.024	1.256		0.055	
PS × VS	0.016	0.023	0.675		0.040	
PS × SNP	−0.014	0.022	−0.668		0.044	
AF × age	0.149	0.036	4.124	***	0.121	
AF × VS	−0.026	0.031	−0.866		0.071	
AF × SNP	0.025	0.026	0.965		0.040	

*$p < 0.05$.
***$p < 0.001$.

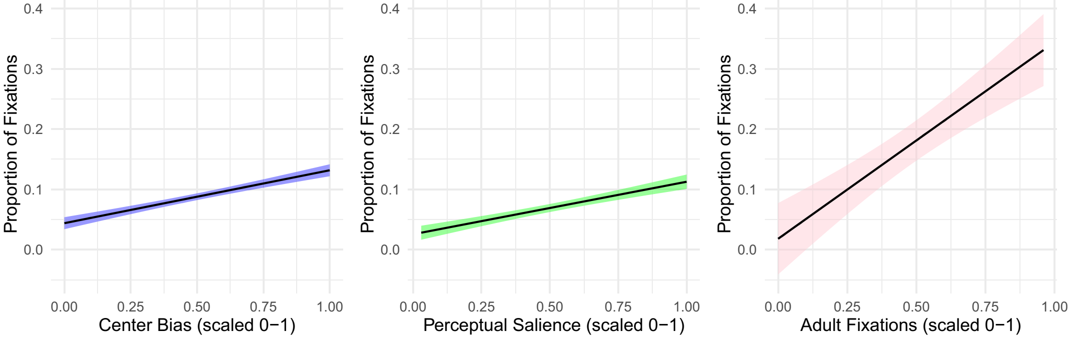

FIGURE 4 Plots of the main effects for the center bias (CB, left panel), perceptual salience (middle panel), and adult fixations (right panel). All have a positive influence meaning that central locations, salient locations, and adult fixation locations are more likely to be fixated than less central, less salient and locations less frequently fixated by adults. The shaded region corresponds to the 95% confidence intervals, and the plotted effects are estimated using the R package effects (Fox, 2003)

3.1.3 | Robustness analysis

In order to assess the robustness of the reported effects, a robustness analysis was performed. This is important as the use of a grid in the analyses and the noisy nature of infant eye movements raise methodological concerns. First, the size of the grid patches is somewhat arbitrary and the results should not depend on the grid size. However, changing the grid size also affects the power as this would lead to less data when increasing the grid size and more data when reducing the grid size. In order to overcome this problem, we decided to redo the analysis with noise added to the fixation locations. By adding noise to the fixation locations, fixation locations near the borders of patches are likely to fall in a different patch after adding noise. Importantly, this should not affect the results as the results should not depend on the somewhat arbitrarily chosen patch borders. Second, adding noise also allows to examine to what extent the results are influenced by noise. With infants that are difficult to calibrate and infant eye movements that are known to be noisy (Wass et al., 2014), it is very likely that there is difference between the fixation locations reported by the eye-tracker and the actual fixation locations. Again, the results should not depend on measurement error, and by adding noise, we gain insight to what extent the results are influenced by noise.

To perform the robustness analysis, random variables with mean 0 and standard deviation of 20 pixels were added to the measured *x*- and *y*-coordinates of the fixation locations. This resulted in a mean displacement of 0.63 ($SD = 0.33$) degrees of visual angle which is similar to the calibration error. The analysis was redone using the noisy fixation locations, and the results were very similar (see Table 3). All main effects of CB, PS, and AF and the interaction between age and AF are still significant. Only the interaction between CB and VS is no longer significant. This implies that the main findings are robust and are unlikely to depend on arbitrarily chosen patch borders and/or measurement error inevitable in infants.

3.1.4 | Conclusions: fixation locations

In sum, this analysis showed that both perceptual salience and adult fixation locations have unique contributions to prediction of fixation locations in infants on top of the influence of CB. Furthermore,

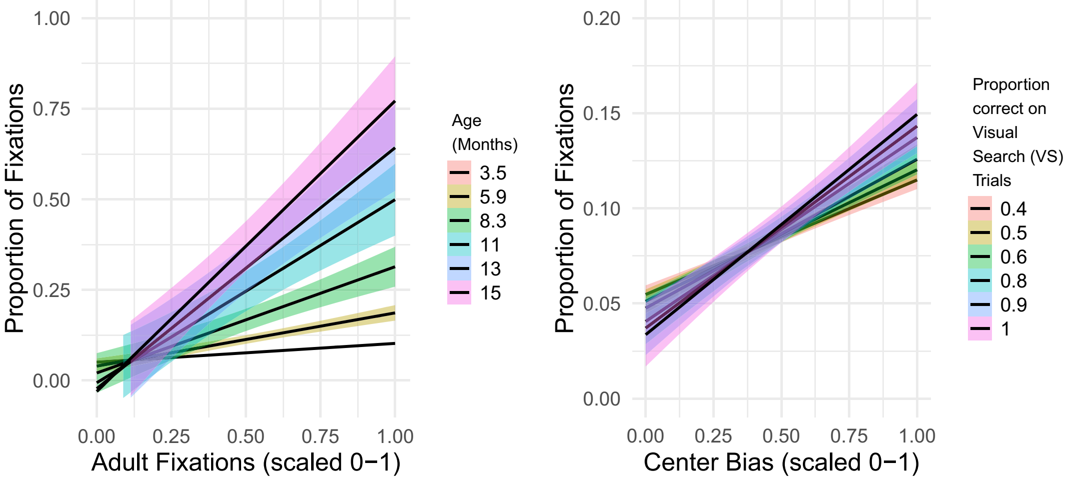

FIGURE 5 Plot of the interaction effects of adult fixations (AF) with age (left panel) and the center bias (CB) with VS (right panel). For older infants, the relation between adult fixations and fixation location is stronger than for younger infants, as can be seen from the steeper slopes for older infants. The interaction effect between CB and VS shows that infants who perform well on the search task make more central fixations than infants with lower search task scores. The shaded region corresponds to the 95% confidence intervals, and the plotted effects are estimated using the R package effects (Fox, 2003). Note that the different levels are for illustration purposes only and all variables are continuous in the actual analysis

the role of AF increases with age. These results were confirmed in a robustness analysis in which noise was added to the fixation locations to show that the results are not an artifact of arbitrarily chosen patch borders and/or measurement error.

3.2 | Fixation durations

A linear mixed model was fitted to the data, with (log-transformed[2]) fixation durations during FV as the dependent variable. The independent variables were CB, PS, AF, age, VS, SNP, and the interactions between CB, PS, AF, VS, SNP, and age. Table 4 shows the summary statistics. There were main effects for AF and age, and there were two interaction effects for age with VS and age with SNP.

3.2.1 | Main effects

Figure 6 shows a graphical representation of the significant main effects in the fixation duration analysis. The left panel shows the main effect of fixation durations that decreased with age. The right panel shows the main effect of fixation durations increasing with adult fixation locations. That is, at regions more frequently fixated by adults, the fixation durations of infants were longer than at regions less frequently fixated by adults.

3.2.2 | Interaction effects

To visualize the interaction effects, the VS and SNP variables were split into six groups. The left panel of Figure 7 shows that fixation durations decreased more rapidly with age for infants with high scores on the VS task relative to infants with low scores on the VS task. The right panel shows that fixation durations of infants that showed an inhibition effect on the SNP task decreased with age, whereas the fixation durations of infants that did not show the inhibition effect on the SNP task increased with age. Again, it is important to note that these plots serve for interpretation purposes only and that the actual analysis was done with continuous variables and not the grouped version that is displayed here.

3.2.3 | Conclusions: fixation durations

In sum, the analysis of fixation durations showed that fixation durations decreased with age as a function of both the VS and SNP tasks. For infants with higher orienting skills, fixation durations were more strongly negatively related to age than for infants who had lower orienting skills. For infants that showed an inhibition effect on the SNP task, the opposite was true. Their fixation durations were uncorrelated with age, whereas fixation durations of infants that did not show an inhibition effect on the SNP task were negatively related to age.

4 | DISCUSSION

In this study, we analyzed the foci of infants' visual attention when viewing static real-world scenes as a function of perceptual salience, adult fixation locations, age, and attentional mechanisms, while accounting for the center bias. We showed that all these factors play a role in guiding infants' eye

[2]Log transformation of fixation durations was necessary to assure normal distribution of the residuals; however, the results were similar compared to the analysis without log-transformed fixation durations.

TABLE 3 Robustness analysis. Beta estimates, standard errors (*SE*), and *z*-values for the center bias (CB), perceptual salience (PS), adult fixations (AF), age, VS, and SNP, and interactions of CB, PS, and AF with age, VS, and SNP

	Fixed effects					
	Beta	*SE*	*z*-value	Sig.	By scene	By participant
Intercept	−2.705	0.049	−55.110	***	0.167	0.212
CB	0.348	0.042	8.317	***	0.184	0.083
PS	0.286	0.046	6.172	***	0.221	0.048
AF	0.307	0.050	6.157	***	0.227	0.101
age	0.066	0.043	1.534		0.058	
VS	−0.044	0.044	−1.000		0.074	
SNP	−0.017	0.041	−0.407		0.071	
CB × age	−0.041	0.030	−1.383		0.089	
CB × VS	0.039	0.027	1.440		0.058	
CB × SNP	−0.016	0.027	−0.571		0.077	
PS × age	0.032	0.024	1.327		0.068	
PS × VS	0.005	0.023	0.215		0.054	
PS × SNP	−0.014	0.022	−0.656		0.052	
AF × age	0.140	0.034	4.065	***	0.114	
AF × VS	0.007	0.029	0.250		0.059	
AF × SNP	0.008	0.028	0.275		0.068	

*$p < 0.05$.
***$p < 0.001$.

movements. Both perceptual salience and locations of adult fixations made unique contributions to attention allocation, after controlling for the center bias. Age also played an important role: Older infants were more likely to fixate parts of scenes more frequently fixated by adults relative to younger infants. The analysis of fixation durations showed that fixations were longer on regions more frequently fixated by adults. The duration analysis also showed a decrease in fixation durations with age, an effect that was moderated by the orienting and selective attention skills of infants. Better orienting skills resulted in a stronger decrease in fixation durations with age than lower orienting skills. Fixation durations of infants that showed an inhibition effect on the selective attention task were not correlated with age, whereas infants that did not show an inhibition effect on the selective attention task showed the expected decrease in fixation durations with age.

As noted, we found that fixation durations decreased with age. This is in line with findings of other studies using similar stimuli (Helo et al., 2016; Wass & Smith, 2014) and extends the findings of Helo et al. (2014) and Açık, Sarwary, Schultze-Kraft, Onat, and König (2010), who found decreasing fixation durations for children in the age range of 3–10 years. An explanation of this decrease is that younger infants may need more time to process the available information in the region they fixate. This explanation is also in line with the finding that fixation durations were longer on regions frequently fixated by adults. These regions are likely to contain more information as adults are known to fixate meaningful objects (Einhäuser et al., 2008; Nuthmann & Henderson, 2010; Stoll et al., 2015). The prolonged fixation durations in these regions are likely to reflect the time needed to process the available information.

TABLE 4 Beta estimates, standard errors (*SE*) and *t*-value for CB, PS, AF, age, VS, SNP and the interactions between CB, PS, AF, VS, SNP and age

	Fixed effects				Random Effects, *SD*	
	Beta	*SE*	*t*-value	Sig.	By-scene	By-participant
Intercept	5.912	0.023	252.218	***	0.040	0.117
CB	−0.016	0.012	−1.391		0.028	0.035
PS	0.009	0.008	1.115		0.010	0.000
AF	0.030	0.011	2.847	**	0.000	0.025
age	−0.082	0.023	−3.551	***	0.016	
VS	0.002	0.025	0.077		0.021	
SNP	−0.009	0.027	−0.340		0.025	
SNP × VS	−0.014	0.024	−0.584		0.000	
age × CB	0.010	0.010	0.964		0.000	
age × PS	−0.013	0.007	−1.819		0.000	
age × AF	0.001	0.010	0.108		0.000	
age × VS	−0.057	0.027	−2.085	*	0.000	
age × SNP	0.074	0.026	2.896	**	0.000	
age × VS × SNP	0.033	0.022	1.485		0.000	

*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.

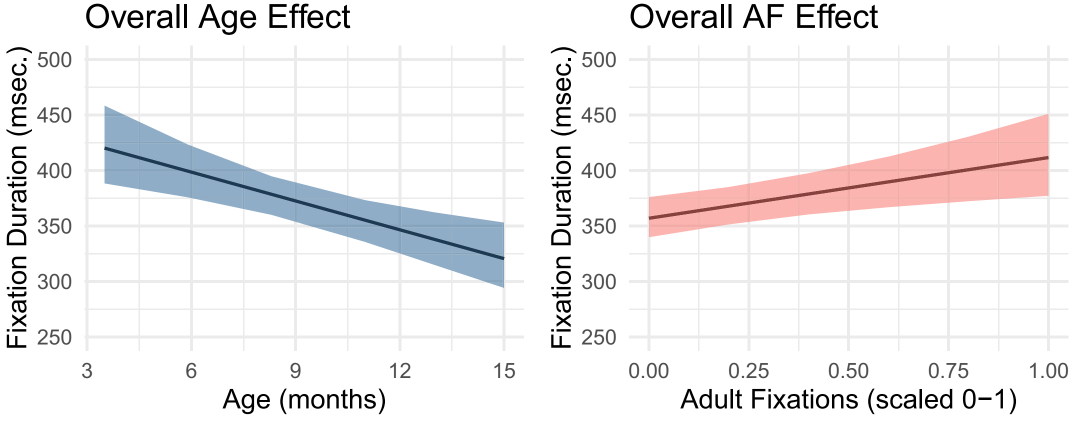

FIGURE 6 Plot of the main effect of age (left panel) and adult fixation locations (right panel) on fixation durations. The shaded region corresponds to the 95% confidence intervals, and the plotted effects are estimated using the R package effects (Fox, 2003)

This study provides insight into the *where* and *when* of infants' attention to static real-world scenes. The analysis of attention allocation addresses the *where*, and the fixation duration analysis addresses the *when*. This distinction is typical in the scene perception literature, as there are studies examining fixation locations *or* durations, but almost never both; exceptions are papers by Tatler, Brockmole, and Carpenter (2017) and Einhäuser and Nuthmann (2016). Since the saliency model of Itti and Koch

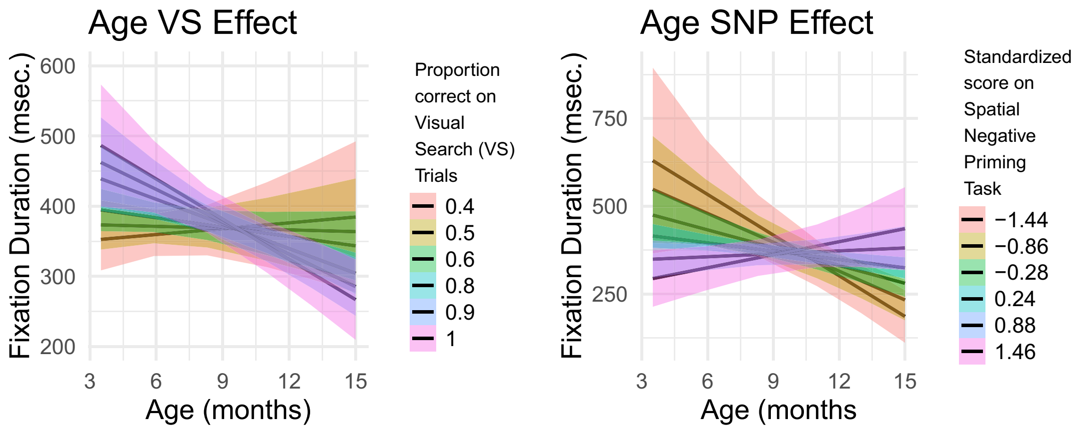

FIGURE 7 Plot of the interaction effects between VS and age (left panel) and SNP and age (right panel) on fixation durations. The shaded region corresponds to the 95% confidence intervals, and the plotted effects are estimated using the R package effects (Fox, 2003). Note that the different levels in the interactions plots are for illustration purposes only and these are continuous variables in the actual analysis

(2000), many newer saliency models have been proposed (for a review see, Borji et al., 2013). These models use scene properties (e.g., contrast, luminance, edges) to predict which parts of scenes are likely to be fixated. As noted previously, Mahdi et al. (2017) compared the performance of different salience maps in explaining infant attention. Similar to the results of the present study, they showed that salience predicts where infants look, although the effects are weaker than in adults. The current study extends these findings by simultaneously controlling for the center bias and assessing attentional control mechanisms. Salience has a unique contribution in driving infants' eye movements and is not entirely an artifact of the co-occurrence with general biases (Tatler & Vincent, 2008, 2009) or adult fixations. Little is known about how salience changes with age, and how different aspects of salience might interact during development (Rogers, Franklin, and Knoblauch (2018), for an analysis of how infants might combine information distributed along multiple, distinct dimensions).

Understanding what factors drive infant attention in static real-world scenes is important because visual attention serves as an important first step for higher order cognition. Regions of scenes that are attended constitute the information that is processed by infants to shape their understanding of the world. How infants perceive, learn, and think about their environment, therefore, is initially shaped by their attention (Johnson, 2011; Johnson et al., 2004). In turn, understanding the cognitive processes in the infant mind requires knowing what information is available to infants for processing. The current study identifies factors that guide infants' attention over complex, static real-world scenes. Selective attention and orienting skills are key elements of executive functioning (Rose, Feldman, Jankowski, & Rossem, 2005) and have predictive value for cognitive abilities at age 11 (Rose, Feldman, & Jankowski, 2012). Here, these factors are shown to influence information processing and attention allocation during infant development. Experiments that examine the interactions between these factors, including social information (Amso et al., 2014; Frank et al., 2014, 2009) and motor activity (cf. Kretch & Adolph, 2015), and their effects on attentional processing are essential to move developmental theories of attention forward.

The current study used an analysis technique that allowed us to incorporate global, local, and individual factors simultaneously. However, the method is still suboptimal, because two important aspects of eye movements were not incorporated into the GLMM analysis technique: time and duration. Incorporating fixation durations in the analysis would make it possible to assess interactions between fixation locations and durations. In a recent study, Einhäuser and Nuthmann (2016) showed that fixation durations on semantically relevant parts of scenes are longer than fixation durations on other parts, in line with our finding that infant fixation durations were longer in regions more frequently fixated by adults. *When* locations are fixated may be an important factor to study as well. For instance, Tatler et al. (2005) showed that fixation locations diverge over time. That is, early fixation locations after scene onset were similar between participants and were likely to fall in the center, but over time, fixation locations diverged. Understanding how attention allocation changes over time is therefore also an important factor to include in future gaze behavior studies.

Analyzing gaze behavior is complex as there are many possible dependent variables. Fixation locations and durations are the most obvious ones, but it is also possible to look at more derivative measures such as the sequence of fixations, the number of fixations, and the amplitude and direction of saccades. Traditional analysis techniques offer few possibilities to include multiple correlated dependent variables, although it is possible using linked linear models. Hohenstein, Matuschek, and Kliegl (2017), for instance, used this technique to study fixation locations and durations simultaneously during reading. In the current study, we also used location information in the model of fixation durations. However, given the complexity of gaze data and the many possible dependent variables, future studies could benefit from using computational models instead of regression techniques. In a computational model, the process hypothesized to generate the data is specified. Behavioral data are then compared with the data of the model to assess whether the hypothesized process is indeed likely to have generated such data. This implies that computational models will make certain predictions about what to expect under different circumstances, allowing for development and testing of more sophisticated theories of infant attention.

The findings of the current study are important to further develop and extend computational models meant to explain fixation durations and locations. Examples of computational models that try to explain fixation durations during scene viewing are the Controlled Random-walk with Inhibition for Saccade Planning (CRISP) model (Nuthmann, Smith, Engbert, & Henderson, 2010) and the Inhibitory Control with Adaptive Timer (ICAT) model (Trukenbrod & Engbert, 2014). These models assume that saccades can be explained by two processes. First, saccades are initiated after random time intervals by an autonomous saccade timer. That is, saccades are triggered independently of what is being processed. Second, processing demands can delay saccade initiation and prolong fixation durations. One of the most consistent findings across the scene-viewing literature is the decrease in fixation durations with age which most likely reflects the decrease in processing time as infants grow older. The computational models such as the ones described above can help to inform which processes underlie this decrease (de Urabain, Nuthmann, Johnson, & Smith, 2017). In addition, these models can be used to inform which processes underlie the interactions with fixation locations, orienting skills, and selecting skills.

5 | CONCLUSION

To sum up, this study describes attention allocation in infants during unconstrained free-viewing of static real-world scenes. We showed that both perceptual salience and adult fixations predict infant attention allocation. In addition, older infants were more likely to attend scene regions more frequently fixated by adults compared to younger infants. Furthermore, we showed that processing time increased at regions frequently fixated by adults, but decreased with age as a function of orienting and selective attention skills. The results and analysis techniques introduced here provide a useful framework to further develop and extend theories of infant visual development when viewing complex, static real-world scenes.

ACKNOWLEDGMENTS

The authors declare no conflicts of interest with regard to the funding source for this study.

ORCID

Daan R. van Renswoude https://orcid.org/0000-0002-2654-081X

REFERENCES

Açık, A., Sarwary, A., Schultze-Kraft, R., Onat, S., & König, P. (2010). Developmental changes in natural viewing behavior: Bottom-up and top-down differences between children, young adults and older adults. *Frontiers in Psychology*, *1*, 207. https://doi.org/10.3389/fpsyg.2010.00207

Althaus, N., & Mareschal, D. (2012). Using saliency maps to separate competing processes in infant visual cognition. *Child Development*, *83*(4), 1122–1128. https://doi.org/10.1111/j.1467-8624.2012.01766.x

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. https://doi.org/10.1371/journal.pone.0085701

Amso, D., & Johnson, S. P. (2005). Selection and inhibition in infancy: Evidence from the spatial negative priming paradigm. *Cognition*, *95*(2), B27–B36. https://doi.org/10.1016/j.cognition.2004.08.006

Amso, D., & Johnson, S. P. (2006). Learning by selection: visual search and object perception in young infants. *Developmental Psychology*, *42*(6), 1236–1245. https://doi.org/10.1037/0012-1649.42.6.1236

Amso, D., & Johnson, S. P. (2008). Development of visual selection in 3-to 9-month-olds: Evidence from saccades to previously ignored locations. *Infancy*, *13*(6), 675–686. https://doi.org/10.1080/15250000802459060

Amso, D., & Scerif, G. (2015). The attentive brain: insights from developmental cognitive neuroscience. *Nature Reviews Neuroscience*, *16*(10), 606–619. https://doi.org/10.1038/nrn4025

Borji, A., Sihite, D. N., & Itti, L. (2013). What stands out in a scene? a study of human explicit saliency judgment. *Vision Research*, *91*, 62–77. https://doi.org/10.1016/j.visres.2013.07.016

Bronson, G. W. (1990). Changes in infants' visual scanning across the 2-to 14-week age period. *Journal of Experimental Child Psychology*, *49*(1), 101–125. https://doi.org/10.1016/0022-0965(90)90051-9

Bronson, G. W. (1994). Infants' transitions toward adult-like scanning. *Child Development*, *65*(5), 1243–1261. https://doi.org/10.1111/j.1467-8624.1994.tb00815.x

Buswell, G. T. (1935). How people look at pictures: A study of the psychology and perception in art. *Journal of Educational Psychology*, *27*, 397–398.

Castelhano, M. S., & Henderson, J. M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *62*(1), 1–14. https://doi.org/10.1037/1196-1961.62.1.1

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(3), 6. https://doi.org/10.1167/9.3.6

Clarke, A. D., & Tatler, B. W. (2014). Deriving an appropriate baseline for describing fixation behaviour. *Vision Research*, *102*, 41–51. https://doi.org/10.1016/j.visres.2014.06.016

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*(1), 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Dannemiller, J. L. (1998). A competition model of exogenous orienting in 3.5-month-old infants. *Journal of Experimental Child Psychology*, *68*(3), 169–201. https://doi.org/10.1006/jecp.1997.2426

Dannemiller, J. L. (2000). Competition in early exogenous orienting between 7 and 21 weeks. *Journal of Experimental Child Psychology*, *76*(4), 253–274. https://doi.org/10.1006/jecp.1999.2551

de Urabain, I. R. S., Nuthmann, A., Johnson, M. H., & Smith, T. J. (2017). Disentangling the mechanisms underlying infant fixation durations in scene perception: A computational account. *Vision Research*, *134*, 43–59. https://doi.org/10.1016/j.visres.2016.10.015

Di Giorgio, E., Turati, C., Altoè, G., & Simion, F. (2012). Face detection in complex visual displays: An eye-tracking study with 3-and 6-month-old infants and adults. *Journal of Experimental Child Psychology*, *113*(1), 66–77. https://doi.org/10.1016/j.jecp.2012.04.012

Einhäuser, W., & Nuthmann, A. (2016). Salient in space, salient in time: Fixation probability predicts fixation duration during natural scene viewing. *Journal of Vision*, *16*(11), 13. https://doi.org/10.1167/16.11.13

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14), 18–18. https://doi.org/10.1167/8.14.18

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, *8*(3), 3–3. https://doi.org/10.1167/8.3.3

Foulsham, T., & Kingstone, A. (2010). Asymmetries in the direction of saccades during perception of scenes and fractals: Effects of image type and image features. *Vision Research*, *50*(8), 779–795. https://doi.org/10.1016/j.visres.2010.01.019

Foulsham, T., & Kingstone, A. (2012). Modelling the influence of central and peripheral information on saccade biases in gaze-contingent scene viewing. *Visual Cognition*, *20*(4–5), 546–579. https://doi.org/10.1080/13506285.2012.680934

Foulsham, T., & Kingstone, A. (2013). Optimal and preferred eye landing positions in objects and scenes. *The Quarterly Journal of Experimental Psychology*, *66*(9), 1707–1728. https://doi.org/10.1080/17470218.2012.762798

Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research*, *48*(17), 1777–1790. https://doi.org/10.1016/j.visres.2008.05.018

Foulsham, T., Teszka, R., & Kingstone, A. (2011). Saccade control in natural images is shaped by the information visible at fixation: Evidence from asymmetric gaze-contingent windows. *Attention, Perception, & Psychophysics*, *73*(1), 266–283. https://doi.org/10.3758/s13414-010-0014-5

Fox, J. (2003). Effect displays in r for generalised linear models. *Journal of Statistical Software*, *8*(15), 1–27. https://doi.org/10.18637/jss.v008.i15

Franchak, J. M., Heeger, D. J., Hasson, U., & Adolph, K. E. (2016). Free viewing gaze behavior in infants and adults. *Infancy*, *21*(3), 262–287. https://doi.org/10.1111/infa.12119

Frank, M. C., Amso, D., & Johnson, S. P. (2014). Visual search and attention to faces during early infancy. *Journal of Experimental Child Psychology*, *118*, 13–26. https://doi.org/10.1016/j.jecp.2013.08.012

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*(2), 160–170. https://doi.org/10.1016/j.cognition.2008.11.010

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the development of social attention using free-viewing. *Infancy*, *17*(4), 355–375. https://doi.org/10.1111/j.1532-7078.2011.00086.x

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy*, *14*(5), 550–562. https://doi.org/10.1080/15250000903144199

Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology*, *4*, 527. https://doi.org/10.3389/fpsyg.2013.00527

Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye tracking in infancy research. *Developmental Neuropsychology*, *35*(1), 1–19. https://doi.org/10.1080/87565640903325758

Helo, A., Pannasch, S., Sirri, L., & Rämä, P. (2014). The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research*, *103*, 83–91. https://doi.org/10.1016/j.visres.2014.08.006

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. https://doi.org/10.1017/S0952523816000110

Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences*, *21*(1), 15–23. https://doi.org/10.1016/j.tics.2016.11.003

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), 10–10. https://doi.org/10.1167/18.6.10

Henderson, J. M., & Luke, S. G. (2014). Stable individual differences in saccadic eye movements during reading, pseudoreading, scene viewing, and scene search. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1390–1400. https://doi.org/10.1037/a0036330

Hessels, R. S., Hooge, I. T., & Kemner, C. (2016). An in-depth look at saccadic search in infancy. *Journal of Vision*, *16*(8), 10. https://doi.org/10.1167/16.8.10

Hohenstein, S., Matuschek, H., & Kliegl, R. (2017). Linked linear mixed models: A joint analysis of fixation locations and fixation durations in natural reading. *Psychonomic Bulletin & Review*, *24*(3), 637–651. https://doi.org/10.3758/s13423-016-1138-y

Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy*, *6*(2), 231–255. https://doi.org/10.1207/s15327078in0602_5

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10), 1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7

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*, *20*(11), 1254–1259. https://doi.org/10.1109/34.730558

Johnson, M. H. (1990). Cortical maturation and the development of visual attention in early infancy. *Journal of Cognitive Neuroscience*, *2*(2), 81–95. https://doi.org/10.1162/jocn.1990.2.2.81

Johnson, M. H. (2008). The development of visual attention: A cognitive neuroscience perspective. In M. H. Johnson, Y. Munakata, & R. O. Gilmore (Eds.), *Brain development and cognition*. https://doi.org/10.1002/9780470753507.ch9

Johnson, S. P. (2011). Development of visual perception. *WIREs Cognitive Science*, *2*(5), 515–528. https://doi.org/10.1002/wcs.128

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds. *Infancy*, *6*(2), 185–201. https://doi.org/10.1207/s15327078in0602_3

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). Sun: Top-down saliency using natural statistics. *Visual Cognition*, *17*(6–7), 979–1003. https://doi.org/10.1080/13506280902771138

Kretch, K. S., & Adolph, K. E. (2015). Active vision in passive locomotion: Real-world free viewing in infants and adults. *Developmental Science*, *18*(5), 736–750. https://doi.org/10.1111/desc.12251

Kwon, M.-K., Setoodehnia, M., Baek, J., Luck, S. J., & Oakes, L. M. (2016). The development of visual search in infancy: Attention to faces versus salience. *Developmental Psychology*, *52*(4), 537. https://doi.org/10.1037/dev0000080

Leppänen, J. M. (2016). Using eye tracking to understand infants' attentional bias for faces. *Child Development Perspectives*, *10*(3), 161–165. https://doi.org/10.1111/cdep.12180

Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition*, *68*(3), 293–308. https://doi.org/10.1016/j.bandc.2008.08.019

Mahdi, A., Su, M., Schlesinger, M., & Qin, J. (2017). A comparison study of saliency models for fixation prediction on infants and adults. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(3), 485–498. https://doi.org/10.1109/TCDS.2017.2696439

Nobre, A. C., Gitelman, D., Dias, E., & Mesulam, M. (2000). Covert visual spatial orienting and saccades: Overlapping neural systems. *NeuroImage*, *11*(3), 210–216. https://doi.org/10.1006/nimg.2000.0539

Nuthmann, A., & Einhäuser, W. (2015). A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences*, *1339*(1), 82–96. https://doi.org/10.1111/nyas.12705

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8), 20. https://doi.org/10.1167/10.8.20

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). Crisp: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382–405. https://doi.org/10.1037/a0018924

Nyström, M., & Holmqvist, K. (2008). Semantic override of low-level features in image viewing–both initially and overall. *Journal of Eye Movement Research*, *2*(2), 1–11. https://doi.org/10.16910/jemr.2.2.2

Papageorgiou, K. A., Smith, T. J., Wu, R., Johnson, M. H., Kirkham, N. Z., & Ronald, A. (2014). Individual differences in infant fixation duration relate to attention and behavioral control in childhood. *Psychological Science*, *25*(7), 1371–1379. https://doi.org/10.1177/0956797614531295

R Core Team. (2017). *R: A language and environment for statistical computing [Computer software manual]*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Reynolds, G. D., Zhang, D., & Guy, M. W. (2013). Infant attention to dynamic audiovisual stimuli: Look duration from 3 to 9 months of age. *Infancy*, *18*(4), 554–577. https://doi.org/10.1111/j.1532-7078.2012.00134.x

Risko, E. F., Anderson, N. C., Lanthier, S., & Kingstone, A. (2012). Curious eyes: Individual differences in personality predict eye movement behavior in scene-viewing. *Cognition*, *122*(1), 86–90. https://doi.org/10.1016/j.cognition.2011.08.014

Rogers, M., Franklin, A., & Knoblauch, K. (2018). A novel method to investigate how dimensions interact to inform perceptual salience in infancy. *Infancy*, *23*(6), 833–856. https://doi.org/10.1111/infa.12260

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2012). Implications of infant cognition for executive functions at age 11. *Psychological Science*, *23*(11), 1345–1355. https://doi.org/10.1177/0956797612444902

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Rossem, R. (2005). Pathways from prematurity and infant abilities to later cognition. *Child Development*, *76*(6), 1172–1184. https://doi.org/10.1111/j.1467-8624.2005.00843.x

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). Labelme: A database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*(1–3), 157–173. https://doi.org/10.1007/s11263-007-0090-8

Slater, A., Morison, V., Somers, M., Mattock, A., Brown, E., & Taylor, D. (1990). Newborn and older infants' perception of partly occluded objects. *Infant Behavior and Development*, *13*(1), 33–49. https://doi.org/10.1016/0163-6383(90)90004-R

Stoll, J., Thrun, M., Nuthmann, A., & Einhäuser, W. (2015). Overt attention in natural scenes: objects dominate features. *Vision Research*, *107*, 36–48. https://doi.org/10.1016/j.visres.2014.11.006

Tatler, B. W. (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. https://doi.org/10.1167/7.14.4

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, *45*(5), 643–659. https://doi.org/10.1016/j.visres.2004.09.017

Tatler, B. W., Brockmole, J. R., & Carpenter, R. (2017). Latest: A model of saccadic decisions in space and time. *Psychological Review*, *124*(3), 267–300. https://doi.org/10.1037/rev0000054

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 5–5. https://doi.org/10.1167/11.5.5

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 1–18. https://doi.org/10.16910/jemr.2.2.5

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*(6–7), 1029–1054. https://doi.org/10.1080/13506280902764539

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (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. https://doi.org/10.1037/0033-295X.113.4.766

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Trukenbrod, H. A., & Engbert, R. (2014). Icat: A computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*, *21*(4), 907–934. https://doi.org/10.3758/s13423-013-0575-0

Unema, P. J., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, *12*, 473–494. https://doi.org/10.1080/13506280444000409

van Renswoude, D. R., Johnson, S. P., Raijmakers, M. E., & Visser, I. (2016). Do infants have the horizontal bias? *Infant Behavior and Development*, *44*, 38–48. https://doi.org/10.1016/j.infbeh.2016.05.005

van Renswoude, D. R., Raijmakers, M. E., Koornneef, A., Johnson, S. P., Hunnius, S., & Visser, I. (2018). Gazepath: An eye-tracking analysis tool that accounts for individual differences and data quality. *Behavior Research Methods*, *50*(2), 834–852. https://doi.org/10.3758/s13428-017-0909-3

van Renswoude, D. R., van den Berg, L., Raijmakers, M. E., & Visser, I. (2019). Infants' center bias in free viewing of real-world scenes. *Vision Research*, *154*, 44–53. https://doi.org/10.1016/j.visres.2018.10.003

Velichkovsky, B. M., Dornhoefer, S. M., Pannasch, S., & Unema, P. J. (2000). *Visual fixations and level of attentional processing*. *Proceedings of the 2000 symposium on eye tracking research & applications* (pp. 79–85). https://doi.org/10.1145/355017.355029

Wang, H.-C., & Pomplun, M. (2012). The attraction of visual attention to texts in real-world scenes. *Journal of Vision*, *12*(6), 26–26. https://doi.org/10.1167/12.6.26

Wass, S. V., 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. https://doi.org/10.1111/infa.12055View

Wass, S. V., Jones, E. J., Gliga, T., Smith, T. J., Charman, T., Johnson, M. H. (2015). Shorter spontaneous fixation durations in infants with later emerging autism. *Scientific Reports*, *5*, 8284. https://doi.org/10.1038/srep08284

Wass, S. V., & Smith, T. J. (2014). Individual differences in infant oculomotor behavior during the viewing of complex naturalistic scenes. *Infancy*, *19*(4), 352–384. https://doi.org/10.1111/infa.12049

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, *45*(1), 229–250. https://doi.org/10.3758/s13428-012-0245-6

Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of Vision*, *14*(1), 28–28. https://doi.org/10.1167/14.1.28

Yarbus, A. L. (1967). *Eye movements during perception of complex objects. In eye movements and vision* (pp. 171–211). New York: Plenum Press.

How to cite this article: van Renswoude DR, Visser I, Raijmakers MEJ, Tsang T, Johnson SP. Real-world scene perception in infants: What factors guide attention allocation? *Infancy*. 2019;24:693–717. https://doi.org/10.1111/infa.12308