



PAPER

Active vision in passive locomotion: real-world free viewing in infants and adults

Kari S. Kretch and Karen E. Adolph

Department of Psychology, New York University, USA

Abstract

Visual exploration in infants and adults has been studied using two very different paradigms: free viewing of flat screen displays in desk-mounted eye-tracking studies and real-world visual guidance of action in head-mounted eye-tracking studies. To test whether classic findings from screen-based studies generalize to real-world visual exploration and to compare natural visual exploration in infants and adults, we tested observers in a new paradigm that combines critical aspects of both previous techniques: free viewing during real-world visual exploration. Mothers and their 9-month-old infants wore head-mounted eye trackers while mothers carried their infants in a forward-facing infant carrier through a series of indoor hallways. Demands for visual guidance of action were minimal in mothers and absent for infants, so both engaged in free viewing while moving through the environment. Similar to screen-based studies, during free viewing in the real world low-level saliency was related to gaze direction. In contrast to screen-based studies, only infants – not adults – were biased to look at people, participants of both ages did not show a classic center bias, and mothers and infants did not display high levels of inter-observer consistency. Results indicate that several aspects of visual exploration of a flat screen display do not generalize to visual exploration in the real world.

Research highlights

- This study bridges the gap between two disparate research paradigms – free viewing using desk-mounted eye trackers and visual guidance of action using head-mounted eye trackers – to examine factors driving visual exploration in infants and adults in the real world.
- Similar to screen-based studies, during free viewing in the real world low-level saliency was related to gaze direction for both infants and adults.
- In contrast to screen-based studies, only infants – not adults – were biased to look at people, participants of both ages did not show a classic center bias, and infants and adults did not display high levels of inter-observer consistency.
- Results suggest that several aspects of visual exploration of a flat screen display do not generalize to visual exploration while moving through a real-world environment.

Introduction

What we see depends on where we look. Visual acuity diminishes in the periphery of the visual field, so eye movements ensure that the high-resolution fovea has access to areas of interest. Vision, thus, involves active selection of information. How do we decide where to point our eyes? And how does visual exploration change over development?

Since the pioneering work of Buswell (1935) and Yarbus (1967), researchers have examined factors that influence natural visual exploration using a *free-viewing* paradigm: Observers look at visual displays on a computer monitor while a stationary, desk-mounted eye tracker records their spontaneous eye movements. The paradigm is simple enough to allow testing of infants (Frank, Vul & Johnson, 2009), people with disabilities (Klin, Jones, Schultz, Volkmar & Cohen, 2002), and animals (Shepherd, Steckenfinger, Hasson & Ghazanfar, 2010). Moreover, the resulting data – even from only a few minutes of free viewing – provide a rich

Address for correspondence: Karen E. Adolph, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA; e-mail: karen.adolph@nyu.edu

source of information about the temporal and spatial properties of spontaneous visual exploration.

Although ostensibly the goal of the free-viewing paradigm is to characterize natural patterns of visual exploration, displays presented to viewers vary widely in their level of realism: objects or disembodied faces on a blank background, static photographs, Hollywood films, professionally produced cartoons, real-world videos collected with head-mounted cameras, and so on. Moreover, the ultimate question of interest is not where observers direct their gaze on computer displays. No matter how realistically a video *depicts* the real world, looking at a two-dimensional depiction is not necessarily the same as visual exploration in the real, three-dimensional world.

Recent advances in head-mounted eye tracking allow researchers to record moment-to-moment direction of gaze in adults, children, and infants as they move freely through the environment and interact with objects, surfaces, and people. In a freely mobile observer, vision serves to guide actions such as posture, locomotion, and reaching. Accordingly, previous work using head-mounted eye trackers focused on *visual guidance of actions* such as steering a car around a bend (Land & Lee, 1994), reaching for objects (Franchak, Kretch, Soska & Adolph, 2011; Pelz, Hayhoe & Loeber, 2001), navigating obstacles (Franchak & Adolph, 2010; Patla & Vickers, 1997), fixing a cup of tea or sandwich (Hayhoe, Shrivastava, Mruczek & Pelz, 2003; Land, Mennie & Rusted, 1999), or hitting a cricket ball (Land & McLeod, 2000). In such cases, looking is tightly linked to ongoing actions and goals.

But guiding action is not the only role of vision in real-world environments. Free viewing – visual exploration of places, objects, people, and events – is concurrent with visual guidance of action. While strolling along a city block, we use vision to steer between pedestrians, step over a puddle or up onto a curb, and avoid getting hit by cars. But we also use vision to admire the display in a store window, inspect our own reflection in the window, survey a building façade, and stare at a teenager with pink hair. Free viewing is especially important in infancy; visual exploration is the earliest developing action system and is infants' primary means for learning about the world. Before infants can reach for objects or walk across a room, they gain access to the world through looking (Gibson, 1988). And even after other action systems become available, infants are physically restrained for much of their day and exploration of the environment is limited to vision. Unfortunately, due to methodological limitations of previous studies, researchers know little about the factors that drive infants' visual exploration in the real world.

Desk-mounted eye-tracking and head-mounted eye-tracking methods each have benefits and shortcomings. With a desk-mounted eye tracker and a computer display, researchers have complete control over the stimulus, and can compare looking patterns across participants and ages. But findings may not generalize to visual exploration in the real world. Conversely, with a head-mounted eye tracker, researchers can study real-world vision and allow observers to select visual information with head and body as well as eye movements. But to the extent that participants go to different places, face in different directions, and engage in different actions, the stimuli are uncontrolled, making comparisons across observers difficult. Combining the benefits of the two approaches seems necessary.

Spatial biases

A robust finding from desk-mounted eye tracking is a bias for observers to point their gaze toward the center of the display ('t Hart, Vockeroth, Schumann, Bartl, Schneider, König & Einhauser, 2009; Buswell, 1935; Dorr, Martinetz, Gegenfurtner & Barth, 2010; Frank *et al.*, 2009; Parkhurst, Law & Niebur, 2002; Tatler, 2007; Tatler, Baddeley & Gilchrist, 2005; Tseng, Carmi, Cameron, Munoz & Itti, 2009). Possibly, this *center bias* arises as an artifact of 'photographer bias' – the fact that interesting objects are often positioned in the center of a scene (Tseng *et al.*, 2009; cf. Tatler, 2007). The center bias could also reflect a strategy for maximizing the information obtained from the scene – looking toward the center of the display rather than a side minimizes the distance of any part of the screen from the fovea ('t Hart *et al.*, 2009; Tatler, 2007; Tseng *et al.*, 2009). In studies where observers' heads are fixed on a chin rest and pointed toward the center of the display, 'orbital reserve', the preference to return the eye to a comfortable position in the center of the orbit, may contribute to the center bias (Fuller, 1996; Pare & Munoz, 2001; Tatler, 2007; Tseng *et al.*, 2009).

The notion of a center bias takes on a different meaning in head-mounted eye-tracking studies. There is no fixed center of a display because the observer is completely surrounded by visual information and the field of view changes every time the observer's head moves; the display at every moment is head-centered. For example, the wall at the end of a hallway is not the center of the display if observers don't keep their heads pointed straight ahead, and indeed, walkers rarely point their gaze at the wall at the end of a hallway (Turano, Geruschat & Baker, 2003). However, gaze does tend to cluster in the middle of the head-centered field of view, because people point their heads where they want to look. Direct comparisons between head-mounted and desk-mounted methods suggest that a head-centered center bias is even stronger than a

display-centered center bias (Foulsham, Walker & Kingstone, 2011). Moreover, in the real world, spatial distribution of gaze depends on the task and vantage point. For example, infants and parents both show a center bias while engaged in a table-top task with objects, but the center bias is more pronounced for parents (Bambach, Crandall & Yu, 2013), possibly because their greater height puts objects at a greater distance from their eyes, allowing the entire object to be explored from the center of their field of view.

Inter-observer consistency

Another highly replicable finding from desk-mounted eye tracking is that different people tend to look in the same place at the same time while watching moving images on a video ('t Hart *et al.*, 2009; Dorr *et al.*, 2010; Goldstein, Woods & Peli, 2007; Hasson, Yang, Vallines, Heeger & Rubin, 2008; Kirkorian, Anderson & Keen, 2012; Marchant, Raybould, Renshaw & Stevens, 2009; Shepherd *et al.*, 2010; Tosi, Mecacci & Pasquali, 1997; X.H. Wang, Freeman, Merriam, Hasson & Heeger, 2012). Inter-observer consistency is higher for professionally produced Hollywood films than for videos of natural scenes (Dorr *et al.*, 2010; Hasson *et al.*, 2008) and consistency increases from infancy to adulthood (Franchak, Heeger, Hasson & Adolph, 2013; Frank *et al.*, 2009; Kirkorian *et al.*, 2012).

Do parts of the natural environment also draw observers' attention in similar ways? This question is difficult to answer because in head-mounted eye-tracking studies, observers move their heads and bodies in different directions at different times. Thus, the scene is not perfectly synchronized between observers, and absolute gaze coordinates correspond to different parts of the scene for each observer. Instead of comparing the sequence of absolute gaze coordinates as researchers do for desk-mounted displays, researchers must score the targets of fixations and compare them categorically over time.

When pairs of observers are engaged in a mutual task or conversation, their looking is highly coordinated. For example, while playing together with objects on a table, infants and parents look at the same object at the same time at greater than chance levels (Yu & Smith, 2013). But in the absence of a mutual task and communicative cues, gaze is coordinated only at chance levels between observers' view as they move through the world wearing a head-mounted eye tracker and observers' free viewing of the videos from the head-mounted eye tracker (Foulsham *et al.*, 2011).

Bottom-up vs. top-down influences

Researchers have focused on two types of influences to account for inter-observer consistency in free-viewing studies with desk-mounted eye trackers: bottom-up,

stimulus-driven factors such as image salience, and top-down, cognitive factors such as goals or social relevance. Image salience refers to low-level features represented early in the visual stream (luminance contrast, orientation changes, motion, etc.) that 'capture' visual attention regardless of content. Formal computational models (e.g. Itti & Koch, 2000) show that salience contributes significantly to observers' fixations during free viewing of static images (Foulsham & Underwood, 2008; Parkhurst *et al.*, 2002; Peters, Iyer, Itti & Koch, 2005; Tatler *et al.*, 2005) and dynamic videos ('t Hart *et al.*, 2009; Itti, 2005; Le Meur, Le Callet & Barba, 2007; Mital, Smith, Hill & Henderson, 2011).

However, in free-viewing situations, low-level salience is often confounded with higher-level factors such as objects (Einhauser, Spain & Perona, 2008b; Elzary & Itti, 2008) or semantically informative portions of a scene (Henderson, Brockmole, Castelhana & Mack, 2007). In fact, other observers' gaze patterns tend to be better predictors of gaze than saliency models (Einhauser *et al.*, 2008b; Henderson *et al.*, 2007; Shepherd *et al.*, 2010), suggesting that inter-observer consistency is driven by more than low-level salience. Moreover, in visual search tasks, image salience has only a minimal effect on observers' gaze (Einhauser, Rutishauser & Koch, 2008a; Foulsham & Underwood, 2007; Henderson *et al.*, 2007; Underwood, Foulsham, van Loon & Underwood, 2005), suggesting that top-down task demands can override the effects of bottom-up salience.

One particularly influential top-down influence on free viewing is the social relevance of stimuli. Adults preferentially look at people and faces in still images (Cerf, Harel, Einhauser & Koch, 2007; Yarbus, 1967), Hollywood movies (Klin *et al.*, 2002; Shepherd *et al.*, 2010), and natural videos (Foulsham, Cheng, Tracy, Henrich & Kingstone, 2010). By 4 months of age, infants preferentially look at faces over objects in static arrays (DeNicola, Holt, Lambert & Cashion, 2013; Di Giorgio, Turati, Altoe & Simion, 2012; Gliga, Elsabbagh, Andravizou & Johnson, 2009; Libertus & Needham, 2011; Schietecatte, Roeyers & Warreyn, 2011). Nonetheless, face preference is less pronounced in infants and increases over infancy and childhood (Amso, Haas & Markant, 2014; Aslin, 2009; Frank *et al.*, 2009).

Thus, social attention may override or modulate the influence of bottom-up salience on free viewing. Six-month-old infants preferentially fixate images of faces and body parts over images of objects, even when objects are more salient (Gluckman & Johnson, 2013). When watching cartoons with social content, 3-month-olds' fixations are best predicted by saliency, whereas fixations of 9-month-olds and adults are better predicted by the locations of faces (Frank *et al.*, 2009). Similarly, for adults, adding a face detection channel to a standard

saliency model consistently improves the predictive ability of the model for fixations of natural images (Cerf *et al.*, 2007). And salience may modulate face preference: Children and adults (but not infants in their first year) are more likely to orient to faces if they are also highly salient (Amso *et al.*, 2014).

However, effects of salience and social attention may be very different in the real world. Adult participants are less likely to fixate people in the real world than videos of people on a computer screen (Laidlaw, Foulsham, Kuhn & Kingstone, 2011). And infants in natural play situations rarely fixate their parents' faces, preferring to look at objects and locomotor obstacles (Franchak *et al.*, 2011; Yu & Smith, 2013). Because head-mounted eye-tracking studies examine gaze during specific tasks, salience typically has a negligible influence on looking (Rothkopf, Ballard & Hayhoe, 2007; Turano *et al.*, 2003), leading some researchers to suggest that looking is based on 'world salience' or behavioral relevance of stimuli, rather than image salience (Cristino & Baddeley, 2009; Tatler, Hayhoe, Land & Ballard, 2011). However, in table-top play, salience is a reasonably good predictor of gaze, and is more influential for adults than infants (Bambach *et al.*, 2013).

Current study

The current study was designed to bridge the gap between two disparate research paradigms: free viewing as typically studied with desk-mounted eye tracking and visual guidance of action while wearing a head-mounted eye tracker. We used head-mounted eye trackers to characterize influences on free viewing in infants and adults in a real-world setting.

Infants and adults generally do not share the same viewpoint, movements, or immediate goals. How then can we directly compare their visual exploration in a natural environment? We capitalized on a common practice – caregivers carrying infants in a forward-facing baby carrier – to equate visual scenes between infant and adult participants. Mothers and infants were thus physically yoked and experienced the same environment and the same body movements from approximately the same viewpoint. In contrast to previous work where adults made sandwiches, steered cars or navigated obstacles, task demands for mothers were minimal – to keep balance while wandering through hallways in the psychology department – and infants were in a free-viewing situation. We asked mothers to walk freely through an indoor space to provide a continually changing, three-dimensional, real-world visual scene to themselves and their infants.

We aimed to examine whether four key findings from free-viewing studies with desk-mounted eye trackers generalize to infants' and adults' visual exploration in

the real world. First, we examined the spatial distribution of eye gaze in environment-centered and head-centered coordinates to ask whether a center bias occurs in natural environments and whether it differs between infants and adults. If the center bias is an artifact of screen-based displays, then participants may direct their gaze equally to all parts of the environment. However, if participants coordinate head and eye movements, we should also see a bias for gaze to cluster in the middle of the head-centered field of view (Bambach *et al.*, 2013; Foulsham *et al.*, 2011). If previously reported differences between infants and adults (Bambach *et al.*, 2013) are due to different viewpoints, we may see similar clustering of gaze for both infants and mothers.

Second, we took advantage of the new yoking method to examine inter-observer consistency in the real world: We asked whether infants and their mothers looked in the same place at the same time. If inter-observer consistency in screen-based free viewing relies on showing a restricted, perfectly identical scene, then the ability to make independent head movements in a panoramic scene may not produce highly synchronized viewing patterns between infants and mothers.

Of particular interest was how infants' and adults' real-world visual exploration is driven by bottom-up and top-down factors. Thus, our third question concerned the influence of low-level saliency on infants' and mothers' gaze. If the capture of visual attention by visually salient areas is an inherent property of free exploration, then saliency should predict gaze in the real world during relatively task-free conditions. Regarding developmental effects, we expected saliency to be a better predictor of gaze for adults than infants, based on previous findings from both desk-mounted (Frank *et al.*, 2009) and head-mounted (Bambach *et al.*, 2013) methods.

Finally, we investigated the influence of top-down factors by asking how frequently infants and adults direct their gaze to socially relevant stimuli. Whereas screen-based free-viewing studies have revealed a high rate of looking to people and faces that increases with age (Amso *et al.*, 2014; Aslin, 2009; Frank *et al.*, 2009), head-mounted eye-tracking studies suggest that looking at other people may be infrequent in the real world for both infants and adults (Franchak *et al.*, 2011; Laidlaw *et al.*, 2011; Yu & Smith, 2013).

Method

Participants

Fifteen 9-month-old infants (range = 8.55–9.01 months, $M = 8.81$; 7 boys) and their mothers (age range = 23.62–

41.07 years, $M = 32.52$) participated. Families were recruited from the maternity wards of local hospitals and received small souvenirs for participation. Families were mostly white and middle class. All infants had experience in strollers and all but one had experience in an infant carrier. Data from five additional dyads were excluded because the camcorders stopped recording ($n = 3$), the infant would not look to the calibration targets ($n = 1$), or the parent did not follow instructions ($n = 1$).

Head-mounted eye trackers

Both infants and mothers wore Positive Science (positivescience.com) head-mounted eye trackers (Figure 1A). The eye trackers consisted of two small cameras mounted on lightweight headgear: the *scene camera* pointed outward from above the right eye to record the

scene, and the *eye camera* pointed inward to record the participant's eye. The scene camera had a field of view of 54.4 degrees horizontal \times 42.2 degrees vertical. The adult headgear was a set of eyeglass frames (Franchak & Adolph, 2010) and the infant headgear was a flexible band attached to the front of a spandex cap (Franchak *et al.*, 2011).

Videos from both cameras were collected at 30 fps and were fed to small camcorders in a backpack worn by the mother. After the session, we used Yarbus software (from Positive Science) to calculate the point of gaze (spatial accuracy ~ 2 degrees) within the scene camera video using estimates of the center of the pupil and the corneal reflection from the eye camera video. The software produced two sources of data: a series of x and y coordinates of the point of gaze within the scene camera video, and a new video with a circular cursor (diameter = 1.5 degrees) overlaid on each frame at the

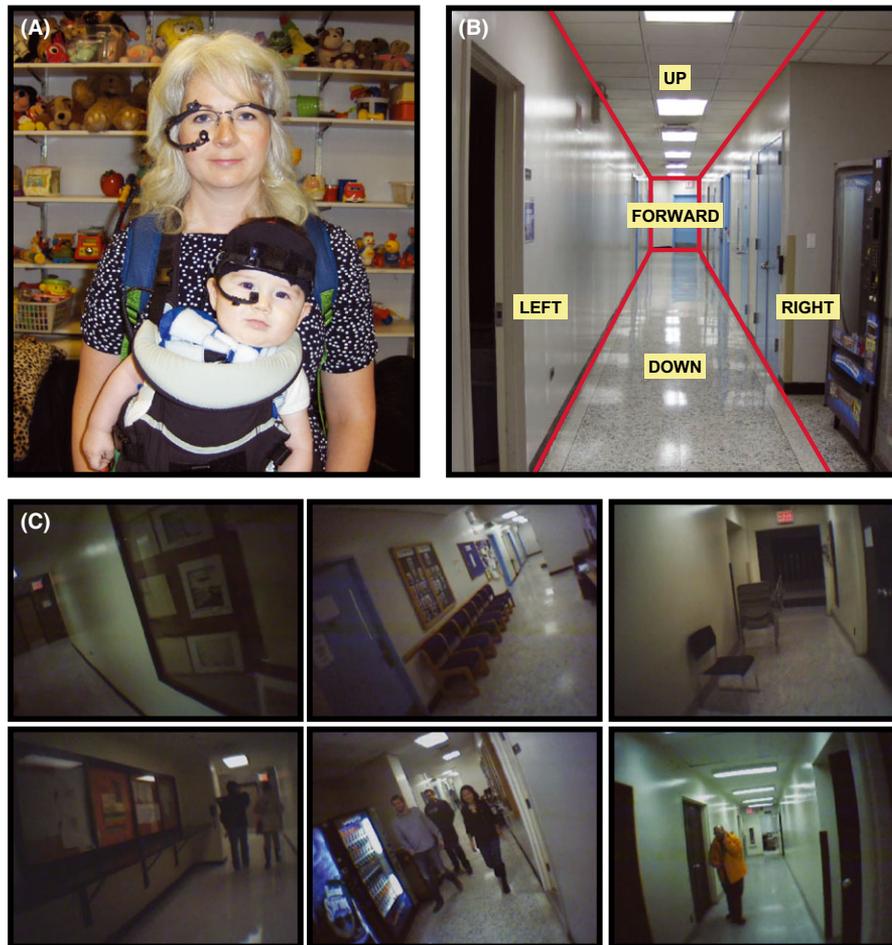


Figure 1 (A) Infant and mother wearing head-mounted eye trackers. (B) Gaze direction coding scheme illustrated for Hallway 3. For each video frame, coders scored which of the five regions contained the gaze cursor. If the gaze cursor fell on a person, coders categorized the person as standing to the right or left side of the hallway. (C) Example video frames from the scene cameras of infants and mothers during the study.

point of gaze. Note that the point of gaze could be any point within the scene camera video. Changes in the scene camera video represent head movements, and changes in the point of gaze within the scene camera video represent eye movements.

Procedure

First, mothers were fitted with the headgear, infant carrier (Infantino Easy Rider), and backpack. The experimenter collected calibration data by asking mothers to look at nine points on a display board. Mothers were asked to hold their heads still for calibration so that the points were spread over the entire scene camera field of view. Then, the experimenter placed the hat and headgear on the infants while an assistant distracted infants with toys. For calibration, the assistant called infants' attention to locations on another display board by presenting noisy toys in cutout windows. The number of calibration points varied between infants (minimum = 4 points), but points were placed so that they were spread over the entire scene camera field of view. Both mothers and infants viewed calibration boards from a distance of approximately 4 ft.

After calibration, the experimenter secured infants in the carrier, and accompanied participants to the first floor of the building. Mothers were told to walk around at a comfortable pace, and that they were free to go anywhere or look at anything they chose but that they should not speak to their infants or point out where to look. The experimenter followed behind mothers with a video camera to capture their movements in the hallway. The third-person video and both first-person gaze videos were synchronized offline for coding.

The first floor of the building consisted of three hallways connected by open vestibules. We selected the three hallways as the target areas to facilitate video coding of gaze direction. Hallway 1 was 24.2 m long \times 2.4 m wide and contained a large display case, a shelf, a pay phone, and several posted signs. Hallway 2 was 32 m long \times 1.7 m wide and contained several classroom doors, vending machines, a table, a bulletin board, a row of chairs, and signs. Hallway 3 was 12.2 m long \times 1.8 m wide and contained two classroom doors, a drinking fountain, a chair, and signs. The experiment was conducted during regular weekday hours and other people frequently walked through the hallways. Figure 1B shows a portion of hallway 3 and Figure 1C shows representative frames from the scene camera video of several participants from all three hallways. Mothers mostly walked continuously without stopping, but walking speed varied between participants (e.g. time to walk down hallway 1 ranged from 17.10 - 53.06 s), and mothers sometimes slowed

down and turned their bodies to the side to look at objects on the walls. The sessions produced six walking sequences per dyad because mothers traveled down each hallway and back; however, one mother did not visit hallway 3, resulting in 88 sequences in total.

Video coding

Although some of our research questions could be answered from the raw x,y coordinates of gaze location within the scene camera video, other questions required us to know the targets of gaze in the world. Thus, we coded the processed videos frame by frame for gaze targets. We scored the location of the gaze cursor to obtain data about where participants pointed their eyes, irrespective of where they pointed their heads. We first coded gaze direction based on the geometry of the hallways (Figure 1B): For each frame, coders scored whether the gaze cursor was on the ceiling, the left wall, the right wall, the wall at the end of the hallway, or the floor (for mothers, looking at the infant was scored as looking down).

In addition, coders noted when participants looked at people and when they looked at people's faces (defined as any part of the head). For inclusion in spatial analyses, looks at people were classified as left or right based on whether the person being fixated was more to the left or right side of the hallway. Because people were not always present, coders also scored whether people were visible in the infants' and mothers' scene camera video.

Finally, for the frames where infants and mothers were looking in the same direction (e.g. both looking left), coders scored whether they were looking at the same object, and if so, what the object was. Objects included chairs, baseboards, doors, bulletin boards or display cases, shelves, vending machines, signs, tables, water fountains, blank walls, the floor, or people.

A primary coder scored 100% of the data, and a second coder scored 25–33% of each participant's data for inter-rater reliability. Coders agreed on 96% of frames for gaze direction, 99% of frames for people in the scene video, and 94% of frames for looking at the same object; disagreements were resolved through discussion.

Saliency maps

To determine whether gaze was related to visual saliency, we selected the first pass down each hallway (44 sequences total) and created a 'saliency map' for each video frame based on the classic algorithm of Itti, Koch and Niebur (1998; using J. Harel, *A Saliency Implementation in Matlab*: <http://www.klab.caltech.edu/~harel/share/gbvs.php>). Using a biologically inspired model of

early visual processing, each pixel of each video frame was evaluated for its distinctiveness over five saliency channels: color, intensity, orientation, flicker, and motion. Color, intensity, and orientation channels were computed based on the current frame alone, and motion and flicker channels were computed based on differences between the current frame and the previous frame. For simplicity, the five channels were combined with equal weights into a single saliency value. These overall saliency values were converted to percentile ranks for analysis, so that the most salient pixel in the frame had a value of 100 and the least salient pixel had a value of 0.

Results

Spatial distribution of eye gaze

Head-centered spatial distribution

We first analyzed the spatial distribution of gaze within the head-centered scene to investigate whether gaze was

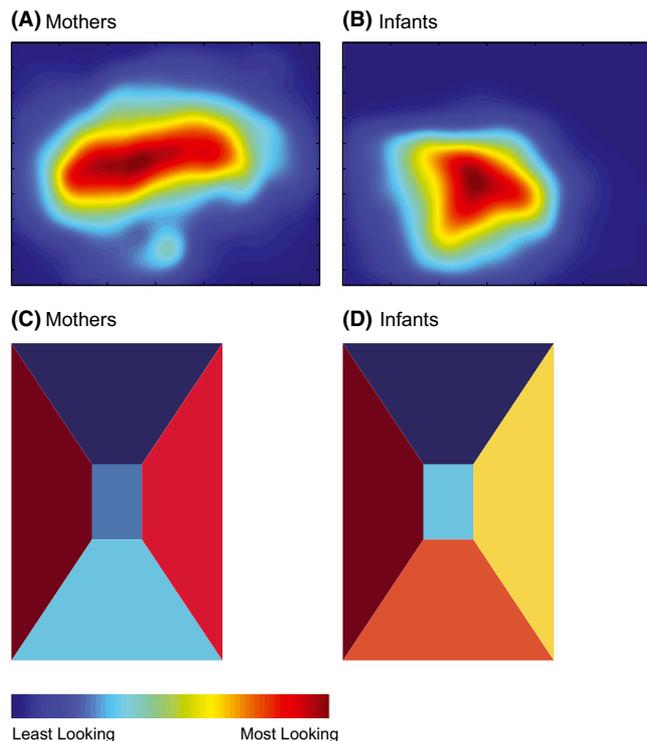


Figure 2 Gaze density heat maps for (A) mothers and (B) infants showing the frequency of gaze at all locations within the scene camera field of view. Gaze density diagrams showing the frequency of gaze for (C) mothers and (D) infants in the five sections of the hallway. Redder shades indicate more looking in that location and bluer shades indicate less looking.

clustered toward the center of the field of view – that is, whether infants and adults pointed their eyes in the same direction as their heads. As shown in Figure 2A–B, gaze was indeed tightly clustered, especially for infants. Note that because the scene camera is placed over the *right* eye, the field of view is shifted toward the right, causing the location of this cluster to be shifted toward the *left* side of the scene. Furthermore, the scene camera on the adult headgear was mounted on eyeglass frames and was lower on the face than the scene camera on the infant headgear, which was mounted on a hat. Thus, the gaze map for infants is shifted toward the bottom left compared to the gaze map for mothers. However, the center of the clusters roughly corresponds to the middle of the head-centered scene. To quantify the extent to which gaze was clustered toward the center vs. spread over the scene, we compared standard deviations of the gaze coordinates in the horizontal and vertical dimensions. Mothers' gaze was more widely distributed in the horizontal dimension than infants' gaze (M standard deviation for mothers = 138.18 pixels/11.75 degrees, M standard deviation for infants = 92.09 pixels/7.83 degrees), $t(14) = 7.61$, $p < .001$, but they did not differ in the vertical dimension (M mothers = 82.40 pixels/7.24 degrees, M infants = 74.71/6.57 degrees), $t(14) = 1.19$, $p = .25$. Thus, infants were more likely to use head movements to direct their gaze from side to side and mostly kept their eyes aligned with their head, whereas mothers sometimes looked to the left and right using only eye movements.

Environment-centered spatial distribution

Because our real-world free-viewing setup allowed participants to move their heads, a head-centered center bias did not necessarily imply a 'display-centered' or environment-centered center bias. In fact, we did not find evidence of a hallway-centered center bias as mothers and infants traveled straight down the hallway. Figure 2C–D depicts the accumulated duration of gaze to the different areas of the hallway using the same color scheme as the head-centered heat maps, with redder shades representing more looking in that direction and bluer shades representing less looking. Looking was clearly not biased toward the center of the hallway; mostly, mothers looked to the walls and down. As shown in Figure 3A, mothers spent more time than infants looking at the side walls, $t(14) = 3.70$, $p = .002$, and infants spent more time than mothers looking down, $t(14) = 2.63$, $p = .02$. Both mothers and infants spent a similarly low proportion of time looking straight ahead ($p = .14$).

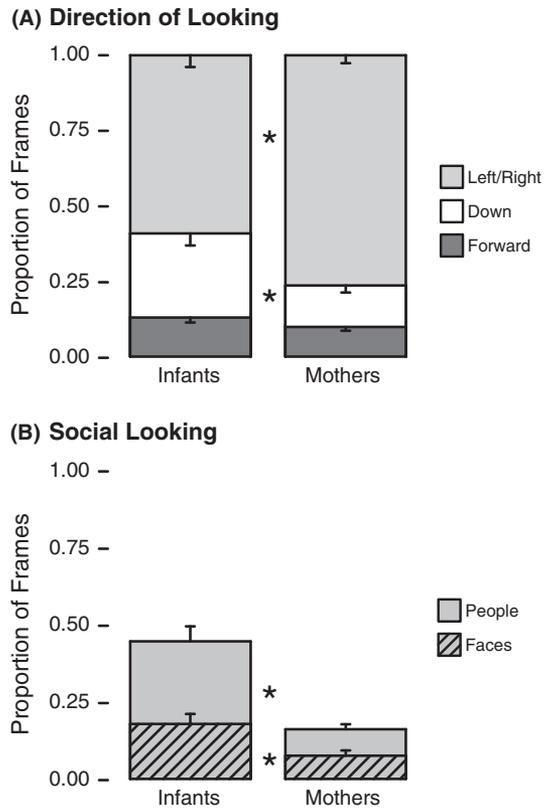


Figure 3 (A) Proportion of video frames (out of the total looking time) in which participants looked at each of the target locations. Looks up at the ceiling were negligible and were included with forward looks. (B) Proportion of video frames (out of the frames where people were visible in the scene camera video) in which participants looked at people and faces. Error bars denote standard errors. Asterisks denote proportions significantly different between mothers and infants ($p < .02$).

Infant–mother synchronization

Although mothers and infants traveled the same path through the same environment with a similar vantage point, their looking patterns could, in principle, differ. In fact, for most video frames, mothers and infants were not looking in the same direction at the same time. Overall, mothers and infants looked in the same direction (i.e. forward, left, right, or down) in $M = 36\%$ of video frames. The proportion of frames looking in the same direction ranged from 21% to 68% between dyads and 1% to 74% between individual hallway sequences. In 60.3% of direction matching episodes, mothers looked in the matching direction before infants, suggesting a tendency for infants' gaze to follow mothers' (binomial $p < .001$).

Although simultaneous looking was relatively infrequent, mothers' and infants' looking patterns were

coordinated at greater than chance levels. To investigate whether mothers' and infants' gaze direction was statistically independent, we examined conditional probabilities for looking in each direction based on where the other member of the dyad was looking (Table 1). This analysis revealed that participants were more likely to be looking in a given direction when the other member of the dyad was also looking there. For example, infants were more likely to be looking to the left if the mother *was* looking to the left than if the mother *was not* looking to the left (and vice versa). This pattern was true for forward, left, and right looks and looks to people. However, infants' and mothers' downward looks were unrelated.

Another way to investigate whether mothers and infants were more coordinated than would be expected from random looking is to compare the direction match rate in actual sequences to the distribution that results from a sample of randomly shuffled sequences. In 47 of 88 sequences (53.4%), the direction match rate was significantly higher than chance; that is, the actual rate was higher than the top .05 of 1000 randomly shuffled sequences.

The data reported in this section thus far reflect the time when mothers and infants were looking in the same broad direction. But the walls of the hallway were large regions of visual space. A dyad could be coded as both looking to the right if, for example, the infant looked at a door up ahead while the mother read a sign as she passed by. More detailed analyses revealed that mothers and infants rarely deployed their visual attention to the same object at the same time. Mothers and infants looked at the exact same object in only $M = 12\%$ of video frames (range 2%–36%). Most frequently, object matching occurred when participants fixated large displays or bulletin boards ($M = 25.7\%$ of the total object matching time). Inspection of the videos suggests that mothers often turned their bodies (and, thus, their infants) to explore these objects of interest. Other objects that elicited matching fixations were the vending machines ($M = 16.4\%$ of object matching time), doors ($M = 13.1\%$), the floor ($M = 13.4\%$), and people ($M = 13.1\%$). Only $M = 4.6\%$ of object matching time consisted of mutual looks to blank areas of the wall, despite these areas taking up a large portion of the field of view.

Social attention

In a subset of video frames, other people walking through the hallways were available as potential gaze targets. Mothers had people in view on $M = 11.7\%$ of frames, and infants had people in view on $M = 12.9\%$ of frames. This small difference was statistically significant, $t(14) = 3.43$, $p = .004$, in part due to the extremely high

Table 1 Conditional probabilities of looking in each direction based on looking direction of other participant

Look direction	P (infant looking mother looking)	P (infant looking mother not looking)	P (mother looking infant looking)	P (mother looking infant not looking)	Odds ratio
Right	.34	.18	.51	.31	3.37*
Left	.43	.29	.49	.35	2.16*
Forward	.26	.11	.19	.09	3.31*
Down	.25	.28	.13	.14	1.05

Note: *Odds ratio significantly different from 1, $p < .05$.

correlation between mothers and their infants in how often people were in view, $r(13) = .99$, $p < .001$, suggesting that infants more than mothers moved their heads to keep people in view.

Infants' gaze was drawn to social stimuli, but mothers' gaze was not. When people were in view, infants looked at them in $M = 45\%$ of video frames, compared with only $M = 17\%$ of frames for mothers, $t(14) = 3.91$, $p = .002$ (Figure 3B). Similarly, infants looked at faces in $M = 18\%$ of frames compared with $M = 8\%$ for mothers, $t(14) = 2.78$, $p = .015$. Social content did not drive synchronization within dyads: Mothers and infants were not more likely to look in the same direction when there were people in view ($p = .82$).

Saliency

Saliency at gaze location

To examine how well low-level saliency predicted gaze for infants and mothers, we compared saliency values (as percentile ranks) at the location of gaze to saliency values at randomly generated gaze points. Saliency values were defined as the maximum value within a circle (radius 1.5 degrees) around the specified location. The average saliency rank at gaze was relatively high ($M = 80.27$ and $M = 80.84$ for infants and mothers, respectively; Figure 4A). In other words, the saliency value at the gaze location was typically higher than 80% of pixels in the frame. For comparison, we generated chance distributions based on random locations using two different procedures: The first set of points was chosen randomly from a uniform distribution over all possible pixels, and the second set was chosen randomly from the actual distribution of gaze locations in the data set, thus adjusting for spatial biases in gaze. Regardless of which criteria were used for the random distribution, 2 (infant vs. mother) \times 2 (gaze vs. random) repeated-measures ANOVAs revealed that saliency at gaze locations was significantly higher than at random locations,

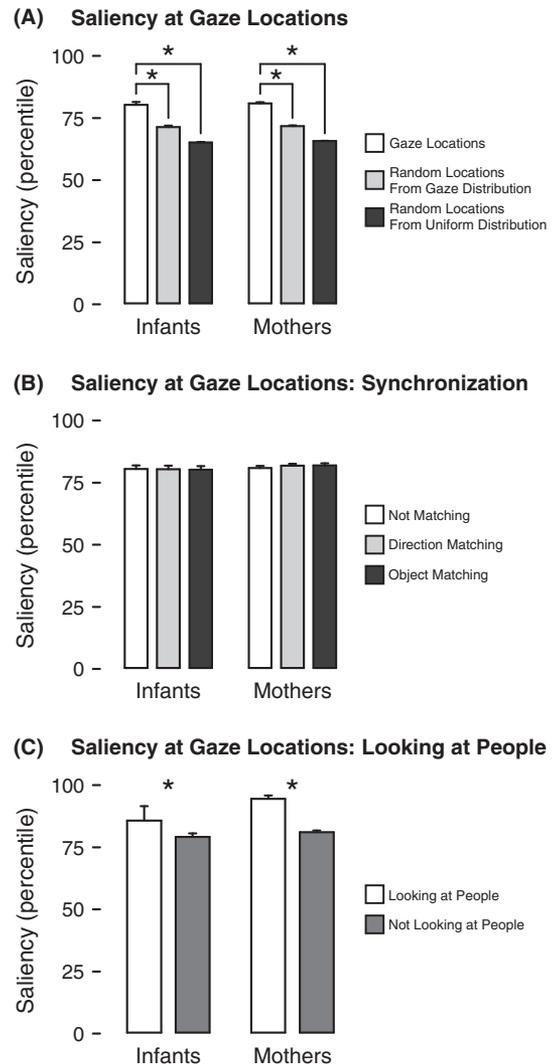


Figure 4 (A) Average saliency value (as percentile of all pixels in the frame) for gaze locations compared with both types of random locations. (B) Average saliency value at gaze location for frames where infants and mothers were not looking in the same direction, were looking in the same direction, or were looking at the same object. (C) Average saliency value at gaze location for frames where participants looked at people and frames where participants looked at other objects. Error bars denote standard errors. Asterisks denote means that were significantly different ($p < .01$).

$F(1, 14) = 595.81$ and 191.58 , $ps < .01$. Gaze locations were equally salient for infants and mothers $F(1, 14) = 0.34$ or 0.63 , $ps > .44$.

However, infants' and mothers' gaze was *not* strongly attracted to the most salient spot in the scene. The location of gaze fell within 1.5 degrees of the most salient pixel in only $M = 1.2\%$ of frames for infants and $M = 1.1\%$ of frames for mothers, $t(14) = 0.54$, $p = .60$.

Although very low, these proportions were significantly higher than those calculated using random points for infants, $ps < .02$, and significantly higher than random points from the uniform distribution for mothers, $p = .01$ (and marginally higher than random points from the mothers' gaze distribution, $p = .07$).

Relation between saliency and synchronization

We hypothesized that areas that were particularly salient might draw simultaneous gaze from both infants and mothers, thus contributing to synchronization. However, we found the location of gaze to be equally salient regardless of whether infants and mothers were looking in the same direction, $F(1, 14) = 0.13$, $p = .72$, and regardless of whether infants and mothers were looking at the same object, $F(1, 14) = 0.30$, $p = .59$ (Figure 4B).

Relation between saliency and social attention

People were particularly salient gaze targets. Three of the dyads did not look at people in the subset of sequences used for saliency computations. For the remaining 12 dyads, saliency at the location of gaze was significantly higher when mothers and infants were looking at people than when they were looking at other objects, $F(1, 11) = 11.49$, $p < .01$ (Figure 4C), suggesting that due to the visual complexity and independent motion of social targets, saliency and social relevance of stimuli may be correlated. The participant (infants vs. mothers) \times target (people vs. other) interaction was not significant, $F(1,11) = 1.51$, $p = .24$.

Discussion

The current study investigated gaze allocation in infants and their mothers using a novel real-world free-viewing paradigm that bridges the gap between previous desk-mounted and head-mounted eye-tracking studies. We found that some classic findings from traditional free-viewing studies based on desk-mounted eye tracking generalize to free viewing in the real world but some do not. Our hybrid approach (free viewing using head-mounted eye tracking) provides new insights into the factors driving spontaneous visual exploration in infants and adults.

Spatial biases in the real world

Orbital reserve (an aversion to moving the eye too far from the center of its orbit) is one proposed explanation

for the classic center bias in screen-based free viewing studies. As in previous head-mounted eye-tracking studies (Bambach *et al.*, 2013; Foulsham *et al.*, 2011), we observed clear evidence of orbital reserve: Both infants and mothers tended to look where their heads were pointing. This tendency was stronger for infants, indicating that infants were more likely to move their heads in the direction they were looking and mothers made more eye movements without corresponding head turns. Because mothers were controlling locomotion, they may have been more reticent to turn their heads away from the direction of motion for long periods of time. This trend is the opposite from that found in object play, where gaze was more clustered in the middle of the head-centered field of view for mothers than infants (Bambach *et al.*, 2013). Together, these findings indicate that the contributions of eye and head movements to gaze direction may be more dependent on viewpoint and task than on age.

In screen-based free-viewing studies, participants are discouraged or physically prevented from moving their heads. Thus, orbital reserve produces a bias to fixate the center of the display. Other proposed contributions to the center bias include the fact that interesting objects tend to be framed in the center of a scene rather than along the edges, and the fact that looking to the center provides the best vantage point to see all parts of the screen. But in the real world, observers are free to move their heads, and the display and objects within it extend 360 degrees around. In the current study, infants and mothers generally looked where their head was pointing, but they did not always point their heads forward. While walking down a hallway, infants and mothers only fixated the wall straight ahead – the central point of the environment – 10–13% of the time. This may seem surprising especially for mothers, who were responsible for maintaining locomotion in the forward direction. However, some models of steering suggest that information from objects off to the side is sufficient for maintaining a constant heading direction (Cutting, Alliprandini & Wang, 2000; R.F. Wang & Cutting, 2004). Similarly, a recent report describing optic flow produced by head and body movements in a subset of these dyads found that neither infants nor mothers directed their gaze close to the focus of expansion, or the visual 'straight ahead' (Raudies, Gilmore, Kretch, Franchak & Adolph, 2012).

Although the most frequent gaze direction for both mothers and infants was to the lateral walls, mothers spent more time than infants looking to the sides and infants spent more time than mothers looking down at the floor. Possibly due to riding in the infant carrier,

infants may have adopted a relaxed head posture that biased gaze toward the floor.

Inter-observer consistency in the real world

Given a similar visual environment, a primary question of interest was whether infants and mothers looked in the same direction at the same time. We found that mothers' and infants' gaze directions were not statistically independent and in many cases were more coordinated than would be expected by chance. In contrast to face-to-face interactions (Yu & Smith, 2013), with the baby carrier, mothers and infants could not see each other's eye gaze direction and mothers did not speak to their infants or point out objects or events using gestures. What, then, accounted for this synchronization in looking? Gaze synchronization did not appear to be driven by low-level saliency or the presence of moving social agents. Possibly, common body movement may have biased infants and mothers to look in the same direction. When mothers turned their bodies to explore something, infants came along for the ride, often filling their visual fields with a similar scene. Infants were more likely to look where mothers were already looking than vice versa, indicating that mothers' visual exploration influenced that of their infants.

Despite evidence of above-chance gaze synchronization, infants' head and eye movements were usually different from those of their mothers. Mostly, infants did *not* look where mothers were looking: 88% of the time they were not looking at the same object, and 64% of the time they were not even looking in the same region of the hallway. This suggests that immersion in a panoramic scene and the ability to make independent head movements limits inter-observer consistency. Thus, the high consistency found in screen-based free viewing may be a laboratory phenomenon reliant on watching two-dimensional, artfully framed movie stimuli under head-fixed conditions.

An interesting developmental implication of passively carried infants and actively walking mothers looking in different directions is that infants' visual exploration is likely to be less tightly correlated with ongoing movement. This may contribute to deficits in visuo-spatial skills seen prior to the onset of self-produced locomotion (Bertenthal & Campos, 1990; Campos, Anderson, Barbu-Roth, Hubbard, Hertenstein & Witherington, 2000). It has been suggested that passively carried infants assume a 'state of "visual idle" – staring blankly straight ahead and not focusing on single objects in the environment' (Campos, Svejda, Campos & Bertenthal, 1982, p. 208). Infants in the current study did not appear to be 'visually idling' – they actively explored their environ-

ment during passive locomotion. However, visual experiences that are uncorrelated with ongoing movement may not be optimal for developing spatial search, position constancy, or other perceptual and cognitive abilities that emerge alongside locomotor development.

Social attention in the real world

In the last few years, a rash of studies has revealed that in screen-based displays, infants preferentially attend to images of people and faces (Amso *et al.*, 2014; Aslin, 2009; DeNicola *et al.*, 2013; Di Giorgio *et al.*, 2012; Frank *et al.*, 2009; Gliga *et al.*, 2009; Gluckman & Johnson, 2013; Libertus & Needham, 2011; Schietecatte *et al.*, 2011). The current findings suggest that this social bias generalizes to real-world free viewing. Despite the cluttered nature of our naturalistic visual environment, infants looked at people about half the time that they were present.

Surprisingly, over half of this social looking was directed at body parts other than the face or head. Many screen-based studies examining infants' social attention have displayed people only from the neck up (DeNicola *et al.*, 2013; Di Giorgio *et al.*, 2012; Gliga *et al.*, 2009; Libertus & Needham, 2011), presumably under the assumption that people's arms, legs, and trunks are not informative social gaze targets. In the current study, not only were whole bodies visible to infants but they were dynamic stimuli: People walked, turned around, held objects, and gestured to one another. Viewing moving human actors is important visual input for developing perceptions of agency (Cicchino, Aslin & Rakison, 2011) and biological motion (Bertenthal, Proffitt & Kramer, 1987). Our findings suggest that infants actively seek out this visual input in real-world free-viewing situations.

But how do we reconcile infants' high rates of social looking in the current study with other head-mounted eye-tracking work showing that infants rarely attend to their parents' faces (Franchak *et al.*, 2011; Yu & Smith, 2013)? In previous work, infants were engaged in compelling natural tasks, navigating locomotor obstacles and actively manipulating objects; thus, these tasks drew the majority of infants' visual attention. Moreover, in free locomotor play, infants are so short that adult faces are typically out of view. Our findings suggest that low rates of looking to faces during natural play are due to physical constraints of infants' bodies and the distractions of immediate tasks, not an inherent lack of interest in faces. Infants do receive considerable visual input about faces, but these opportunities for learning likely occur in free-viewing situations, rather than active playtime.

A striking finding was the dramatic reversal of the age trend seen in previous screen-based studies (Amso *et al.*,

2014; Aslin, 2009; Frank *et al.*, 2009): Infants looked at people and faces far more often than did adults. Why do adults look almost exclusively at faces in screen-based displays (Cerf *et al.*, 2007; Foulsham *et al.*, 2010; Klin *et al.*, 2002; Shepherd *et al.*, 2010; Yarbus, 1967), but not in the real world? One possibility is that mothers' desire to look at people was overridden by the visual requirements for balance and steering. However, the navigation task was not a difficult one (walk straight down a hallway and back) and, in fact, most obstacles that had to be avoided were other people – in virtual reality environments, observers look at pedestrians to avoid collision (Jovancevic, Sullivan & Hayhoe, 2006). A more likely interpretation is that mothers had internalized social rules about staring at strangers and actively avoided looking at people, especially their faces (Laidlaw *et al.*, 2011). On this account, infants are oblivious to social rules so stare directly at faces.

Another possible contributor to this age difference is the development of peripheral visual acuity. Some studies suggest that acuity in peripheral parts of the visual field may not be adult-like by 9 months of age (Mohn & van Hof-van Duin, 1986; Sireteanu, Fronius & Constantinescu, 1994). Thus, infants may need direct fixation to gain high-resolution information about people, whereas adults may obtain adequate resolution without looking directly at people.

Visual saliency in the real world

As in screen-based free viewing, both infants and mothers directed their gaze to salient parts of the scene – areas that stood out from the surround on the basis of color, intensity, orientation, flicker, and motion. However, gaze was not driven purely by saliency: Infants and mothers rarely looked at the most salient part of the scene, indicating that other top-down factors play a role in infants' and adults' visual exploration of natural, real-world scenes.

Moreover, our data suggest that low-level saliency is correlated with higher-level content: In this case, people were especially salient gaze targets. Therefore, findings that gaze is related to saliency do not necessarily imply that bottom-up visual saliency is the factor driving gaze (Cristino & Baddeley, 2009; Einhauser *et al.*, 2008b; Elzary & Itti, 2008; Henderson *et al.*, 2007; Tatler *et al.*, 2011).

Unlike previous work where saliency was more predictive for adults than infants (Bambach *et al.*, 2013; Frank *et al.*, 2009), we found no difference between mothers and infants in the visual saliency at gaze locations. Perhaps saliency values at gaze were decreased for mothers in the current study because they

avoided looking at visually salient people. In addition, in face-to-face play (Bambach *et al.*, 2013), infants and parents – due to differences in height – experienced different viewpoints of the tabletop scene, resulting in very different visual displays; in the current study, mothers and infants had comparable viewpoints.

It is important to note that temporal aspects of bottom-up and top-down visual attention were not addressed in the current analyses. In lab studies, orienting based on top-down processes occurs rapidly upon presentation of a scene, and bottom-up processes play a greater role during subsequent exploration (Parkhurst *et al.*, 2002). Thus, in the real world, we might expect the influence of bottom-up factors to be stronger when the visual environment changes – when the viewer turns around or a new person enters the room, for example. Because real-world visual scenes are constantly changing and are actively controlled by the viewer, the issue of how lower- and higher-level influences on visual attention unfold over time is an important question for further study using head-mounted eye tracking in natural environments.

Conclusions

Our findings reveal that real-world free viewing differs from both real-world visual guidance of action and screen-based free viewing. In real-world free viewing, infants and adults freely move their heads to sample the panoramic scene and generally keep their eyes in line with their heads, and they look at relatively salient – but not the *most* salient – parts of the environment. Mutual body movement may influence the direction of looking, but infants and adults rarely look at the same object at the same time, and infants but not adults look frequently at social stimuli. Our findings also indicate that infants are actively engaged in obtaining rich visual input during imposed free-viewing situations such as passive locomotion while being carried by their caregivers.

Acknowledgements

The project was supported by Award # R37HD033486 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to Karen Adolph and by Graduate Research Fellowship # 0813964 from the National Science Foundation to Kari Kretch. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development, the National Institutes of Health, or the National Science Foundation. Portions

of this work were presented at the 2010 meeting of the International Society for Developmental Psychobiology and the 2012 International Conference on Infant Studies. We gratefully acknowledge Julia Brothers and members of the NYU Infant Action Lab for helping to collect and code data.

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.
- Aslin, R.N. (2009). How infants view natural scenes gathered from a head-mounted camera. *Optometry and Vision Science*, *86*, 561–565.
- Bambach, S., Crandall, D.J., & Yu, C. (2013). Understanding embodied visual attention in child–parent interaction. *Proceedings of the IEEE Conference on Development and Learning and Epigenetic Robotics*.
- Bertenthal, B.I., & Campos, J.J. (1990). A systems approach to the organizing effects of self-produced locomotion during infancy. In C.K. Rovee-Collier & L.P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 6, pp. 1–60). Norwood, NJ: Ablex.
- Bertenthal, B.I., Proffitt, D.R., & Kramer, S.J. (1987). Perception of biomechanical motions by infants: implementation of various processing constraints. *Journal of Experimental Psychology: Human Perception and Performance*, *13* (4), 577–585.
- Buswell, G.T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago, IL: University of Chicago Press.
- Campos, J.J., Anderson, D.I., Barbu-Roth, M.A., Hubbard, E.M., Hertenstein, M.J., & Witherington, D.C. (2000). Travel broadens the mind. *Infancy*, *1*, 149–219.
- Campos, J.J., Svejda, M.J., Campos, R.G., & Bertenthal, B. (1982). The emergence of self-produced locomotion: its importance for psychological development in infancy. In D. Bricker (Ed.), *Intervention with at-risk and handicapped infants* (pp. 195–216). Baltimore, MD: University Park Press.
- Cerf, M., Harel, J., Einhauser, W., & Koch, C. (2007). Predicting human gaze using low-level saliency combined with face detection. *Advances in Neural Information Processing Systems*, *20*, 1–7.
- Cicchino, J.B., Aslin, R.N., & Rakison, D.H. (2011). Correspondences between what infants see and know about causal and self-propelled motion. *Cognition*, *118*, 171–192.
- Cristino, F., & Baddeley, R. (2009). The nature of the visual representations involved in eye movements when walking down the street. *Visual Cognition*, *17*, 880–903.
- Cutting, J.E., Alliprandini, P.M.Z., & Wang, R.F. (2000). Seeking one's heading through eye movements. *Psychonomic Bulletin & Review*, *7*, 490–498.
- DeNicola, C.A., Holt, N.A., Lambert, A.J., & Cashon, C.H. (2013). Attention-orienting and attention-holding effects of faces on 4- to 8-month-old infants. *International Journal of Behavioral Development*, *37*, 143–147.
- Di Giorgio, E., Turati, C., Altoe, 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*, 66–77.
- Dorr, M., Martinetz, T., Gegenfurtner, K.R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10* (10), 28, 1–17.
- Einhauser, W., Rutishauser, U., & Koch, C. (2008a). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8* (2), 2, 1–19.
- Einhauser, W., Spain, M., & Perona, P. (2008b). Objects predict fixations better than early saliency. *Journal of Vision*, *8* (14), 18, 1–26.
- Elzary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, *8* (3), 3, 1–15.
- Foulsham, T., Cheng, J.T., Tracy, J.L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: effects of social status and speaking. *Cognition*, *117*, 319–331.
- Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception*, *36*, 1123–1138.
- Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8* (2), 6, 1–17.
- Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what, and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*, 1920–1931.
- Franchak, J.M., & Adolph, K.E. (2010). Visually guided navigation: head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research*, *50*, 2766–2774.
- Franchak, J.M., Heeger, D.H., Hasson, U., & Adolph, K.E. (2013, April). Free-viewing gaze behavior in infants and adults. Poster presented at the meeting of the Society for Research in Child Development, Seattle, WA.
- Franchak, J.M., Kretch, K.S., Soska, K.C., & Adolph, K.E. (2011). Head-mounted eye tracking: a new method to describe infant looking. *Child Development*, *82*, 1738–1750.
- Frank, M.C., Vul, E., & Johnson, S.P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*, 160–170.
- Fuller, J.H. (1996). Eye position and target amplitude effects on human visual saccadic latencies. *Experimental Brain Research*, *109*, 457–466.
- Gibson, E.J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology*, *39*, 1–41.
- Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy*, *14*, 550–562.
- Gluckman, M., & Johnson, S.P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology*, *4*, 527.

- Goldstein, R.B., Woods, R.L., & Peli, E. (2007). Where people look when watching movies: do all viewers look at the same place? *Computers in Biology and Medicine*, **37**, 957–964.
- Hasson, U., Yang, E., Vallines, I., Heeger, D.H., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, **28**, 2539–2550.
- Hayhoe, M.M., Shrivastava, A., Mruczek, R.E.B., & Pelz, J.B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, **3** (1), 49–63.
- Henderson, J.M., Brockmole, J.R., Castelano, M.S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R.P.G. van Gompel, M.H. Fischer, W.S. Murray & R.L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford: Elsevier.
- Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, **12**, 1093–1123.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, **40**, 1489–1506.
- Itti, L., Koch, B., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, 1254–1259.
- Jovancevic, J., Sullivan, B.T., & Hayhoe, M.M. (2006). Control of attention and gaze in complex environments. *Journal of Vision*, **6**, 1431–1450.
- Kirkorian, H.L., Anderson, D.R., & Keen, R. (2012). Age differences in online processing of video: an eye movement study. *Child Development*, **83**, 497–507.
- Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, **59**, 809–816.
- Laidlaw, K.E.W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences, USA*, **108**, 5548–5553.
- Land, M.F., & Lee, D.N. (1994). Where we look when we steer. *Nature*, **369**, 742–744.
- Land, M.F., & McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nature Neuroscience*, **3** (12), 1340–1345.
- Land, M.F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, **28**, 1311–1328.
- Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research*, **47**, 2483–2498.
- Libertus, K., & Needham, A. (2011). Reaching experience increases face preference in 3-month-old infants. *Developmental Science*, **14**, 1355–1364.
- Marchant, P., Raybould, D., Renshaw, T., & Stevens, R. (2009). Are you seeing what I'm seeing? An eye-tracking evaluation of dynamic scenes. *Digital Creativity*, **20**, 153–163.
- Mital, P.K., Smith, T.J., Hill, R.L., & Henderson, J.M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, **3**, 5–24.
- Mohn, G., & van Hof-van Duin, J. (1986). Development of the binocular and monocular visual fields of human infants during the first year of life. *Clinical Vision Sciences*, **1**, 51–64.
- Pare, M., & Munoz, D.P. (2001). Expression of a re-centering bias in saccade regulation by superior colliculus neurons. *Experimental Brain Research*, **137**, 354–368.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of saliency in the allocation of overt visual attention. *Vision Research*, **42**, 107–123.
- Patla, A.E., & Vickers, J.N. (1997). Where and when do we look as we approach and step over an obstacle in the travel path? *NeuroReport*, **8**, 3661–3665.
- Pelz, J.B., Hayhoe, M.M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, **139** (3), 266–277.
- Peters, R.J., Iyer, A., Itti, L., & Koch, B. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, **45**, 2397–2416.
- Raudies, F., Gilmore, R.O., Kretch, K.S., Franchak, J.M., & Adolph, K.E. (2012). Understanding the development of motion processing by characterizing optic flow experienced by infants and their mothers. *Proceedings of the IEEE Conference on Development and Learning*.
- Rothkopf, C.A., Ballard, D.H., & Hayhoe, M.M. (2007). Task and context determine where you look. *Journal of Vision*, **7** (14), 16, 1–20.
- Schietecatte, I., Roeyers, H., & Warreyn, P. (2011). Can infants' orientation to social stimuli predict later joint attention skills? *British Journal of Developmental Psychology*, **30**, 267–282.
- Shepherd, S.V., Steckenfinger, S.A., Hasson, U., & Ghazanfar, A.A. (2010). Human–monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Current Biology*, **20**, 649–656.
- Sireteanu, R., Fronius, M., & Constantinescu, D.H. (1994). The development of visual acuity in the peripheral visual field of human infants: binocular and monocular measurements. *Vision Research*, **34**, 1659–1671.
- 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, 1–17.
- Tatler, B.W., Baddeley, R.J., & Gilchrist, I.D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research*, **45**, 643–659.
- Tatler, B.W., Hayhoe, M.M., Land, M.F., & Ballard, D.H. (2011). Eye guidance in natural vision: reinterpreting saliency. *Journal of Vision*, **11** (5), 5, 1–23.
- 't Hart, B.M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., Konig, P., & Einhauser, W. (2009). Gaze allocation in natural stimuli: comparing free exploration to head-fixed viewing conditions. *Visual Cognition*, **17**, 1132–1158.
- Tosi, V., Mecacci, L., & Pasquali, E. (1997). Scanning eye movements made when viewing film: preliminary observations. *International Journal of Neuroscience*, **92**, 47–52.

- Tseng, P.H., Carmi, R., Cameron, I.G.M., Munoz, D.P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision*, **9** (7), 4, 1–16.
- Turano, K.A., Gerguschat, D.R., & Baker, F.H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research*, **43**, 333–346.
- Underwood, G., Foulsham, T., van Loon, E., & Underwood, J. (2005). Visual attention, visual saliency and eye movements during the inspection of natural scenes. In J. Mira & J.R. Alvarez (Eds.), *Artificial intelligence and knowledge engineering applications: A bioinspired approach, LNCS 3562* (pp. 459–468). Berlin: Springer-Verlag.
- Wang, R.F., & Cutting, J.E. (2004). Eye movements and an object-based model of heading perception. In L. Vaina, S. Beardsley & S. Rushton (Eds.), *Optic flow and beyond* (pp. 61–78). Dordrecht, The Netherlands: Kluwer Academic Press.
- Wang, X.H., Freeman, J., Merriam, E.P., Hasson, U., & Heeger, D.H. (2012). Temporal eye movement strategies during naturalistic viewing. *Journal of Vision*, **12** (1), 16, 1–27.
- Yarbus, A.L. (1967). *Eye movements and vision*. New York: Plenum.
- Yu, C., & Smith, L.B. (2013). Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye–hand coordination. *PLoS ONE*, **8** (11), e79659.

Received: 21 October 2013

Accepted: 18 August 2014