CHAPTER THREE

Visual exploratory behavior and its development

John M. Franchak*
Department of Psychology, University of California, Riverside, CA, United States
*Corresponding author: e-mail address: franchak@ucr.edu

Contents

Abstract

Visual exploratory behavior refers to actively gathering visual information through coordinated eye, head, and body movements—looking around with a purpose. In this article, I compare the study of visual exploration using stationary, screen-based tasks vs mobile, naturalistic tasks. In doing so, I discuss how ecological validity depends on variations in three factors: visual characteristics of stimuli, the opportunity for participants to interact with their surroundings, and participants' ability to move their bodies. In particular, the long-standing reliance on stationary, screen tasks has precluded progress in understanding the last factor—how anatomical, biomechanical, and physiological aspects of the body influence where observers look. I argue that each factor is vital to revealing the flexibility, adaptiveness, and efficiency of everyday visual exploration.

Psychology of Learning and Motivation, Volume 73
ISSN 0079-7421
https://doi.org/10.1016/bs.plm.2020.07.001

1. Introduction

For sighted individuals, carrying out the activities of daily life relies on gathering information through visual perception. The eyes shift from one place to another 2–3 times per second—approximately 200,000 times each day—to attend to people, objects, and places in the world. Although we may, at times, "stare off into the distance," or "space out," most looking events result from *visual exploratory behavior*—purposeful orientation of the visual system to acquire information relevant to one's current goals and actions (Gibson, 1988, 1966). The word "exploration" warrants emphasis. Perception is not passive sensation, but a process of active information gathering—exploring. What behaviors comprise visual exploration? Eye movements certainly facilitate visual exploration, but the eyes alone can only shift gaze within the head-centered visual field. Movements of the eyes within the head, the head within the body, and the body in space must be coordinated for an observer to distribute gaze in all directions in a complex, natural environment (Gibson, 1979; Land, 2004). As such, visual exploratory behavior refers to active, whole-body coordination of the visual system to support the informational needs of the observer (Franchak, 2020).

Visual exploration is a *selective* process. The binocular visual field spans approximately 180 degrees horizontally, meaning that only a portion of the world is in view at any moment. Moreover, the quality of information varies substantially across the visual field—the high-acuity fovea comprises only the central 2 degrees of the retina. Thus, when gaze is directed at one location, other locations are either unseen (out of view completely) or they are seen through low-acuity peripheral vision. Given these constraints, researchers have long sought to understand the processes that govern where people look (e.g., Bahill, Adler, & Stark, 1975; Yarbus, 1967). Why are some locations attended instead of others? Bottom-up influences and top-down influences have been the primary focus of past work. *Bottom-up influences* refer to looking behavior that is driven by the appearance of stimuli, regardless of task relevance or meaning. For example, a flashing light on the dashboard might pull the driver's gaze away from the road. In contrast, *top-down influences* reflect how an observer's task and goals guide looking. A driver searching for a particular street will fixate street signs even when there are more visually salient targets in the field of view (e.g., a red car zooming past). Of course, these influences can occur simultaneously: The flashing light on the dashboard may attract attention both because of its appearance (i.e., a bright, flickering light) and because of its task relevance (i.e., the driver knows the dashboard provides important information about the vehicle).

The first goal of this article is to describe how a third category—*embodied influences*—shapes visual exploratory behavior. For example, a flashing light on a low dashboard panel requires the observer to tilt both the eyes and the head down, temporarily orienting the view away from the road. Multiple aspects of the body are relevant to understanding visual exploratory behavior. The *anatomy* of the body—such as the height of the eyes from the ground plane when sitting or standing—determines what is in view. Visual exploration also depends on *biomechanical* considerations, such as the range of movement of the eyes and head and how those movements are coordinated. Finally, *physiological* factors, such as the energy required to make different types of exploratory movements, may influence where people choose to look. Relative to bottom-up and top-down influences, little is known about embodied influences. In part, this is due to methodology. Many aspects of bottom-up and top-down influences can be studied in screen-based tasks, in which *screen eye trackers* (SETs) record where people look on a computer monitor (Fig. 1A). However, screen-based tasks are ill-suited for studying embodied influences: With a seated observer viewing images that are placed directly in front of the head, it is impossible to know how the body and head contribute to looking behavior. Studying embodied influences on exploratory behavior requires *mobile eye trackers* (MET) to record gaze in observers who are free to move eyes, head, and body (Fig. 1B).

The second goal of this article is to discuss how development, from infancy to adulthood, changes visual exploration. Although studies of infants' and children's visual exploration in SET tasks have been conducted since the 1980s, advances in MET technology over the past decade have forged a new literature describing the development of embodied visual exploration in naturalistic tasks. Physical and motor development—changes in anatomy due to growth and the acquisition and refinement of motor skills—mean that embodied influences on visual exploration are considerable. Greater homogeneity in adults' bodies and motor abilities compared with infants and children, both between individuals and over time, masks the role of embodied influences on visual exploration and makes them more difficult to study. Situating visual exploration within the context of physical and motor development reveals how the body shapes what we see and what we learn in the first years of life.

Finally, naturalistic studies of visual exploration from participants across the lifespan show that generalization from screen-based tasks to everyday

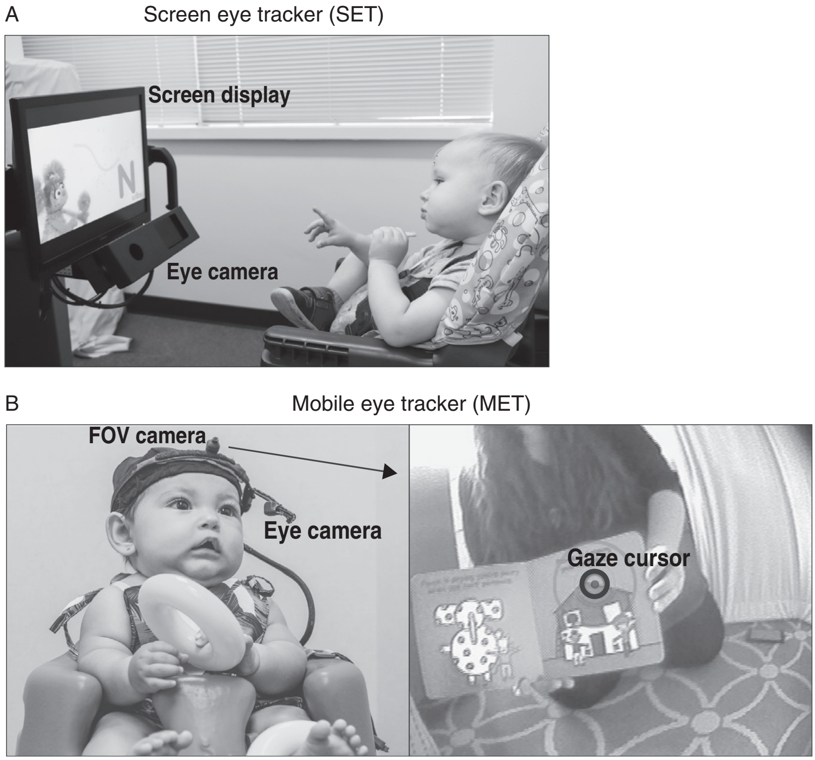

Fig. 1 (A) Screen eye tracker (SET) with infant observer sitting in a high chair. The eye camera detects the infant's eyes to detect where gaze is directed within the screen display. (B) Mobile eye tracker (MET). The *left image* shows an infant wearing the MET headgear, which contains an eye camera to record eye movements and a field of view (FOV) camera that records the head-centered perspective. The *right image* shows an infant's field of view video with a *black circle* indicating the infant's gaze location.

situations is not assured (Land & Tatler, 2009). Accordingly, the third goal of this article is to characterize the features of both paradigms that impact ecological validity. This article is organized around three aspects of ecological validity. First, how closely do the *visual characteristics* of what observers see in a research study approximate the visual characteristics of the real world? Second, what is the degree of *interactivity* afforded to the observer? Third, to what extent are *embodied influences* able to affect visual exploration? Isolating exactly which aspects of screen-based and real world tasks are most relevant to ecological validity requires deeper analysis because SET and MET studies vary in many ways. Box 1 reviews the features of screen eye tracking (SET) and mobile eye tracking (MET) technology that are relevant to understanding their contributions to studying visual exploration.

BOX 1 Screen eye tracking vs mobile eye tracking

Eye tracking systems measure *gaze*—the location the observer is looking at on a screen or in the world—through video recording of eye position. A screen eye tracker (SET) uses a specialized eye camera to detect the eye movements of a stationary observer (Fig. 1A). Although some SET systems may require the head to be immobilized in a chin rest, other SET systems detect the observer's head and then detect the eyes within the head. Even in head-free eye tracking, head movement is limited because the participant cannot move beyond the fixed field of view of the eye tracking cameras. Modern SET systems boast superb spatial accuracy (0.02–0.5 degrees) and temporal sampling rates (500–2000 Hz). These specifications are made possible by constraining the observer's movement and restricting the trackable region to a screen in a static location.

Aside from superior data quality, SET systems collect data that are more easily analyzed because gaze and stimuli coordinates share a single frame of reference. SET systems measure eye gaze in screen (e.g., pixel) coordinates, and researchers present stimuli that can be referenced in screen coordinates, making it straightforward to determine where people are looking. This also facilitates calculating patterns of behavior across multiple observers: For two different observers who watched the same stimulus on the same screen, their gaze locations are comparable. That is not to say that analysis of SET data is *easy*. As stimulus complexity increases, it becomes more difficult to determine how best to compare spatial and temporal exploration across observers and conditions.

Mobile eye trackers (METs) are wearable devices that use miniature cameras mounted on a pair of glasses or a headband to record gaze (Fig. 1B). As with SET systems, MET devices use specialized cameras (often infrared) to detect movements of the eye (i.e., *eye cameras*). Computer vision algorithms analyze eye images to detect the location of the pupil and/or cornea as markers of eye position. However, since a MET is not integrated with a screen, a second camera (i.e., *field of view camera*) is needed to capture the "stimulus"—that is, what is in view for the observer (right image in Fig. 1B). The FOV camera resides on the headgear, facing outward, to capture the observer's head-centered view. As the observer moves the head, the FOV camera moves along with the head to record the corresponding changes in the field of view. Likewise, as the observer walks from one place to another, the FOV camera translates in space to capture the perspective from the new location. A calibration process links eye position from the eye camera to gaze position in the FOV camera to determine where observers look in the head-centered field of view. Compared to SET, an MET's spatial accuracy (0.5–2 degrees) and temporal resolution (30 Hz to 200 Hz) is worse. Although technological advances in MET systems continue to improve the accuracy and sampling rate, METs will always be at a disadvantage compared to SETs because they need to isolate pupil movement from the noise of body movement.

Continued

BOX 1 Screen eye tracking vs mobile eye tracking—cont'd

Another important difference is that an MET system records gaze in FOV video coordinates (rather than screen coordinates). The contents of the FOV camera—and what location each pixel corresponds to in the world—changes with every rotation and translation of the head. Each observer creates their own "stimulus video" through the unique movements of their heads and bodies, meaning that targets of interest will move in and out of view and occupy different locations in the FOV camera at different times (and for different observers). Thus, it is more difficult to compare visual exploration across a sample of observers for an MET study compared with an SET study.

2. Visual characteristics contribute to ecological validity

The complexity of visual stimuli—from more artificial to more life-like—impacts our understanding of how bottom-up and top-down factors influence visual exploration. Bottom-up influences on eye movements are typically measured by characterizing the *salience* of potential gaze locations based on computational models of low-level visual features, such as luminance, color, and motion (Borji & Itti, 2013). Top-down influences on eye movements are measured by determining whether people look at *semantically meaningful targets* such as faces, objects, actions. Top-down influences also refer to how an observer's tasks and goals influence where they choose to look, which will be reviewed in the next section. I will argue that interaction between bottom-up and top-down features is inherent in real-world visual exploration, and that studies lacking realistic visual characteristics fail to capture such interaction.

2.1 Static images on screens

Traditional vision science research involves highly controlled, elemental stimuli (e.g., Gabor patches, simple shapes) displayed on computer monitors. For example, in a study testing 3- to 10-year-olds' eye movements while viewing illusory contours (e.g., edges that are not defined by a change in luminance or texture, but rather by the global context of the image), my colleagues and I showed children a set of four black "pacman" inducers on a light gray background to evoke different illusory shapes (Nayar, Franchak, Adolph, & Kiorpes, 2015). Using basic stimuli that lack meaning reduces top-down influences on visual attention to isolate bottom-up influences. However, it is difficult to generalize from visual stimuli that are so dissimilar from everyday scenes. Moving away from the extreme of simple shapes, one step toward ecological validity is showing photographs of real objects. For example, Kwon, Setoodehnia, Baek, Luck, and Oakes (2016) presented arrays of faces and objects (e.g., cups, stuffed animals) on a computer monitor to measure which targets held infants' attention for the longest amount of time. Eight-month-old infants looked longer at faces compared with objects even when the objects had greater bottom-up salience, presumably because faces have more top-down relevance. However, 4-month-olds looked at the most brightly colored, salient target regardless of whether it was an object or a face, suggesting that bottom-up influences outweighed top-down influences for young infants. Other studies using arrays of photographs found that infants preferentially look at faces over nonface stimuli by 6 months of age (Di Giorgio, Turati, Altoe, & Simion, 2012; Gliga, Elsabbagh, Andravizou, & Johnson, 2009); this increase in face looking is thought to reflect developmental improvements in top-down attention because infants preferentially select to look at meaningful locations. As we will see in Section 2.2, the development of visual exploration is more complicated than merely looking more often at semantically meaningful targets.

A disadvantage of presenting simple shapes or photographs of objects on solid backgrounds is that such stimuli are unlike everyday scenes. Many researchers have turned to "natural scenes"—photographs of everyday environments—to improve ecological validity (e.g., Henderson & Hollingworth, 1999). Whereas a shape on a neutral background contains prominent, high-contrast edges (e.g., a black "pacman" on a gray background), natural scene photographs have visual features that more closely approximate those of the real world. Studies using natural images showed that visual exploration depends on interactions between top-down and bottom-up features that cannot be easily dissociated in artificial displays (Einhauser, Spain, & Perona, 2008; Henderson, Brockmole, Castelhano, & Mack, 2007). Observers first narrow the search space based on expectations about the search target (e.g., cars are on the road, mugs are on counters) (Torralba, Oliva, Castelhano, & Henderson, 2006). After the search space is narrowed, bottom-up visual features guide eye movements to search for the target.

Another drawback of simple stimulus arrays is that they lack spatial regularities that observers use to guide looking behavior. In natural scenes (and photographs of natural scenes), visual features are distributed asymmetrically. Buildings, objects, and people must rest on the ground, so targets of interest are typically distributed along the horizon—horizontally—in an image rather than vertically. Observers' viewing patterns are sensitive to these spatial regularities. Both infants (van Renswoude, Johnson, Raijmakers, & Visser, 2016) and adults (Foulsham, Kingstone, & Underwood, 2008) distribute gaze more widely along the horizontal as opposed to the vertical axis. When images are artificially rotated—putting the horizon at an angle—adults adjust their saccade direction to follow the rotated horizon (Foulsham et al., 2008). Infants (Mahdi, Su, Schlesinger, & Qin, 2017; van Renswoude, van den Berg, Raijmakers, & Visser, 2019) and adults (Tatler, 2007) also show a *center bias* by preferentially looking to the middle of a photograph rather than the edges. One explanation for the center bias reveals the limitations of using so-called natural scene photographs: Photographers are biased to center interesting things in view when composing a photograph, resulting in images that contain a higher density of objects, and thus salient visual features, in the middle (Tseng, Carmi, Cameron, Munoz, & Itti, 2009). Yet, visual feature distributions cannot fully explain the center bias. Infants and adults spend more time exploring the center even when an image has a higher density of salient features around the edges (Tatler, 2007; van Renswoude et al., 2019), suggesting that the center bias is a visual exploration strategy and/or may result from embodied influences. These explanations will be revisited in Sections 2.2 and 4.3.

2.2 Dynamic movies on screens

Using dynamic movies as opposed to static images provides a substantial improvement to ecological validity. When investigating a picture, we can shift gaze in sequence and, with the luxury of time, capture the entire scene. But in everyday situations, choosing to look at one event means we miss seeing others. Thus, observers must prioritize where to look not just in space but also in time. SET studies have examined how observers view dynamic scenes primarily through displaying television shows and movies. Although this increases ecological validity by incorporating motion, using "found" stimuli like television shows means that relevant properties—such as the number/type of agents/objects, the size and spacing of different targets, and the narrative content in a scene—are selected for but not perfectly controlled or counterbalanced.

When watching television shows and movies, adults show a high degree of *attentional synchrony*—the degree to which multiple observers look at the same place at the same time—suggesting that similar bottom-up and top-down processes guide the visual exploration of different observers (Shepherd, Steckenfinger, Hasson, & Ghazanfar, 2010; Wang, Freeman, Merriam, Hasson, & Heeger, 2012). Indeed, attentional synchrony is influenced, in part, by motion (Mital, Smith, Hill, & Henderson, 2011; Smith & Mital, 2013): Observers' gaze is drawn to the same moving targets that stand out from the background of the scene. But attentional synchrony cannot be reduced only to bottom-up features—synchrony depends on fixating meaningful locations in a scene (e.g., objects and faces) (Franchak, Heeger, Hasson, & Adolph, 2016; Shepherd et al., 2010). Synchrony also relates to observers' understanding of a scene: Attentional synchrony in adults is disrupted when viewing the shots of a television show in a scrambled as opposed to intact order (Kirkorian & Anderson, 2018). Better understanding of a scene facilitates looking toward meaningful areas. While watching clips from a tennis match, observers who had more knowledge about tennis were more likely to anticipate the movement of the ball compared with nonexperts (Taya, Windridge, & Osman, 2013).

With age, observers visually explore scenes more similarly. Attentional synchrony *within* groups of infants and children increases with age when watching *Charlie Brown* and *Sesame Street* clips (Franchak et al., 2016; Frank, Vul, & Johnson, 2009; Kirkorian, Anderson, & Keen, 2012). That is, a group of 4-year-olds are more likely to look at the same place at the same time, but 1-year-olds are more likely to explore in an idiosyncratic way. Measuring attentional synchrony *between* infants and adults shows how developing visual exploration becomes more adult-like: Younger infants' (6 and 9 months) eye movements weakly correlated with adults', but 12- and 24-month-olds' eye movements were more strongly correlated with adults' (Franchak et al., 2016). In current work, we showed that this finding generalized across seven video clips with varying content and across a wide range of ages (6 months to 11 years) (Kadooka & Franchak, in preparation). Logarithmic fits (Fig. 2, top row) indicated that adult-like visual exploration increased rapidly in infancy and more slowly through childhood.

What drives the development of adult-like exploration? One idea is that there is a global shift from attending primarily to bottom-up, salient features toward attending to meaningful, top-down locations. Support for this theory comes from the static image studies reviewed above (Di Giorgio et al., 2012; Gliga et al., 2009; Kwon et al., 2016) as well as studies that used dynamic scenes (Franchak et al., 2016; Frank et al., 2009). For example, whereas a saliency model was more predictive of young infants' eye

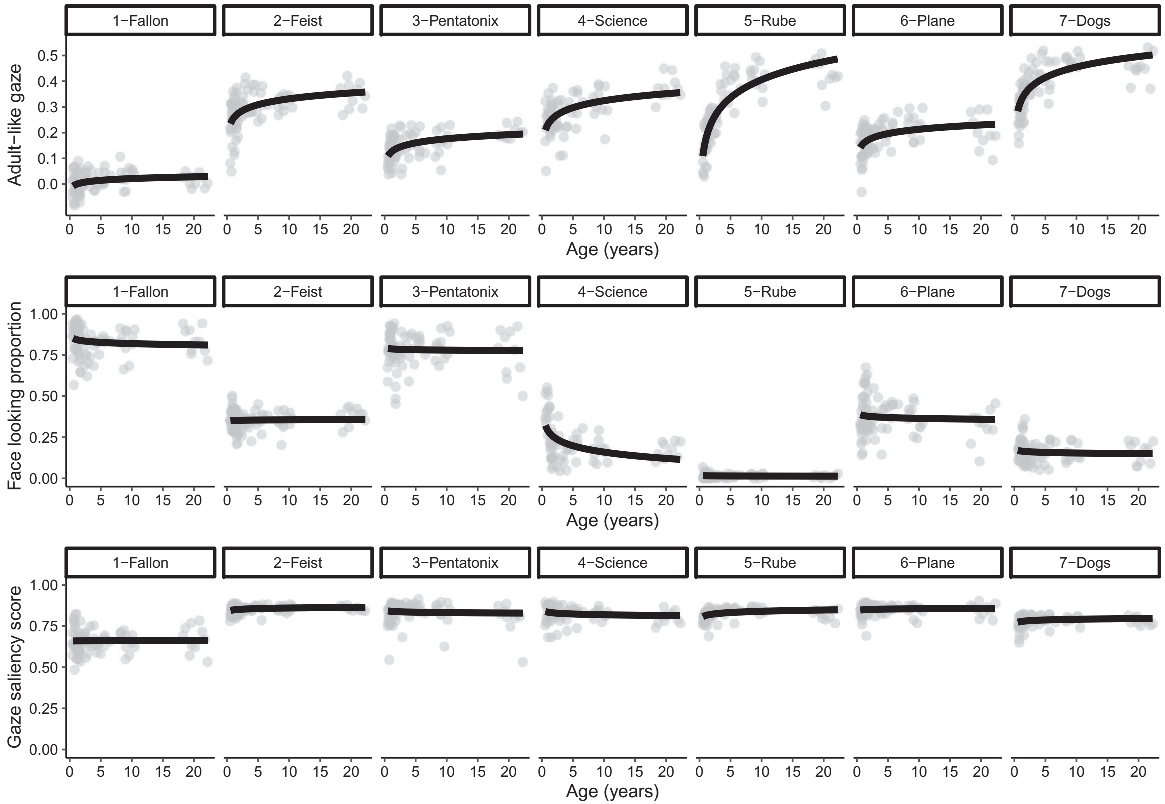

Fig. 2 Visual exploration of seven different video clips (*numbered columns*) by observers aged 6 months to 11 years (*x*-axis). Each *gray point* indicates one participant's data. *Top row* shows observers' adult-like gaze score—the degree to which each observer tended to look at the same location at the same time as a group of adult observers (Kadooka & Franchak, in preparation). *Middle row* shows the proportion of time that each observer spent looking at faces during the video (Kadooka & Franchak, in press). *Bottom row* shows the gaze saliency score for each participant—how often they looked at more salient areas of the video (closer to 1) vs less salient areas (closer to 0) (Kadooka & Franchak, in press). Whereas adult-like gaze scores increased with age for every video (*black lines* show logarithmic fit lines), there were no consistent changes in face looking or gaze saliency scores across videos. Example clips from each video can be viewed at: https://nyu.databrary.org/volume/1007.

movement patterns, older infants' and adults' eye movements were more drawn to meaningful locations (faces) (Frank et al., 2009). Furthermore, my colleagues and I found infants' tendency to fixate faces predicted their adult-like gaze scores even after controlling for age (Franchak et al., 2016). However, in follow-up work that tested across a more diverse set of stimulus videos (example clips are available on Databrary, https://nyu.databrary.org/volume/1007), this finding did not replicate (Kadooka & Franchak, in press). In fact, we found no consistent age-related changes in either looking at faces (Fig. 2, middle row) or looking at salient areas (Fig. 2, bottom row) across seven videos. Only one stimulus video ("4-Science") showed an age effect for face looking—face looking *decreased* with age, contrary to the global shift hypothesis. Only two videos ("2-Feist" and "5-Rube") showed age-related increases in looking at salient areas, but the effect sizes were small relative to the change in adult-like gaze. Clearly, the increases in adult-like gaze seen in the top row of Fig. 2 cannot be attributed to simply looking more at faces and less at salient areas.

Instead, the development of visual exploration means learning to *prioritize* which visual features to attend. Temporal analyses over the course of each of the seven videos showed that adults modulated how much they attended to salient areas and faces from moment to moment (Kadooka & Franchak, in press). At some times adults prioritized faces, and at others they prioritized salient areas (such as moving, nonface objects). Infants and children showed similar prioritization to adults at some times but not at other times. Observers' prioritization of where to look may depend on how different features intersect. For example, Wass and Smith (2015) showed that the speaking character's face is often the most visually salient location in children's television programs, serving as a potential scaffold to infants' attention when viewing media. Yet, past work with static images suggested that young infants fail to take advantage of saliency as a cue to look at faces. Whereas children and adults look more often at salient compared to nonsalient faces, younger infants do not (Amso, Haas, & Markant, 2014). In recent work, we found that these results from static images did not translate to dynamic videos. When watching television shows, saliency cues face-looking behavior equally well for infants, children, and adults (Franchak & Kadooka, under review). Thus, bottom-up and top-down features are not separable influences on eye movements—from an early age, infants integrate both types of features to determine where to look in a complex scene.

Even though television programs and movies have the advantage of being dynamic, they also carry some features that impede ecological validity.

For example, cuts between scenes are ubiquitous in video media, but are not a feature of everyday visual experiences. As consumers of video media, children, and adults learn media-specific exploration strategies that do not translate beyond screen viewing. Both children (Jing, Kadooka, Franchak, & Kirkorian, 2019; Rider, Coutrot, Pellicano, Dakin, & Mareschal, 2018) and adults ('t Hart et al., 2009) reorient their gaze to the center of the screen following a cut, and then afterward move their eyes to a target of interest. Centering on the screen also guides visual attention to characters' faces. Important characters are often framed by the cinematographer to be in the center of view. Infants show a slight bias to fixate centered faces, but older children and adults relied strongly on centering as a cue to face looking (Franchak & Kadooka, under review). The strategy of looking at centered faces was related to media exposure: Infants who spend more time watching television were more likely to use centering as a cue for face looking. Although centering is an important feature for understanding media viewing, it does not translate to real-world environments in which there is no center with respect to a "stimulus". The center of view is wherever the observer points their head, so center changes from moment to moment.

2.3 Visual characteristics in mobile tasks

Simply using a MET system does not ensure that the visual characteristics of the stimuli will be more realistic than in SET studies. Indeed, the illusory contour example in the beginning of this section was collected using an MET system to ensure that children raising their hands to reach a touchscreen would not interfere with gaze recording (Nayar et al., 2015). Similarly, early MET studies presented basic shapes on a pegboard to investigate how observers coordinate eyes, head, and hands to copy block patterns (Pelz, Hayhoe, & Loeber, 2001; Smeets, Hayhoe, & Ballard, 1996). Limiting the visual characteristics of a scene in an MET study can also facilitate data analysis. In studies of infant–caregiver play with objects (Yu & Smith, 2013), participants sat in a white room, at a white table, and were dressed head-to-toe in white clothing so that computer vision algorithms could segment the three objects (solid-colored toys in red, blue, and green) for automated coding.

MET studies of observers "in the wild" have the highest ecological validity because observers are looking at real-world environments. Examples include driving on a road (Land & Lee, 1994), walking through a college campus (Foulsham, Walker, & Kingstone, 2011; McGee, Blanch, & Franchak, in preparation) or city street (Einhauser et al., 2007; Tomasi, Pundlik, Bowers, Peli, & Luo, 2016), making a cup of tea in a kitchen (Land, Mennie, & Rusted, 1999), searching the mailroom (Foulsham, Chapman, Nasiopoulos, & Kingstone, 2014) or hallways (Kretch & Adolph, 2015; Turano, Geruschat, & Baker, 2003) of an office building, hiking over uneven terrain (Matthis, Yates, & Hayhoe, 2018), and observing geological phenomena during a field expedition (Keane, Cahill, Tarduno, Jacobs, & Pelz, 2014). Studying diverse environments and situations reveals that the visual features encountered in daily life vary substantially from one location to the next. Analyzing field of view camera recordings from observers who visited locations such as an art museum, an urban city street, a forest, and a hospital, Schumann et al. (2008) found differences between "open" outdoor environments that allowed observers to see far off in the distance (e.g., beaches, deserts) compared with "closed" outdoor environments that contained buildings that obstructed one's view. Beyond environmental differences, the results of MET recordings of natural scenes show how visual features differ when recording from person-perspective rather than camera perspective. Whereas a photographer's (or experimenter's) bias determine how features are distributed an image, the observer's own head movements with respect to the environment determine how features are distributed in the field of view. Indeed, person-perspective recordings from METs found that visual features tended to cluster toward the top of view, not the center, suggesting that typical SET stimuli presentations—even "natural image" photographs—fail to convey the spatial distributions of features experienced in real situations (Schumann et al., 2008). Across different environments, there was a tendency for observers to direct their eyes toward salient visual features (Schumann et al., 2008). Saliency also predicted infants' and caregivers' eye movements at better-than-chance levels while walking (with infants worn in a baby carrier on the caregiver's chest) through the hallway of a university building (Kretch & Adolph, 2015). Of note, participants in the aforementioned studies were not given an explicit task, but simply walked (and looked) around, approximating the free-viewing situations employed by the SET studies reviewed above. As we will see in the next section, the introduction of more demanding tasks often limits or negates the influence of salient visual features. Instead, observers actively look toward objects and places necessary to carry out a task.

3. Interactivity contributes to ecological validity

Although at times we may visually explore a scene as passive observers—viewing the action without taking part in it—more frequently our eye movements support interaction with the environment. As active observers, the ever-changing informational needs of ongoing tasks shape where and when we need to look. Moreover, our actions have consequences on the environment (and our perspective therein), changing the visual scene. In this section, I will review how varying an observer's task and interactivity leads to different understanding of visual exploration in both SET and MET studies. Limiting or constraining how observers can interact removes the typical demands on attention that are present in daily life, lessening ecological validity. By studying visual exploration in the context of real-life action, we can better understand how observers prioritize where to look.

3.1 Passive vs active screen-based tasks

Many of the SET studies reviewed in the previous section employed a "free-viewing paradigm," in which observers watched photographs or videos on a screen without any specific instructions. Although free viewing may be representative of one everyday activity—watching television—it lacks the demands on visual attention that most other everyday tasks present. Even using a computer, another sedentary screen task, requires users to control a cursor on the screen. The pioneering work of Yarbus (1967) showed that the spatial distribution of eye movements when exploring a painting changed depending on the observer's task. A modern replication of Yarbus' early work found that participants distributed gaze more broadly around a photograph when asked to memorize its contents, whereas participants who were asked to search for a specific object quickly narrowed their focus and only explored the relevant area (Castelhano, Mack, & Henderson, 2009). Task instructions may shift how observers prioritize areas with certain features (Smith & Mital, 2013). Observers who were instructed to discern the location that a video was filmed were less drawn to areas of higher bottom-up salience compared with observers who viewed the video without instructions. However, there are disparate findings in the literature about how task instructions moderate attention to salient areas. Peacock, Hayes, and Henderson (2019) found that even when participants were given a task that should incite them to prioritize saliency (e.g., rate the brightness of the photograph), eye movements were better predicted by meaning compared with saliency. Regardless, a consistent result across SET studies is that observers adapt multiple aspects of their visual exploration—the spatial area to search, the timing of eye movements, and the types of features to prioritize—to the task at hand.

Studying the development of task-driven visual exploration is difficult because infants and toddlers are not apt to understand or follow instructions. Thus, top-down attention is typically inferred from looking at meaningful areas like objects and faces (as reviewed in the previous section). However, this approach makes the assumption that infants will look at what adult experimenters consider “meaningful”. One way that researchers have moved beyond rates of face or object looking is to test whether previous experience with targets within a session guides later exploration. Tummeltshammer and Amso (2018) enticed infants to engage in a search task by showing a simple array of shapes, and then turning the shapes over to reveal a rewarding image under one of the shapes. Over trials, infants learned reward locations in different shape displays, and looked faster to the location where the rewarding image would appear. Creating screen events that respond to observers’ eye movements—gaze contingency—is a way to allow infants to more actively participate in an SET study (Wang et al., 2012). In recent work, we showed 8-, 12-, and 18-month-olds cartoon images of a flower and a sun during a training phase (McGee, Michaels, & Franchak, 2020). Fixating the contingent target triggered a rewarding animation, whereas fixating the noncontingent target had no effect. Afterward, infants viewed television clips in which both target images were digitally inserted at random locations. Prior training modulated infants’ attention when viewing videos—they more often looked at the contingent target compared with the noncontingent target—demonstrating a top-down goal effect on exploration in a complex visual scene.

3.2 Active observers in naturalistic tasks

Although SET studies can increase interactivity by assigning tasks or creating contingent targets, the options for interactivity are limited compared with MET studies. An observer who can use their hands to touch real objects and can walk from one place to another can engage in a wide variety of natural tasks. MET also allows for an additional source of interactivity not present in SET tasks—the ability to change what is in view. Even in an interactive, gaze-contingent SET study, the observer can only see what is presented on a screen. But in naturalistic settings, an observer selects what is in view through movement.

Researchers have compared how active and passive observers visually explore using a hybrid SET/MET approach. An MET measured active observers' visual exploration while completing a real-world task, such as walking through an outdoor environment (Foulsham et al., 2011; 't Hart et al., 2009) or making a cup of tea (Tatler et al., 2013). FOV camera videos, which show how the active observer used the head to select what is in view, were then played on a screen and viewed by passive observers. An SET recorded passive observers' gaze to compare their visual exploration to the active observers' gaze. The gaze of active observers who walked through a college campus clustered more in the center of view compared to passive viewers' gaze (Foulsham et al., 2011). Since active observers used their heads to choose what was in view, their eyes could stay centered and look where the head pointed. In contrast, passive observers did not choose what was in view, so they used their eyes to explore areas across the screen, centered or not. Active observers also fixated the walking path more frequently, presumably to control walking movements, compared with passive observers who did not need to guide locomotion. In another study, active observers making a cup of tea spent more time fixating task-relevant objects (e.g., the tea kettle, tea pot) compared with passive observers who watched the videos without the demands of controlling hand movements to make tea (Tatler et al., 2013).

Actively engaging in a task affects how observers prioritize visual exploration. When making a sandwich or a cup of tea, observers look at the object that is important at that step in the process or glance ahead to the next item that will be used (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land et al., 1999). When walking along a path with oncoming pedestrians, "rogue" pedestrians whose paths would lead to collisions with the observer were fixated more frequently than "safe" pedestrians who avoided collisions (Jovancevic-Misic & Hayhoe, 2009). Top-down task relevance outweighs the influence of bottom-up visual salience on eye movements. When observers in a virtual reality study walked along a path to pick up "litter" (solid-colored rectangles) while avoiding "obstacles" (rectangles of another color), inserting visually salient objects in the background of the scene rarely drew gaze away from the task-relevant litter and obstacles (Rothkopf, Ballard, & Hayhoe, 2007). However, bottom-up cues that align with task goals can facilitate visual exploration. In a real-world search task, placing a colorful frame on a mailbox led to it being fixated more quickly, and participants who were told to expect that the mailbox would be highlighted found it even more quickly (Foulsham et al., 2014).

Studying infants' visual exploration during play reveals that how they prioritize information in the real world is different compared to when they watch stimuli on screens. Whereas infants spend over 50% of the time gazing at faces while viewing videos (Franchak et al., 2016; Frank et al., 2009), infants spend little time looking at caregivers' faces when playing with toys and caregivers at a table (12%, Yu & Smith, 2013) or on the floor (4%, Franchak, Kretch, & Adolph, 2018). Even when caregivers spoke to infants, infants responded by looking at the caregiver's face only 8% of the time (Franchak, Kretch, Soska, & Adolph, 2011). Instead of gazing at faces, infants spend more time looking at toys in both tabletop play (60%, Yu & Smith, 2013) and play on the floor (38%, Franchak et al., 2018). Just as adults making tea or a sandwich fixate task-relevant objects (Hayhoe et al., 2003; Land et al., 1999), infants' object-directed eye movements facilitate reaching and object manipulation during play. Like adults, infants and toddlers fixate objects to guide the trajectory of the hand (Franchak et al., 2011; Franchak & Yu, 2015; Soska, Rachwani, von Hofsten, & Adolph, 2019). With age, children learn how to move the eyes to guide increasingly difficult tasks that require precise movements, such as orienting shapes to fit through openings (Ossmy, Han, Cheng, Kaplan, & Adolph, 2020) and copying letters while writing with a pen (Fears & Lockman, 2019).

Another component of real-world interactivity that is often absent in screen studies is how concurrent tasks compete for attention. For example, controlling locomotion—choosing one's path, avoiding hazards on the ground, and monitoring for collision with other agents—is a subtask we visually guide amidst the demands of completing more primary tasks. In the pick-up-litter task, actors needed to monitor obstacles while also searching for litter (Rothkopf et al., 2007), and actors in the tea-making task needed to walk between different areas of the kitchen to gather supplies (Land et al., 1999). When locomotion is made more challenging, such as when traversing uneven terrain, observers compensated by directing more eye movements toward hazards on the ground (Matthis et al., 2018; 't Hart & Einhauser, 2012; Thomas, Gardiner, Crompton, & Lawson, 2020). But in less demanding environments with uniform surfaces, walkers rely more on peripheral vision to guide locomotion (Franchak & Adolph, 2010). When searching around a room during a scavenger hunt, both 4- to 8-year-old children and adults frequently stepped up, down, and over obstacles without ever fixating the obstacles in advance. Using peripheral vision of obstacles allowed observers to keep their eyes off the ground to better search for the hidden targets around the room. Newly walking infants traverse obstacles without fixating them, even though they typically fixate objects before reaching (Franchak et al., 2011). Even when crossing a narrow bridge over a drop-off, the bridge only warranted a brief glance from infants before crossing; mostly, infants kept their eyes on the destination on the other side of the bridge (Kretch & Adolph, 2017).

4. Embodied factors contribute to ecological validity

The previous section reviewed how interacting with the environment shapes visual exploration. Eye movements guide manual and locomotor actions. However, perception and action are linked in a continual cycle: While perception guides ongoing actions, each movement of the body changes what we see. It follows, then, that what we see depends on how the body can move and whether actors are willing to expend the effort to move. The first section will describe how embodied factors influence how we move the eyes to explore in screen tasks. Yet, screen tasks can reveal little about embodied influences because they do not permit head movement, locomotion, or changes in posture. The remaining sections describe anatomical, biomechanical, and physiological influences on visual exploration in real-world tasks.

Before doing so, it is worth raising the broad question of why we move the eyes the way we do. I began the article describing how eye movements (embedded in head and body movements) facilitate exploration by allow us to direct the higher-resolution fovea to areas of interest in the environment. We stabilize gaze in one location (fixation) between rapid movements that bring the eyes between location (saccades). Counterintuitively, most vertebrate animals still use the fixation-saccade strategy despite lacking foveas (Land, 2006). A more likely explanation for why we move eyes is rooted in their embodied function, which is to stabilize vision in the context of movement. Without the ability to move the eyes, every slight change in body position would shift the direction of gaze, preventing us from continuing to look where we want to look. Furthermore, moving the eyes to compensate for movement prevents blur, which would make it difficult to detect movement of other targets in the environment. Thus, moving eyes facilitate exploration in part because they can scan a scene, but also because they can maintain their position amidst a moving body. For a more extension description of different types of eye movements and their functions, see Land and Tatler (2009).

4.1 Embodied influences in screen tasks

As reviewed above, SET research on visual exploration focuses primarily on how bottom-up and top-down factors influence where people choose to look. Basic phenomena, such as the horizontal and center biases, can be explained in part by those influences: Images contain features that tend to be distributed along the horizon or in the center, and observers explore in directions that match their knowledge about spatial distributions of targets. However, it is important to consider how horizontal and center biases can also be explained by embodied aspects of eye anatomy and biomechanics. Six muscles control eye movements, but horizontal eye movement only require recruiting two muscles, whereas vertical and diagonal eye movements require activation of more eye muscles (Viviani, Berthoz, & Tracey, 1977). This suggests that horizontal eye movements are preferred for their efficiency and the simplicity of control. Furthermore, keeping the eyes at the anatomical limits of their rotation is inefficient for visual exploration (Tatler, 2007). When the eyes are centered in the head, they can move easily in any direction to track a target of interest. But when they are rotated near or at the limit of movement, they can no longer track a target in all directions—dynamic targets can more easily move beyond view when they are not centered.

Although embodied influences are evident even in SET tasks, I argue that screen tasks do more to obscure than reveal how movement influences visual exploration. Indeed, the center bias phenomenon is an artifact of the testing environment. Although "center" can be meaningfully described with respect to an image or a screen, in the three-dimensional environment there is no extrinsic center to view. In real-life visual exploration the observer's head and body movements define what is centered. Generally, in SET studies the anatomy and movements of the body are treated as nuisance variables. Movement is undesirable, and must be minimized to ensure the observer stays in the eye camera's tracking space: The observer sits in a chair, the screen and cameras are oriented to be perfectly positioned relative to the observer, and the head is often immobilized in a chin rest. However, by setting the body's position relative to the stimulus, the experimenter (rather than the observer) has done much of the work of visual exploration. The head and body do not need to coordinate with the eyes to select where to look. Instead, only the eyes need orient within the bounds of the screen. As adults, we may take the ability to coordinate these movements for granted, but for infants orienting with eyes, head, and body can pose a challenge.

Furthermore, SET studies put infants, children, and adults on equal footing (or rather, seating) by giving participants the same visual perspective of each stimulus—the perspective of the photographer or cinematographer who captured the scene. In the real world, height differences and motor skills shape observers' visual perspective differently across development.

4.2 Anatomical influences on visual exploration in real-world tasks

The anatomy of the body shapes our visual perspective of the surrounding world. For humans, the relevant anatomy of the visual system is frontal eyes resting within a head atop an upright, bipedal body. Whereas some animals have laterally positioned eyes that gather visual information in all directions (i.e., they can see the approach of a predator coming from behind), frontal eyes mean that the human visual field is limited to what is in front and to the sides of the body. For all animals, the height of the eyes off the ground determines how much of the ground surface is in view when looking straight ahead. Increasing the height of the eyes from the ground—such as by growing taller—changes the visual perspective. Although this may not be a striking difference for many quadrapedal species (growth lengthens the body more than increases its height while on all fours), growth alters visual perspective in bidpedal humans substantially.

We investigated the role of observer height in visual exploration during obstacle navigation by comparing children and adults stepping up, down, and over obstacles on the ground (Franchak & Adolph, 2010). Obstacles far in the distance are in view, but disappear from view as they become closer to the feet. Looking at a close obstacle requires pointing the head down, which is inefficient for navigating around the room. Consequently, we found that adults fixated obstacles only 31.2% of the time before stepping up, down, or over them, whereas children—who are shorter and could more easily see obstacles on the ground—fixated them 58.9% of the time. Height also predicted age differences in the timing of fixating obstacles before stepping. Children fixated obstacles closer to the moment of stepping, whereas adults fixated obstacles from a greater distance. Presumably, this difference is due to obstacles remaining in view longer for shorter children during the approach. A reasonable objection to claiming that these differences are about height is that developmental differences in motor skill may also play a role—adults are better at controlling locomotion and may rely less on visual information to guide foot placement. However, we found that children were no more likely to make motor errors when fixating obstacles vs guiding locomotion over the same obstacles with peripheral vision. Moreover, when comparing within a group of infants who were proficient crawlers but inexperienced walkers, infants fixated obstacles more often when crawling (with the eyes close to the ground) compared to walking (with the eyes higher off the ground), even though walking was the more challenging skill (Franchak et al., 2011).

Other aspects of body anatomy are relevant to visual exploration. A striking observation from MET studies of infant object play is that infants' visual experiences of objects are shaped by the short length of their arms (Suanda, Barnhart, Smith, & Yu, 2018). When an adult holds an object at arms' reach, it is farther away from the eyes compared to when an infant holds an object. Since infants frequently pick up objects to explore them with both hands and eyes, held objects appear larger in infants' view. One potential benefit of infants' short arms is that held objects are dominant in the field of view at the moment that caregivers speak about them, reducing the visual clutter that may lead infants to link an incorrect word with an object. Hearing a new word paired with seeing an object—while it fills the visual field—provides a strong cue that infants can exploit to learn the reference of a new word.

4.3 Biomechanical influences on visual exploration in real-world tasks

Whereas anatomical influences refer to how the size and configuration of body structures shape what we see, biomechanical influences consider how those structures can move. With the eyes nested within the head and the head resting on top of the body, gaze shifts require coordinated movements of eyes, head, and body (Land, 2004). These multiple degrees of freedom mean that the same gaze shift (i.e., of a particular size and direction) can be accomplished in different ways. Even though the eyes can rotate as much as 50–55 degrees (Guitton & Volle, 1987; Land, 2006), head movements keep the eyes from adopting extreme positions for reasons previously reviewed. In laboratory settings with observers asked to shift gaze between pre-specified targets, smaller shifts of gaze (less than 20–30 degrees) are accomplished primarily through eye movements, while larger shifts of gaze increasingly rely on head movement (Freedman, 2008; Guitton & Volle, 1987). For example, the head might contribute only 1 degree of a 15 degree gaze shift, but may contribute 30 degrees of a 60 degree gaze shift.

Using the eyes and head together to shift gaze involves a coordinated sequence of movements (Land, 2004). Typically, the eyes first make a saccade to orient toward a target before the head moves. As the head rotates to center the target in view, the eyes counter-rotate to remain on the target amidst the movement of the head (termed "vestibulo-ocular reflex", however, I avoid the term because it is an adaptable behavior, not a simple reflex). Otherwise, the eyes would be carried past the target due to the head's rotation. Most likely, the eyes-before-head pattern results from eye movements' faster speed compared with head movements—it is efficient to quickly move the eyes to initially scan a target of interest. If the area warrants further examination, then the slower head and body can reorient the entire view. In other animals, such as peafowl, eye and head movements are equally fast, so the eyes and head move synchronously to a target without the need for compensatory movement (Yorzinski, Patricelli, Platt, & Land, 2015).

Although eye–head coordination may seem stereotyped in more constrained tasks, coordination patterns are flexible and adaptive. For example, observers are more likely to move their heads when shifting gaze to a target when they expect to make a second, subsequent gaze shift in the same direction (Oomen, Smith, & Stahl, 2004), showing that the contribution of the head is planned, not reflexive. Even though the eyes alone typically execute small shifts of gaze in more constrained tasks, Pelz et al. (2001) observed variability in how much the head contributes to gaze shifts in a task where participants arranged blocks in a pattern to match a model. While looking back and forth between the model and the replica they were building, participants' head contribution ranged between 1 and 10 degrees for gaze shifts of 15 degrees. Eyes and head must also be coordinated in the context of locomotion. An early study that used mobile electro-oculography to record eye position during a walk through a college campus found that the eyes rarely rotated more than 15 degrees—far less than the 50–55 degree limit of their range—suggesting that head movements accompanied both smaller and larger shifts of gaze as observers walked (Bahill et al., 1975). A more recent investigation using a modern MET system paired with inertial sensors to measure head rotation found that the head accompanied (temporally) longer gaze shifts, but was recruited less often for quick glances that could be accomplished with eyes only (Tomasi et al., 2016).

To determine how observers adapted the coordination of eyes and head to meet varying task demands in the context of outdoor locomotion, we collected MET measurements of eye position, inertial sensing of head movement, and GPS measurement of walking paths (McGee et al., in

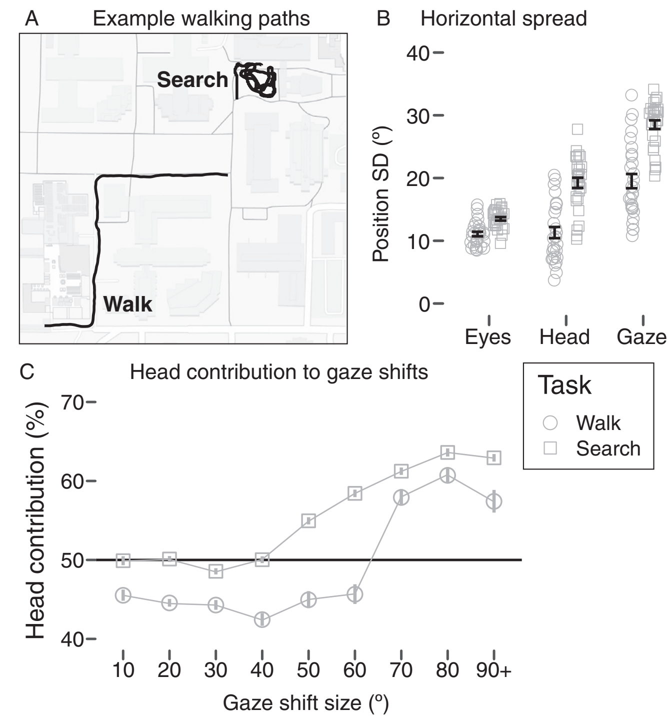

Fig. 3 The coordination of eyes, head, and locomotion in walking and searching tasks (McGee et al., in preparation). (A) Example GPS data showing the walking path of a participant in the walk task and search task. (B) Horizontal spread (SD) of eye position, head position, and gaze position (eyes + head) in degrees according to task. *Gray symbols* indicate data from each participant (*circles* = walking task, *squares* = searching task), with *black error bars* indicating 1 SE around the mean. (C) Relative contribution of the head to gaze shifts of varying magnitude in each task. *Horizontal black line* at 50% indicates that the eyes and head each contributed to half of a gaze shift; *symbols* above the *black line* indicate that the head played a larger role in shifting gaze.

preparation). College-aged adults completed a walking task—when the goal was simply to navigate along a pathway—and a search and retrieval task—in which the goal was to find and pick up hidden targets in a courtyard. Fig. 3A shows example walking and searching paths from one participant's GPS recording, and Video 1 (https://nyu.databrary.org/volume/1147) shows clips from each task with synchronized eye position, head position, and GPS data visualizations. When walking, observers simply stayed on the pathway with little demand on visual attention. In contrast, observers weaved through the courtyard and changed direction while searching for hidden targets. Observers distributed gaze more broadly along the horizontal axis

Video 1 Example excerpts (https://nyu.databrary.org/volume/1147) from the walking task (0–16 s) and the search and retrieval task (17–37 s) on the left side for a participant in (McGee et al., in preparation). The videos show the participant's field of view video overlaid with a gaze cursor (concentric circles). The right side of the figure shows synchronized data for horizontal eye position (*top*), head rotation (*middle*) and GPS (*bottom*). *Vertical red lines* on the eye and head position figures indicate data at the moment corresponding to the video frame. *Positive numbers* indicate rotation to the observer's right, whereas *negative numbers* indicate rotation to the observer's left. The *red asterisk* on the GPS graph shows the observer's current position relative to their entire path (*black line*).

when searching compared to when walking (Fig. 3B). Although both eyes and head rotated more overall during the search task, head movements were adapted by a larger degree compared with eye movements when switching from walking to searching. The two example clips in Video 1 demonstrate the increase in head movement from walking to searching. We also replicated previous findings showing that the head contributes more for larger as opposed to smaller shifts of gaze (Guitton & Volle, 1987), but additionally found that the head contributed more in the search task for both small and large gaze shifts (Fig. 3C). Taken together, the results suggest that eye–head coordination is adapted to meet the informational demands of different tasks within the context of full-body movement.

Over development, infants need not only learn how to prioritize where to look with their eyes, but how to choose what is in view by moving the head. For decades, researchers have known that the burgeoning ability to sustain attention to a target—looking for a long duration with a stable view of the target centered in view—is an important determinant for later learning and school achievement (Razza, Martin, & Brooks-Gunn, 2010; Ruff, Capozzoli, & Weissberg, 1998). Likewise, adults benefit in cognitive tasks when looking with the eyes centered within the head (Nakashima & Shiori, 2014). But only the recent advent of infant MET systems has enabled precise measurement of how infants orient their view with both eye and head movements (Franchak, Smith, & Yu, under review). We annotated the locations of toys and faces in the FOV camera view of 9- to 24-month-olds to determine how infants used head movements to center each type of target (Fig. 4A shows annotated locations of a face

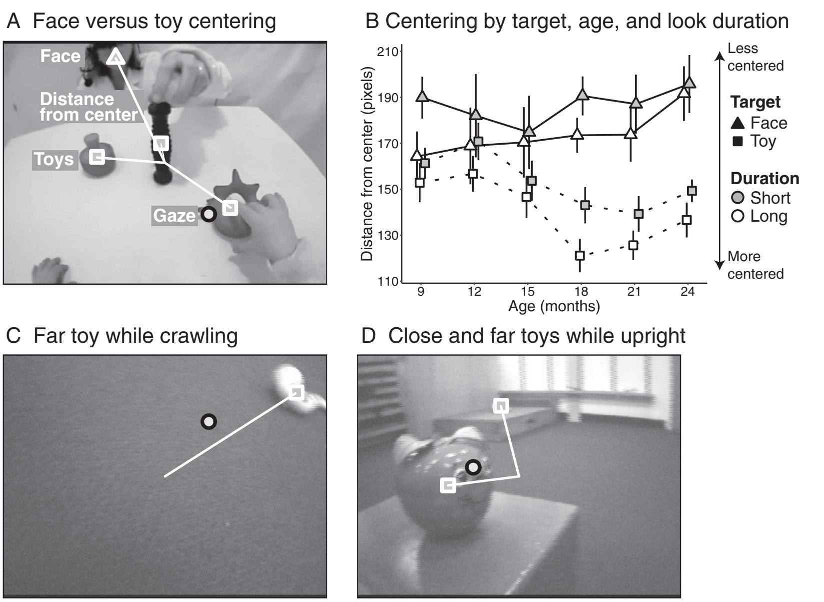

Fig. 4 Head movements structure infants' view of toys and faces. (A) Example frame from an infant's field of view while sitting at a table. *Triangle* indicates the location of the face and squares indicate the location of three toys. The *white lines* show the distance between each target and the center of view (*shorter lines* indicate better centering). (B) Age-related changes in centering of faces and toys during a seated play task (Franchak et al., under review). With age, toys (*squares*) became better centered than faces (*triangles*). Better centering of all targets was observed during longer looks (*white symbols*) compared to shorter looks (*gray symbols*). (C) and (D) Examples of object centering for a far toy while crawling (C) and a close and far toy while standing upright (D) from Luo and Franchak (in preparation).

and three toys). We measured the distance from each target to the center of the view; smaller distances (white lines in Fig. 4A) indicate better target centering. With age, infants preferentially centered toys in view at the expense of faces (smaller values for square symbols vs triangle symbols in Fig. 4B). Most likely, centering toys in view allowed older infants to better reach for and manipulate objects, and faces could be viewed with a quick glance up (without head movement). Importantly, targets were better centered in view during longer (white symbols in Fig. 4B) vs shorter looks (gray symbols in Fig. 4B)—suggesting that over development infants learn how to coordinate eye and head movements to sustain attention.

4.4 Physiological influences on visual exploration in real-world tasks

So far, I have discussed how visual exploration depends on the movement capabilities of the body—shaped by the body's anatomy and biomechanics. Eye, head, and body movements are adapted to meet the informational needs of different tasks. A final, related consideration is how observers *choose* to explore given physiological demands. Exploration is voluntary movement, and movement requires effort. Eye movements are less energetically costly compared to head movements, which are less costly than whole-body movements. Efficient visual exploration means moving no more than necessary, but what is necessary depends on the informational demands of a task. As already reviewed, children and adults often refrain from pointing their heads down to look at obstacles while walking (Franchak & Adolph, 2010). Since they can navigate successfully without fixating obstacles, energy is conserved by keeping the head in a neutral position rather than tilting it up and down with each upcoming obstacle. But when locomotion becomes more demanding (e.g., an uneven, rocky path), it may be worth expending the energy to look down to prevent costly trips and falls (Matthis et al., 2018).

In recent work, we examined how observers chose to gather task-relevant information when the act of looking was made more or less effortful (Luo, Lat, & Franchak, in preparation). Adapting the model-copying task from Pelz et al. (2001), we presented participants with models of randomized spatial configurations of magnets on a grid and asked participants to copy each pattern using a second set of magnets in their workspace. The model was placed 45, 90, 135, or 180 degrees to the left or right of the participants to increase the motor cost of looking. At 45 degrees, a slight head turn was sufficient to look, whereas at larger angles a large head turn or a head-plus-

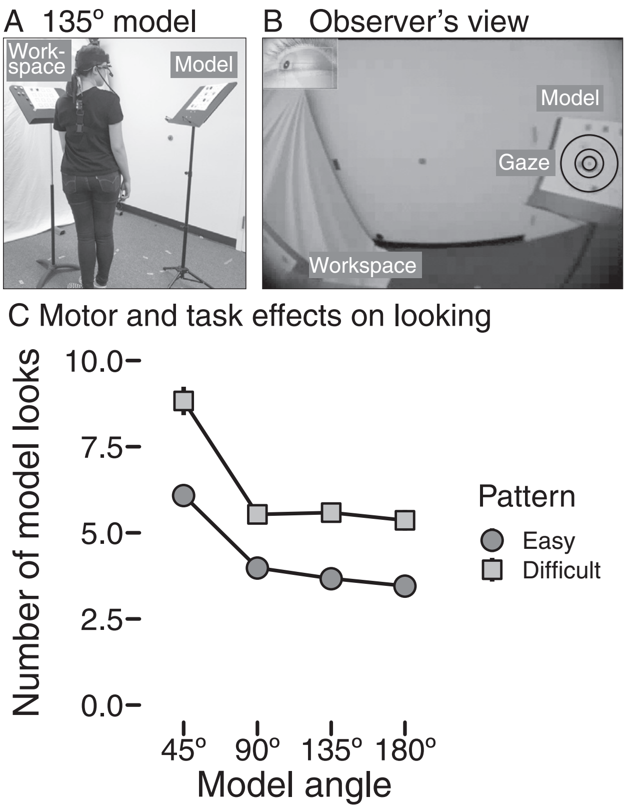

Fig. 5 Motor and task difficulty affect participants' willingness to visually explore (Luo et al., in preparation). (A) Participant turning the head and torso away from the workspace to look at a model placed at a 135 degree rotation. (B) The participant's field of view and gaze location while turning to view the 135 degree model. (C) The number of looks to the model depended on the model's angle (*x*-axis) and the difficulty of the pattern.

body movement was required. Fig. 5A and B shows a participant turning from the workspace to look at the 135 degree model. On some trials, participants completed "easy" trials in which all magnets were the same color, meaning that they only need to memorize the spatial locations of magnets. On "difficult" trials magnets were different colors, so participants needed to remember both the location and color of magnets to copy the pattern correctly. Fig. 5C shows that participants looked more often to the model when it was motorically easier (requiring only a slight shift of the eyes and head at 45 degrees) vs when it was more difficult to look (moving head and/or body for models 90 degrees and greater). Overall, participants looked more often at the model when copying the difficult pattern, but pattern difficulty did not interact with motor difficulty. Most likely, participants chose to rely more on their memory of the pattern when the model was far to reduce costly movement, but did so to a lesser extent when the pattern was more difficult to memorize. Ongoing work investigates how children, with both immature memory and motor skills, navigate these informational and energetic tradeoffs.

The prior examples showed how increasing motor costs of looking within a posture (e.g., standing upright) discourage visual exploration, even when it means deciding not to look at information relevant to ongoing tasks. Motor costs also govern how infants visually explore in different postures. As infants master new skills—sitting, crawling, standing, and walking—they can transition between different body positions that change their visual perspective. We used METs to measure the approximate head-centered visual fields of 13-month-olds while walking vs crawling (Kretch, Franchak, & Adolph, 2014). While walking, infants can see farther into the distance and their visual field captures higher locations off the ground. In contrast, while crawling infants' view of distant and higher places is compromised, and their view is often dominated by the floor. Although infants can crane their necks to look up while crawling, rotating the head against gravity to an extreme angle is effortful and infants choose to do so only intermittently. By 12 months, infants spend about 50% of the time on the floor in a typical day (Franchak, 2019), rendering much of the world "above the ground" difficult to see, especially when crawling. The motor costs of looking in different postures have consequences on visual exploration during play. We found that, while crawling, 12-month-old infants fixated caregivers' faces less often compared to when sitting and upright (Franchak et al., 2018). Possibly, changes in infant posture over the first year of life relate to the frequency that faces are seen in daily life. Field-of-view camera recordings (without eye tracking) of infants in their homes found that faces are in view more frequently in the first months of life—when infants spend more time held off the ground or laying on their backs—but decline in frequency as infants become older and spend more time crawling and sitting on the ground (Fausey, Jayaraman, & Smith, 2016; Jayaraman, Fausey, & Smith, 2015).

Of course, the motor costs of looking depend not just on the observer's body posture, but the observer's spatial position relative to potential gaze targets. Even when sitting and upright, faces of relatively taller adults may not be easy to view. Fig. 1B shows an example of how an adult's face is out of view for a sitting infant with the head in a neutral position. Changing the height of the caregiver's face relative to the infant affects the rate of face looking. Across infant postures, infants looked at seated caregivers' faces more often compared to standing caregivers' faces (Franchak et al., 2018). The distance of a gaze target also determines how easy it is to view. Yamamoto, Sato, and Itakura (2019) used caregiver-worn METs to record how often infants and caregivers made eye contact with each other. Eye contact was difficult to achieve when infants and caregivers were either too close or too far from one another; the optimal distance for eye contact was about 1 m. Infants' view of objects during play also depends on the distance between infants and objects as well as infants' posture (Luo & Franchak, in preparation). Infants could keep close toys (within 1 m) well-centered in view regardless of their body position (crawling, sitting, and upright). However, infants struggled to center distant toys in view while crawling compared with while sitting and upright. Fig. 4C and D shows examples of infants' view of close and far objects when crawling vs upright. Taken together, these results show how the natural, spatial configurations of both observers and environments influence visual exploration.

5. Conclusion

To summarize, visual exploration is the active acquisition of information through coordinated movements of the eyes, head, and body. Most studies of visual exploration employ screen eye trackers that record the gaze of a seated observer who looks at computer display, however, researchers are increasingly using mobile eye trackers to record the gaze of moving observers in real-world tasks. Regardless of the type of eye tracker used, studies vary in their ecological validity depending on how they (1) present realistic visual stimuli, (2) afford opportunities for interaction, and (3) consider the role of the body in shaping where we look. Throughout the article, I reviewed examples to demonstrate how understanding the mechanisms of visual exploration requires more ecologically valid tasks. Observers continually reprioritize bottom-up and top-down features, but this behavior emerges only when viewing complex stimulus sets that vary in their composition (Kadooka & Franchak, in press). Attention to social agents and objects shifts from moment to moment depending on task relevance, but only if observers are engaged in a task (Franchak et al., 2018; Land et al., 1999; Rothkopf et al., 2007). The anatomy, biomechanics, and physiology of the body shape our perspective and constrain what we see, but only in observers who are free to move their bodies (Franchak & Adolph, 2010; Kretch et al., 2014; Suanda et al., 2018). Studying visual exploratory behavior in participants across development also improves generalizability. Infants, children, and adults vary in the types of tasks they perform, their expertise in choosing what is relevant to a task, and in their body size and coordination, demonstrating the flexibility of the visual exploratory system across the lifespan.

Generally speaking, mobile eye tracking studies provide advantages for all three factors that contribute to ecological validity: Observers view real-world environments, engage in everyday tasks, and move their bodies. Without MET tasks, we would continue to overlook the role the body plays in shaping visual exploration. Although I separated embodied influences into three categories—anatomical, biomechanical, and physiological—it is important to note that the three are not independent. For example, the energy required to execute a given exploratory movement (e.g., raising up the head to look from a crawling posture) depends on the anatomy of the structures that need to move and the individual's ability to control the muscles needed to raise the head. In turn, the observer's willingness to execute the movement depends on the amount of effort (cost) with respect to the importance of the exploratory movement to the task at hand (gain). Over development, all three aspects of embodied influences change dramatically, with consequences on visual exploration that we are only now beginning to understand.

What is evident across both SET and MET studies of visual exploration is that flexibility is the rule, not the exception. This should not be surprising, given that the visual exploratory system has many degrees of freedom. The eyes and head can move in concert to center a target in view, or the eyes can move alone to fixate a target on the edge of the head-centered view. But SET and MET systems only measure gaze. Focusing solely on where people look does not reveal how looking was accomplished. More research that simultaneously records head movements using motion trackers or calculates the centering of targets will help identify the flexibility in how looking is accomplished through different coordinations of eyes and head.

More work is also needed to understand how and why observers adapt eye, head, and body movements to the demands of different tasks. Although the few naturalistic tasks that have been studied, such as driving, food preparation, and walking, show how visual exploration differs compared with passively viewing screens, we are still far from understanding just how much visual exploration needs to adapt across the diverse activities of daily life. Do observers learn specific coordination patterns for each different task (Hayhoe, 2000, e.g., visual routines), or do they adapt eye and head movements on the fly from a general-purpose pattern? Over development, how do infants and children learn to adapt their visual exploration in the context of changing embodied influences, cognitive abilities, and task expertise? Although challenging, these questions will bring us closer to understanding how visual exploration meets the functional needs of daily life in observers of all ages.

Acknowledgments

The author thanks Chuan Luo, Kellan Kadooka, and Brianna McGee for their help in providing comments on this chapter.

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*, 1–7. https://doi.org/10.1371/journal.pone.0085701.

Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology & Visual Science*, *14*(6), 468–469.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*, 185–207. https://doi.org/10.1109/TPAMI.2012.89.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*, 1–15. https://doi.org/10.1167/9.3.6.

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.

Einhauser, W., Schumann, F., Bardins, S., Bartl, K., Boning, G., Schneider, E., & Konig, P. (2007). Human eye-head co-ordination in natural exploration. *Network: Computation in Neural Systems*, *18*, 267–297.

Einhauser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*, 1–26.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107. https://doi.org/10.1016/j.cognition.2016.03.005.

Fears, N. E., & Lockman, J. J. (2019). Using head-mounted eye-tracking to study handwriting development. *Journal of Motor Learning and Development*, *1*, 1–17.

Foulsham, T., Chapman, C., Nasiopoulos, E., & Kingstone, A. (2014). Top-down and bottom-up aspects of active search in a real-world environment. *Canadian Journal of Experimental Psychology*, *68*, 8–19. https://doi.org/10.1037/cep0000004.

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.

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. https://doi.org/10.1016/j.visres.2011.07.002.
Franchak, J. M. (2019). Changing opportunities for learning in everyday life: Infant body position over the first year. *Infancy*, *24*, 187–209.
Franchak, J. M. (2020). Looking with the eyes and head. In J. B. Wagman & J. J. C. Blau (Eds.), *Perception as information detection: Reflections on Gibson's ecological approach to visual perception* (pp. 205–221). Routledge.
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. https://doi.org/10.1016/j.visres.2010.09.024.
Franchak, J. M., Heeger, D. J., Hasson, U., & Adolph, K. E. (2016). Free viewing gaze behavior in infants and adults. *Infancy*, *21*, 262–287. https://doi.org/10.1111/infa.12119.
Franchak, J. M., & Kadooka, K. (under review). Age differences in orienting to faces in dynamic scenes depend on face centering, not visual salience. Manuscript Submitted for Publication.
Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant–caregiver social looking during locomotor free play. *Developmental Science*, *21*, e12626. https://doi.org/10.1111/desc.12626f.
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. https://doi.org/10.1111/j.1467-8624.2011.01670.x.
Franchak, J. M., Smith, L. B., & Yu, C. (under review). Developmental changes in how head orientation structures infants' visual experiences. Manuscript submitted for publication.
Franchak, J. M., & Yu, C. (2015). Visual-motor coordination in natural reaching of young children and adults. In *Proceedings of the 37th annual meeting of the Cognitive Science Society*.
Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*, 160–170. https://doi.org/10.1016/j.cognition.2008.11.010.
Freedman, E. G. (2008). Coordination of the eyes and head during visual orienting. *Experimental Brain Research*, *190*(4), 369.
Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology*, *39*, 1–41.
Gibson, J. J. (1966). The senses considered as perceptual systems. Boston: Houghton-Mifflin.
Gibson, J. J. (1979). The ecological approach to visual perception. Boston, MA: Houghton Mifflin Company.
Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy*, *14*, 550–562. https://doi.org/10.1080/15250000903144199.
Guitton, D., & Volle, M. (1987). Gaze control in humans: Eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of Neurophysiology*, *58*, 427–459.
Hayhoe, M. M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, *7*, 43–64.
Hayhoe, M. M., Shrivastava, A., Mruczek, R. E. B., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*, 49–63.
Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford: Elsevier. https://doi.org/10.1016/b978-008044980-7/50027-6.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*(1), 243–271.
Jayaraman, W., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PLoS ONE*, *10*, e0123780. https://doi.org/10.1371/journal.pone.0123780.
Jing, M., Kadooka, K., Franchak, J. M., & Kirkorian, H. (2019). The effect of comprehensibility on saliency-based gaze prediction for children and adults watching sesame street. In *Poster presented at the biennial meeting of the Cognitive Development Society, Louisville, KY*.
Jovancevic-Misic, J., & Hayhoe, M. M. (2009). Adaptive gaze control in natural environments. *The Journal of Neuroscience*, *29*(19), 6234–6238. https://doi.org/10.1523/JNEUROSCI.5570-08.2009.
Kadooka, K., & Franchak, J. M. (in press). Developmental changes in infants' and children's attention to faces and salient regions vary across and within video stimuli. *Developmental Psychology*.
Kadooka, K., & Franchak, J. M. (in preparation). Developmental changes in adult-like gaze across varied stimulus videos. Manuscript in Preparation.
Keane, T. P., Cahill, N. D., Tarduno, J. A., Jacobs, R. A., & Pelz, J. B. (2014). Computer vision enhances mobile eye-tracking to expose expert cognition in natural-scene visual-search tasks. *Human Vision and Electronic Imaging XIX*, *9014*, 90140F.
Kirkorian, H. L., & Anderson, D. R. (2018). Effect of sequential video shot comprehensibility on attentional synchrony: A comparison of children and adults. *Proceedings of the National Academy of Sciences*, *115*, 9867–9874. https://doi.org/10.1073/pnas.1611606114.
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. https://doi.org/10.1111/j.1467-8624.2011.01719.x.
Kretch, K. S., & Adolph, K. E. (2015). Active vision in passive locomotion: Real-world free viewing in infants and adults. *Developmental Science*, *18*, 736–750. https://doi.org/10.1111/desc.12251.
Kretch, K. S., & Adolph, K. E. (2017). The organization of exploratory behaviors in infant locomotor planning. *Developmental Science*, *20*, e12421.
Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, *85*, 1503–1518. https://doi.org/10.1111/cdev.12206.
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*, 537–555. https://doi.org/10.1037/dev0000080.
Land, M., & Tatler, B. (2009). Looking and acting: Vision and eye movements in natural behaviour. Oxford University Press.
Land, M. F. (2004). The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research*, *159*, 151–160. https://doi.org/10.1007/s00221-004-1951-9.
Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, *25*, 296–324.
Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*, 742–744.
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.
Luo, C., & Franchak, J. M. (in preparation). Head and body structure infants' visual experiences during mobile, naturalistic play. Author manuscript.
Luo, C., Lat, Y., & Franchak, J. M. (in preparation). Trade-off between eye-head movement and visual working memory. Author manuscript.

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*, 485–498. https://doi.org/10.1109/tcds.2017.2696439.
Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology*, *28*, 1224–1233.
McGee, B., Blanch, G., & Franchak, J. M. (in preparation). Coordination of eye and head movements for outdoor visual exploration differs by task. Author Manuscript.
McGee, B., Michaels, K., & Franchak, J. M. (2020). Associative learning modulates infants' attentional selection in a free-viewing task. In *Poster presented at the 2020 International Congress on Infant Studies*.
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. https://doi.org/10.1007/s12559-010-9074-z.
Nakashima, R., & Shiori, S. (2014). Why do we move our head to look at an object in our peripheral region? Lateral viewing interferes with attentive search. *PLoS ONE*, *9*, e92284.
Nayar, K., Franchak, J. M., Adolph, K. E., & Kiorpes, L. (2015). From local to global processing: The development of illusory contour perception. *Journal of Experimental Child Psychology*, *131*, 38–55.
Oomen, B. S., Smith, R. M., & Stahl, J. S. (2004). The influence of future gaze orientation upon eye-head coupling during saccades. *Experimental Brain Research*, *155*, 9–18.
Ossmy, O., Han, D., Cheng, M., Kaplan, B. E., & Adolph, K. E. (2020). Look before you fit: The real-time planning cascade in children and adults. *Journal of Experimental Child Psychology*, *189*, 104696.
Peacock, C. E., Hayes, T. R., & Henderson, J. M. (2019). Meaning guides attention during scene viewing, even when it is irrelevant. *Attention, Perception, & Psychophysics*, *81*(1), 20–34.
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. https://doi.org/10.1007/s002210100745.
Razza, R. A., Martin, A., & Brooks-Gunn, J. (2010). Associations among family environment, sustained attention, and school readiness for low-income children. *Developmental Psychology*, *46*, 1528–1542.
Rider, A. T., Coutrot, A., Pellicano, E., Dakin, S. C., & Mareschal, I. (2018). Semantic content outweighs low-level saliency in determining children's and adults' fixation of movies. *Journal of Experimental Child Psychology*, *166*, 293–309. https://doi.org/10.1016/j.jecp.2017.09.002.
Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, *7*, 1–20. https://doi.org/10.1167/7.14.16.
Ruff, H. A., Capozzoli, M., & Weissberg, R. (1998). Age, individuality, and context as factors in sustained visual attention during the preschool years. *Developmental Psychology*, *34*, 454–464. https://doi.org/10.1037/0012-1649.34.3.454.
Schumann, F., Einhäuser, W., Vockeroth, J., Bartl, K., Schneider, E., & Koenig, P. (2008). Salient features in gaze-aligned recordings of human visual input during free exploration of natural environments. *Journal of Vision*, *8*(14), 12–12.
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. https://doi.org/10.1016/j.cub.2010.02.032.
Smeets, J. B. J., Hayhoe, M. M., & Ballard, D. H. (1996). Goal-directed arm movements change eye-head coordination. *Experimental Brain Research*, *109*, 434–440.
Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*, 1–24. https://doi.org/10.1167/13.8.16.

Soska, K. C., Rachwani, J., von Hofsten, C., & Adolph, K. E. (2019). Infants plan prehension while pivoting. *Developmental Psychobiology*, *61*(7), 1048–1063.
Suanda, S. H., Barnhart, M., Smith, L. B., & Yu, C. (2018). The signal in the noise: The visual ecology of parents' object naming. *Infancy*, *24*, 455–476.
't Hart, B. M., & Einhauser, W. (2012). Mind the step: Complementary effects of an implicit task on eye and head movements in real-life gaze allocation. *Experimental Brain Research*, *223*, 233–249.
't Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., Konig, P., et al. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition*, *17*, 1132–1158.
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, 1–17. https://doi.org/10.1167/7.14.4.
Tatler, B. W., Hirose, Y., Finnegan, S. K., Pievilainen, R., Kirtley, C., & Kennedy, A. (2013). Priorities for selection and representation in natural tasks. *Philosophical Transactions of the Royal Society B*, *368*, 20130066.
Taya, S., Windridge, D., & Osman, M. (2013). Trained eyes: Experience promotes adaptive gaze control in dynamic and uncertain visual environments. *PLoS ONE*, *8*(8), e71371.
Thomas, N. D. A., Gardiner, J. D., Crompton, R. H., & Lawson, R. (2020). Look out: An exploratory study assessing how gaze (eye angle and head angle) and gait speed are influenced by surface complexity. *PeerJ*, *8*, e8838.
Tomasi, M., Pundlik, S., Bowers, A. R., Peli, E., & Luo, G. (2016). Mobile gaze tracking system for outdoor walking behavioral studies. *Journal of Vision*, *16*(3), 1–15.
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*, 766–786. https://doi.org/10.1037/0033-295X.113.4.766.
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*, 1–16. https://doi.org/10.1167/12.13.3.
Tummeltshammer, K., & Amso, D. (2018). Top-down contextual knowledge guides visual attention in infancy. *Developmental Science*, *21*, e12599.
Turano, K. A., Geruschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research*, *43*, 333–346.
van Renswoude, D. R., Johnson, S. P., Raijmakers, M. E. J., & Visser, I. (2016). Do infants have the horizontal bias? *Infant Behavior and Development*, *44*, 38–48.
van Renswoude, D. R., van den Berg, L., Raijmakers, M. E. J., & Visser, I. (2019). Infants' center bias in free viewing of real-world scenes. *Vision Research*, *154*, 44–53.
Viviani, P., Berthoz, A., & Tracey, D. (1977). The curvature of oblique saccades. *Vision Research*, *17*(5), 661–664.
Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U., & Heeger, D. J. (2012). Temporal eye movement strategies during naturalistic viewing. *Journal of Vision*, *12*, 1–27. https://doi.org/10.1167/12.1.16.
Wang, Q., Bolhuis, J., Rothkopf, C. A., Kolling, T., Knopf, M., & Triesch, J. (2012). Infants in control: Rapid anticipation of action outcomes in a gaze-contingent paradigm. *PLoS ONE*, *7*(2), e30884.
Wass, S. V., & Smith, T. J. (2015). Visual motherese? Signal-to-noise ratios in toddler-directed television. *Developmental Science*, *18*, 24–37. https://doi.org/10.1111/desc.12156.
Yamamoto, H., Sato, A., & Itakura, S. (2019). Eye tracking in an everyday environment reveals the interpersonal distance that affords infant-parent gaze communication. *Scientific Reports*, *9*(1), 1–9.
Yarbus, A. L. (1967). Eye movements and vision. New York: Plenum.

Yorzinski, J. L., Patricelli, G. L., Platt, M. L., & Land, M. F. (2015). Eye and head movements shape gaze shifts in indian peafowl. *Journal of Experimental Biology*, *218*(23), 3771–3776. https://doi.org/10.1242/jeb.129544.

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*, e79659. https://doi.org/10.1371/journal.pone.0079659.