Title
The Influence of Target Regularity and Task on Screen-Based and Real-World Visual Exploration

Author
McGee, Brianna Lynn

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

The Influence of Target Regularity and Task on Screen-Based and Real-World Visual Exploration

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Brianna L. McGee

September 2022

Dissertation Committee:

Dr. John M. Franchak, Chairperson
Dr. Chandra Reynolds
Dr. David Rosenbaum

The Dissertation of Brianna L. McGee is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am grateful for the wisdom and support of my advisor, John Franchak. John, you made a place for me when I thought no one would, that meant everything to me. I made it to the finish line because of you and your persistent and unyielding belief in me (and your seemingly endless patience). Thank you for everything.

This dissertation would not have been possible without the love and support of my family. I dedicate my dissertation to each of you.

For my incredible husband Stephen: please know that it was your strength and faith in me that kept me going– this dissertation would not have been possible without you. I am better for knowing you and I am far better for loving you.

For my parents, Lynn and Dale: thank you for everything, but especially for my siblings, three of my favorite people in all the world.

For my godparents, Beth and Chris: I have no words to express how grateful I am to you both for your unending love and support. And I will always love you more.

Each of you were my first teachers and my biggest supporters. I would not be here without you.

Momma: you are my hero. Pradeep, pradeep, pradeep, pradoo.

ABSTRACT OF THE DISSERTATION

The Influence of Target Regularity and Task on Screen-Based and Real-World Visual Exploration

by

Brianna L. McGee

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, September 2022
Dr. John M. Franchak, Chairperson

Visual attention is thought to be influenced by two categories of factors, those that are *top-down* and endogenous (e.g., prior knowledge, current goals, etc.) and those that are *bottom-up* and exogenous (e.g., properties of external stimuli such as color, contrast, and orientation). The primary aim of the three studies included in this dissertation is to assess how particular types of top-down information (underlying environmental regularities and prior knowledge) influence attention across development (Chapter 2) and in conjunction with motor factors (Chapters 3 and 4). Chapter 2 focuses on the development of the influence of top-down information on infant free-viewing of dynamic scenes. Chapters 3 and 4 both assess aspects of visual attention during real-world search. Specifically, Chapter 3 investigates the influence of underlying environmental regularities on real-world search efficiency and Chapter 4 asks how exploratory eye and head movements are differentially adapted to the varying demands on attention created by different tasks and environments.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Real world visual exploration is a multi-system process that engages both visual and motor systems [65]. Prior work on visual attention has revealed two main categories of influences (bottom-up and top-down) although they are not always easy to disentangle and may in fact overlap. As I will review, bottom-up influences refer to those that are exogenous to the individual whereas top-down influences are endogenous and varied. Furthermore, motor factors influence where people look because the act of "looking" is inherently a motor action that must be controlled by a physical body. Both the attentional and motor systems change over development and these changes open up new ways to view and interact with the world.

The goal of this dissertation is to look at how particular aspects of top-down information (regularities and tasks) influence attention over development, as well as how they interact with motor factors. Chapter 2 looks at the role of regularity detection in guiding visual exploration and how this process likely changes throughout infancy. Chapter

3 investigates how environmental regularities and motoric factors shape adults' real-world exploration. Finally, Chapter 4 aims to assess how a second type of top-down information (task) influences real-world exploration.

1.1 Influences on visual attention

The visual world is cluttered and complex which creates a challenge for the many systems involved in attention. How do we know what to (visually) attend to and when to do so? Attentional selection is the process of determining what locations in the environment, stimuli, or information to attend to. The human brain is thought to have a limited capacity for the processing of sensory information, which is among the reasons that attention is thought to be a selective process [8]. Research indicates that this is an ability that develops [14, 60, 33], as there are age related changes in both the mechanisms of visual attention (e.g., attentional selection) as well as factors that influence said attention (e.g., knowledge). Additionally, it is understood that attentional selection is a developing ability and the prioritization of certain visual information over others not only changes across the lifespan, but is also contextually dependent [50].

Early studies of visual attention began with simple screen-based stimuli (e.g., lines oriented in different directions, basic shapes, etc.) on a neutral background to isolate which factors influence where we look [100]. From these foundational studies the visual attention literature has determined two primary categories that influence attention, bottom-up and top-down attentional selection. Bottom-up attentional selection refers to attention that is driven by exogenous factors such as an object's contrast, luminance or flicker– taken together these factors are sometimes referred to as the "physical salience" of an object. Conversely, top-down attentional selection is endogenous to the individual such as their prior knowledge or goals.

1.1.1 Bottom-up influences on visual attention

Bottom-up attentional selection, or attention to physically salient objects or regions, is thought to be predictive of human visual behaviors under certain circumstances. Computational models of human visual attention have successfully replicated performance on screen based visual search tasks using an exclusively saliency-based, or bottom-up, model (e.g., saliency maps; [46]). In these purely stimulus driven models, objects are deconstructed into feature maps of their representative channels of information, such as the orientation of lines, intensity of color, or contrast. From these feature maps a single saliency map is then created which delineates between areas of high and low salience. This in turn allows for attentional selection mechanisms to then direct attention to areas of high physical salience without any top-down direction.

However, this model best replicates static image viewing or search for a single feature (search for items that "pop-out"). This so called pop-out effect is thought to capture attention because these features generate strong "attend to me" signals [88]. This attentional capture likely reflects the early processing of basic visual information such as contrast or orientation. Some theories, such as the guided model of visual search [10, 109] assert that attentional control mechanisms will first prioritize bottom-up factors for parallel processing and after that initial selection top-down factors such as personal goals will then influence further serial processing.

Models of visual attention are more predictive of human behavior when they include temporal elements such as motion or flicker [73]. Flicker is differentiated from motion in that it describes changes to a stimulus' luminence or color while motion refers to movement [23, 73]. Motion can induce flicker, and flicker can be perceived as motion without any movement actually occurring (flicker induced motion) [23]. Similarly, objects that flicker at particular rates can also suppress the perception of motion (flicker induced motion suppression). In dynamic scenes existing people are free to move at will and exert their influence on the elements within the scene and new objects and people at times appear within the scene. All of which creates changes in both motion and flicker. These temporal elements are not possible in static scenes, but represent the dynamic nature of the natural world quite well (with the exception of artificial film elements such as scene cuts). These temporal characteristics are the only known visual features that influence exogenous attention regardless of endogenous factors [73]. However, temporal characteristics such as motion and flicker are highly correlated with top-down factors such as the semantics associated with the scene.

Development of attention to bottom-up features

Visual attention early in infancy is thought to be driven by primarily exogenous factors such as the complexity or physical salience of a stimulus [13]. These factors are also thought to be related to how quickly they draw the infants' attention (attention getting – physical salience of the stimulus) and how long infants look at an object (attention holding – complexity of the stimulus).

Evidence of a weak pop-out effect on visual attention has been documented in infants as young as 2 months [14]. The pop-out effect on visual attention is thought to be driven by exogenous factors and by three months of age infants have shown a pop-out effect on the same time scale as adults (milliseconds)– even with several competing distractors [2]. These findings suggest that within just a few months bottom-up attentional processing is online and is comparable to that of adults in terms of how quickly items that "pop-out" can "grab" infant attention.

Between 4 and 8 months, an object's physical salience appears to have a decreasing influence on infant's attentional selection. When presented with multi-object arrays, 4-month-old infants looked first and longest to the most physically salient item while 6 and 8-month-old infants spent longer looking to the faces [60]. Similarly, the fixations made by 3-month-old infants were best predicted by a perceptual salience model, while those made by 9-month-old infants were best predicted by the location of faces, even after correcting for the videos in which the faces were highly physically salient [33]. These findings demonstrate that by the second half of the first year the factors that influence infant attention begin to change. However, in studies such as these, top-down attention is measured by infant attention to faces and previous work has demonstrated that faces tend to be highly physically salient [44, 107].

1.1.2 Top-down influences on visual attention

Top-down attention involves using what you know to inform visual guidance for what you are doing. This knowledge comes in many forms, from regularities detected in the environment [48], contextual or semantic knowledge [15], or even previous experience with the demands of a particular task [80]. Furthermore, decades of research has indicated there is a bi-directional relationship between attention and memory. Such that, what is attended to is further processed and thus is likely to create a stronger representation in memory *and*, knowledge is both stored and retrieved from memory and influences what is attended [12, 93]. Furthermore, the utility of stored knowledge is constrained by both access to and the quality of the representation [70].

Attentional (or target) templates are short term descriptions of information that are thought to be stored in visual working memory (VWM) that are used as a mechanism for attentional control as they guide attention to inputs in the visual field that match the description [20, 17]. The specificity of the attentional template has been found to influence search behaviors [70]. When prompted with pictures of the target (as opposed to words) participants in one study fixated on fewer regions of the scene before finding their target. Overall, participants made fewer, shorter, fixations in the picture condition and consequently, the overall duration of the search task was shorter.

Semantic knowledge influences temporal dynamics of visual exploration such as *when* particular categories of objects are fixated [15]. Time series analysis of gaze behavior during a static visual search task has demonstrated that under typical visual search conditions (e.g., instructions for the target given before the trial onset) initial orienting was biased towards visually similar objects. However, when participants were given a preview of the trial this pattern was reversed and initial orienting was biased towards the semantically relevant object. Furthermore, semantic knowledge can be used to guide category search (search for any item that falls within a particular category) on a time scale similar to that of exemplar search (search for a particular item; [75]).

Semantic knowledge also influences spatial elements of visual exploration [105]. When search targets are embedded in natural scenes repeated search trials showed a drastic improvement in search times even when there had been several other search trials in between. Interestingly, no such benefit was found from previewing the target object or even when asked to memorize scenes involving the target before the search task. The authors include that while memory-guided search can be beneficial for repeated search, viewing the target before search does not appear to induce memory guided search and instead participants rely on general semantic knowledge.

Similar to findings from screen-based studies, a virtual reality based search task found that semantic knowledge influences attention allocation to relevant surfaces [68]. Such that, within the first trials participants largely restricted their fixations to counter surfaces. The authors concluded that the adult participants used their prior knowledge of typical scene structure (e.g., objects tend to be placed on surfaces) to guide their search from the very first trial.

Importantly, in studies such as those described above, semantic knowledge was never formally assessed for the individual participants. Instead, relevant semantic knowledge (e.g., where to look for a jam jar in a kitchen scene) was simply assumed. For older children and adults assumptions such as these are less of a concern as some assumptions can be reasonably made as to their general contextual and semantic knowledge. However, for younger children or infants assumptions such as these can be problematic.

Development of attention to top-down features

Knowledge of semantic categories (e.g., faces, objects) is used to guide attention in an endogenous, top-down manner. However, attention to semantic categories in infancy does not necessarily reflect semantic knowledge. It is a well documented phenomenon that newborn infants tend to preferentially look to faces and face-like stimuli, although this preference appears to decline around the first month [49]. It is unclear whether this preference is due to innate knowledge about faces (e.g., reflects top-down attentional selection) or because of their physical salience (areas of high interest, such as eyes, tend to be contain a great deal of contrast and movement and as such rate highly in physical salience). A possible explanation is that humans are born with an early face processing system that ensures enough experience with face like patterns to support the later developing cortical face-processing system used by older children and adults

Within the second half of the first year, infants shown more endogenous control over their visual attention. This is thought to be due to a number of internal factors including the development of frontal areas of the brain involved in alertness, spatial orienting, and object recognition [14]. Around 6 months infants appear to demonstrate a revived interest in looking to faces. In a static array of complex images, 6 and 8-month-old infants looked first and longest to faces even when the competitors were much more physically salient [60]. Importantly, 4-month-old's only looked to faces in arrays of two images when the second image was not particularly salient. Other work has reported a strong correlation with increased face looking in a dynamic display to performance in a static visual search task [34]. This suggests that developing attentional abilities (such as attentional inhibition and shifting) may facilitate social attention [49].

Regularity detection

Real-world studies of visual attention have demonstrated that adults can detect regularities present in their environment which in turn can lead to an attentional bias to spatial locations likely to hold their target [111]. Participant's initial head turns were overwhelmingly biased in the direction of the "rich" quadrant (which contained the target in 50% of trials). This indicates that the participant's were sensitive to the location probability of the target. However, the search area used in this study was relatively small, the participants were able to see the entire search area without moving their bodies. What remains to be seen is whether similar attentional modulation or search efficiency would be apparent in studies of real-world search that better replicate search in natural environments– that is search in a cluttered visual environment that necessitates full body movements.

Development of regularity detection in guiding attention There are elements in the natural world that change in predictable ways over time. The ability to detect such regularities is thought to develop early in humans. Infants as young as two months of age have demonstrated the ability to detect simple visual regularities (alternating left and right displays) as demonstrated by anticipatory eye movements [7]. Additionally, infants as young as 2 months of age appear to be sensitive to the transitional probabilities defining pairs of shapes as evidenced by a novelty preference during a habituation paradigm [53]. Similarly, 9-month-old infants successfully discriminate between shape pairs that reliably co-occured together and those that did not, demonstrating their sensitivity to the co-occurrence frequency of visual elements [24].

The ability to detect environmental regularities is a powerful learning mechanism. This ability likely involves two components, temporal co-occurrence and the ability to detect the probabilities that underlie particular groupings. While the studies reviewed above all used timing structures appropriate for infant detection, the individual visual elements were all static and presented centered in the screen. While this is important for baseline understanding of the developing ability, it begs the question whether infants can learn probabilistic event sequences across space as well as time. A series of experiments have demonstrated that temporal order statistics that involve spatial relations likely come online during the first year of life [54]. These studies showed that 11-month-old infants were able to use location statistics whereas younger infants did not. The youngest infants tested did not demonstrate sensitivity to locations; however, they were able to use probabilistic sequences of both color and shape. Furthermore, 8-month-old infants showed a sensitivity to location statistics only when color and shape both provided redundant cues. These results suggest that while young infants may be sensitive to temporal associations, it appears likely that spatial, or location based associations, are later developing.

The natural world is considerably more cluttered and inherently more variable than most laboratory experiments could possibly allow. When variability and statistical noise are considered, unsurprisingly, 8-month-old infants show more errors (fixating the incorrect window) when the transitional probability of visual events decreases (1.0, 0.75, and 0.50 events; [102]). Furthermore, the authors also reported that in the noise condition infants spent significantly longer looking to the correct window across all three probabilities. However, with regard to anticipatory fixations, infants only showed above chance performance with anticipatory fixations in the high and deterministic probability events. Post hoc comparisons revealed that infants showed even more anticipatory fixations on highly probable events compared to deterministic events.

Taken together these results suggest that variability and noise might modulate infant's attentional allocation, as well as indicate a potential preference for more variable relations. The authors suggested one potential reason for this interesting finding is that the high probability events may have been more interesting to infants than the other conditions (which may have been perceived as repetitive or random). Furthermore, this enhanced interest demonstrated during the high probability events may actually facilitate learning as indicated by the increase in anticipatory fixations.

Recent work has shown that infants as young as 6 months can learn contextual information within a laboratory session by detecting the regularities in visual pairs and then use that knowledge to guide their attention in a visual search task of static images [101]. These results confirm what prior research has suggested, the use of top-down factors to guide attentional selection begins to develop around 6 months. However, thus far these results have been limited to static displays which do not fully replicate the amount of visual information available at any one time in the real-world, nor do they contain attention grabbing temporal characteristics inherent in dynamic displays. An important next step is to determine whether infant viewing patterns show the same influence of top-down factors, such as knowledge, when free viewing dynamic stimuli.

1.1.3 Comparing bottom-up vs. top-down influences

Models of gaze allocation within complex scenes have used a bottom-up approach with mixed success [96]. On it's own, bottom-up attentional selection explains very little about free-viewing behavior of dynamic scenes [89], and there is mixed evidence as to the unique explanatory power of bottom-up models in free-viewing behavior of static scenes [47, 41]. Further analysis of the contributions of low level features to fixation selection have demonstrated that the amount of fixation behavior accounted for by physical salience is rather low [95]. Moreover, this modest influence decreases even further when the observer is given a task such as visual search [44]. A comparison of computational models representing human search behavior found that a purely top-down model was the most representative of visual search performance when compared to any model that comprised bottom-up features [112].

Comparisons of bottom-up (image salience), scene independent spatial biases (SISB; center bias), and top-down (cognitive guidance theories) models of human gaze behavior have demonstrated that the fit of each model is influenced by the viewing task [41]. The authors concluded that tasks such as memorization or aesthetic judgment produce a center bias that is best predicted with a purely SISB model and inclusion of low-level features explained less unique variation in scene fixation density compared to the SISB model. Importantly, the same pattern of results were found when only the earliest fixations were considered which suggests that even early fixations are unlikely to be driven by low-level or bottom-up features other than scene-independent spatial biases. Furthermore, the authors found that the top-down model explained the most unique variance across all three tasks (aesthetic judgment, memorization, and visual search) above and beyond SISB and full image saliency models.

The findings outlined above have compared the influence of bottom-up and top-down attentional control in screen-based studies. How well these findings extend to real-world studies remains largely unknown. One real-world search study investigated the influence of bottom-up and top-down cues on search performance [25]. Experiment one of this study manipulated the physical salience of the target and found that the manipulation did not influence the overall amount of time to find the target (a measure of search efficiency). However, increasing the physical salience did decrease the overall amount of time the participant's took to fixate the target once it was within their visual field. Study two manipulated top-down aspects of the search (knowledge related to the target) which resulted in significantly more efficient search (participants found the target faster and from farther away). These findings suggest that top-down factors (knowledge) influence visual exploration in the real-world to a greater degree than bottom-up factors (physical salience) alone.

Comparing bottom-up and top-down influences across development

Bottom-up factors such as physical salience appear to have less influence over infant visual attention in the second half of the first year of life [60, 14]. One study assessed infant attention (4, 6, and 8-month-old infants) to arrays of static images that included everyday objects such as flowers, shoes, sippy cups, teddy bears and human faces [60]. Each stimulus array was assessed for the most salient objects or regions of the scene using three separate computational models. Faces were included in these arrays as a proxy of top-down attention given that faces contain social information. Four-month-olds looked first and longest to the physically salient item in the six item array while 6 and 8-month-old infants looked first and longest to the faces. The authors concluded that between 4 and 8 months physical salience has a decreasing influence on infant attention.

In studies such as these, faces are often used as a proxy of top-down attentional selection due to the social relevance of human faces. However, images of semantic categories such as faces tend to be physically salient in addition to socially relevant. This makes differentiating between the reasons for infant attention to semantic categories such as faces quite challenging. Yet, assessing the prior knowledge of infants is similarly as challenging. Instead of making assumptions as to the likely knowledge held by infants prior to the experimental session, future studies should consider the well documented ability of infants to rapidly acquire knowledge through mechanisms such as regularity detection during the experimental session itself [101]. In this manner the relative contributions of bottom-up and top-down factors to infant attentional selection will be unambiguous.

1.2 The role of eyes and head in real-world visual attention

Much of the visual attention literature to date has used stationary, screen-based tasks, to track characteristics of the eye movements made when given specific task instructions. However, the findings from screen-based tasks may not generalize to the real-world as, at least historically speaking, we do not live our lives in front of screens. In reality our bodies play a much larger role in visual processes as we continuously move about to bring more of the visual world into view [64]. This can be as simple as turning your head to look to who is speaking, or physically moving (walking, crawling, rolling in your desk chair) locations to peer into the next room or look out the window. Engaging in any combination of these behaviors necessitates, at a minimum, the coordination of eyes and head.

1.2.1 Visual exploration in the real-world depends on coordination of eyes and head

The visual-motor system is an active system in which the eyes *actively* seek out information and the motor system proceeds in a manner that is consistent with current goals. To support this active system our eyes can change direction upwards of three times a second [66]. Given the nested nature of the integral components (eyes nested in the head situated atop the body) there are innumerable ways in which the components of the oculomotor system can operate. How exactly the components of our visual and motor systems coordinate appears to be, at least to some degree, dependent upon the biomechanical constraints of each component in addition to the particulars of the task.

Humans tend to move our bodies in such a way as to minimize the energy being expended [69]. Extreme rotations of the eyes, head, or body, as well as rotations of larger body parts, are more costly in terms of energy expenditure [108, 64]. It stands to reason that the movements that we make to support vision are going to depend on the size of the rotation required in addition to any biomechanical constraints [64]. That is to say that if a head movement will suffice, we are more likely to turn our heads to bring the necessary area into view rather than turn our entire trunk or body. However, this explanation only covers eye or body movements within a certain range as extreme rotations are uncomfortable and we operate under constraints other than these biomechanical ones. For example, while the eyes can physically rotate approximately 50-55^{o} horizontally, head movements tend to accompany eye movements larger than approximately 20^{o} [64]. Similarly, the neck can rotate approximately 90^{o} horizontally, but extreme head rotations are uncomfortable for prolonged periods of time (and so likely necessitate rotations of the trunk or body for extended viewing). In situations in which head movements are necessary for task completion, participants may rely more on their working memory to reduce costly head movements [80]. During a block copying task participants were more likely to move their head when engaging in fine motor activities such as picking up or putting down the block pieces. Importantly, the authors note that the participants were less likely to make head movements for referencing the model and instead were likely spending more time committing the model to memory to reduce the need for unnecessary head movements.

Studies of gaze in natural contexts have shown that gaze tends to be more widely spread horizontally compared to vertically [26, 92, 98]. In studies of walking, participants spread their gaze more to the left and right (14^{o}) when compared to up and down (7^{o}) [92]. Furthermore, the contribution of eye movements to the overall spread of gaze in this study was only 4-5^{o}, presumably leaving the rest up to the head or body. This is especially important given the capacity of the eyes to make much larger rotations (approximately 50-55^{o}). A second study found similar findings in that the horizontal rotation of the head of their participants was between 37-46% of their total gaze shift [98]. Taken together it is reasonable to conclude that while the eyes may have the capacity to rotate up to 55^{o} horizontally we tend to recruit head movements for much smaller rotations.

It is clear from both screen-based and real-world studies that an individual's task influences their visual exploratory behavior. Furthermore, we know that we use eye and head movements differently to fulfill our visual exploration needs. However, how our eye and head movements are adapted to meet the demands of top-down factors such as task remains unknown. Chapter 4 of this dissertation addresses this question by comparing the coordination of eye and head across two locomotor tasks, walking and walking while searching.

1.2.2 Task affects visual exploration

Visual exploration in everyday activities is task dependent. This means that the eye, head, and body movements used to support visual exploration are made in service of that task. Studies of walking have established that walkers don't often look where they are stepping when on level ground [78, 71]. Indeed, when asked to walk over novel, but uninterrupted terrain, walkers looked to the ground less than 10% of the time [71]. On uneven ground, or when there are potential hazards in the path, visual sampling increased. For example, when walking over terrain with a hole in the path, ground sampling increased to approximately 40%. When walking on terrain that included obstacles, the standard deviation of the amount of time in between looking to the ground was significantly higher than when walking on less difficult terrain [78]. This suggests that there is more variability in the visual exploratory behaviors used when traversing uneven or difficult terrain than there is when the terrain is flat. This increase in variability suggests that there are individual differences in visual exploratory strategies.

Furthermore, when walking over uneven terrain walkers tend to look a few steps ahead [71]. Even newly walking infants will travel over obstacles without actively looking at the obstacle [30]. Similarly, when crossing a narrow bridge infants primarily looked toward their goal and only fixated the bridge briefly before crossing [58]. One adult study found that participants looked approximately 2 steps, or 0.8-1 s, ahead on uneven terrain [79]. Looking ahead serves an important role in information gathering. These results suggest that sufficient visual information had been previously gathered to safely traverse the occupied ground. Looking ahead to gather immediately relevant information is not only efficient but in this case also strategic and important for safety.

Gaze behavior during well known tasks such as hand washing [80], tea making [67], or block copying [81] suggest that people look at objects that are relevant for the task at hand. The participant's familiarity with the task appears to influence temporal elements of visual exploration such as when they look to these objects as well. In an "over learned" task such as hand washing, participants demonstrated look-ahead fixations to objects that they would soon use such as the soap or towel dispensers [80]. These look ahead fixations are possible because participants are well aware of the sequential steps necessary to complete this daily activity. Because of this prior knowledge, they were able to gather the visual information necessary to efficiently finish the task by preparing for the next steps, which often required a change in motor activity such as a reach to the soap dispenser before dispensing soap. Importantly, these look ahead fixations did not replace the guiding eye movements made approximately 500 ms before the aforementioned reaches.

In the studies outlined above, the participant's knowledge of the task was not assessed prior to the experimental session. While it is highly likely that their prior knowledge influenced their gaze behaviors, these studies were not set up to directly assess the contribution of top-down factors such as prior knowledge. To address this gap in our understanding, the second study of this dissertation found in chapter 3 directly assessed the role of top-down factors such as environmental regularities (and the ability to detect them) on real-world search efficiency.

Chapter 2

Rapidly Acquired Knowledge Modulates Infant Attention During Free-Viewing of Dynamic Scenes

2.1 Abstract

Attentional selection is thought to be influenced by two categories of factors: *bottom-up factors* that are exogenous to the individual (e.g., physical salience) and *top-down factors* which are endogenous (e.g., knowledge or current goals). While it is understood that the influence of these factors likely changes considerably during infancy [14, 33, 60] and early childhood [1, 43], it is still unknown exactly how adult-like visual exploration develops during the first several years of life. To date, the majority of existing assessments of infant attention to top-down factors has used attention to human faces as a proxy for top-down attention. However, faces are often physically salient as well as socially relevant and so disentangling whether attention to faces reflects endogenously or exogenously guided attention is often impossible. The present study considers the development of the influence of top-down factors on infant free-viewing of dynamic scenes. To manipulate infant attentional selection we used a gaze-contingent paradigm to equip 8, 12, and 18-month-old infants with the ability to control the presentation of stimuli. In this manner infants were given the opportunity to learn the association between their attention to the gaze-contingent object and the subsequent stimuli presentation. To assess the influence of this rapidly acquired knowledge on infant attention, infants were next presented with 30 s Baby Einstein video clips which included the digitally inserted gaze-contingent and non-contingent objects. All infants, regardless of age, spent more time attending to the gaze-contingent object more during free-viewing than the non-contingent object. However, no significant age-related differences, or interactions, were found. The successful modulation of infant attention during free-viewing confirms the influence of top-down factors on infants as young as 8-months

during free-viewing of dynamic scenes. The current sample remains under powered to assess the influence of age on infant attention; however, future directions include assessing individual differences both within and between age groups.

2.2 Introduction

The natural world includes such a wealth of information (visual, auditory, tactile, etc.) to learn from that parsing which information is most relevant at any given moment can be challenging. Attentional selection, or the process of determining where to attend, is an integral component of visual attention. Through this process we are able to rapidly sift through competing sources of visual information and attend to what is most pressing or most informative [8]. Two categories of factors are thought to influence attentional selection: *bottom-up factors* that are exogenous to the individual (e.g., luminence, contrast, flicker, etc.), and *top-down factors* which are endogenous (e.g., knowledge or current goals). Prior research indicates that attentional selection is a process that changes considerably during infancy [14, 33, 60, 4]. The viewing behaviors of young infants are heavily influenced by properties of external stimuli with regard to where they look, when they look, and for how long [13]. However, there is conflicting evidence regarding whether attention to top-down factors increases during infancy, as some prior work has reported this change beginning between 4 and 8-months [60] while others have not [3, 4, 31, 50]

One potential reason for these conflicting findings is the continued use of attention to faces as a metric of top-down attention. Human infants show very early visual preferences for areas of high contrast, certain colors, and clusters of objects that resemble the configuration of eyes and mouth (two on top, one on the bottom; [74]). Elements such as these are not only bottom-up factors, but are also prevalent in human faces (e.g., the iris and the sclera next to one another often creates an area of high contrast). This early preference for faces and face-like stimuli appears to change within the first few months as preferences emerge for mother over stranger [77, 16], female over male [84] and own race versus other race faces (for a review see [86]). While faces are undoubtedly meaningful and convey social information, they are often physically salient as well [44, 107] and so attention to faces cannot be a pure metric of top-down attention. In the present study we aim to directly assess the influence of top-down knowledge on infant free-viewing of dynamic stimuli through use of a gaze-contingent eye tracking paradigm. In this paradigm each infant will control the presentation of their own stimuli, and in this manner are afforded the opportunity to implicitly learn the association between their own gaze and the stimuli presented to them.

2.3 Attention to static faces during infancy

Previous work has shown that top-down factors appear to influence attentional selection during infancy [60]. A study of 4, 6, and 8-month old infants showed that 6 and 8-month-old infants looked first and longest to faces embedded within static arrays of object photos, while 4-month-olds looked most to the physically salient item. These results suggests that between 4 and 8-months a transition may occur between the influence of top-down and bottom-up factors, such that the influence of bottom-up factors decreases while the influence of top-down factors increases. This result extended previous work with children and adults which found that younger children (2-6 years) are more highly influenced by saliency throughout scene viewing, but for older children (8+) and adults saliency was only found to influence their gaze during initial scene inspection [43]. Similarly, work with older children (7-9 years) and adults found that the children used more of a bottom-up fixation strategy than adults which led the authors to conclude that with age top-down factors are more heavily prioritized when viewing natural scenes [1].

While some studies have found evidence for a decrease in the influence of bottom-up features within the first year, others have found that bottom-up features become more influential on infant attentional selection over time [4, 31, 52]. Amso and colleagues (2014) [4]) found that children and adults differentially attended to physically salient faces over non-salient faces, while infants did not. The authors concluded that visually salient faces attract bottom-up orienting more than non-salient faces only after infancy, suggesting an age-related change in attention to bottom-up factors. Importantly, this work manipulated the saliency of the faces which allowed for a direct comparison between attention to highly salient faces versus attention to faces that were *less* physically salient. Similarly, Kelly and colleagues (2015) [52] found that 3-month-old infants consistently detected faces in static images of naturalistic scenes, although salient faces were more likely to be detected (fixated) than less salient faces and there were age related improvements in salient face “detection”.

These studies further our prior work in which we found that with age, all observers (infants and adults) spent more time fixating on the human actors face when freeviewing Sesame Street video clips [31]. Furthermore, faces in these videos were more salient than other areas of the video that the participants fixated (regardless of age) *and* face looking was significantly correlated with intersubject correlations between adult and infant eye movements (a measure of adult-like viewing). There appears to be little to no cohesion in the evidence presented in the developmental literature with regard to age related changes in the influence of bottom-up and top-down factors on infant attentional selection. However, one consistent element from each of these studies is their use of infant attention to faces. Given past evidence on the physical salience of faces henderson2007e, WassSmith2015, infant attention to faces cannot be considered a pure metric of top-down attention.

Evidence from the adult literature demonstrates that task relevant goals and instructions are highly influential in guiding attention on the screen [89] and in the real-world [67, 81, 80]. In addition to task instructions, increasing evidence suggests that the content of the scene is highly influential in determining how both infants and adults allocate their attention [50]. However, these top-down factors remain nearly impossible to assess with infants as we can neither ask infants what they know nor assign them a task. To better understand the influence of top-down and bottom-up factors on attentional selection, we must be able to accurately manipulate rather than assume infant's prior knowledge.

2.4 Regularity detection as a means of rapidly acquiring new information

While there are many mechanisms thought to underlie learning, the ability to detect underlying regularities in environmental input appears to be present from birth [6]. By two months of age infants can learn the pattern associated with alternating visual displays [7] as well as in the transitional probabilities that define pairs of shapes [53]. Young infants are not only able to detect the probabilities that underlie the co-occurrence of visual stimuli that are paired together on the same display during an experimental trial (i.e., two shapes repeatedly paired together in a habituation paradigm), but also objects or events that are temporally contiguous and sequential [7]. As such, this mechanism for learning can afford the opportunity for real-time learning to occur within a laboratory session [106, 101].

Infants as young as six months of age have demonstrated the ability to learn the association between their gaze and a change in visual display in a gaze-contingent trial in as few as three trials [106]. The 8-month-old infants in this sample demonstrated their knowledge by triggering the gaze-contingent event significantly more often than the 6-month-olds. Analysis of the average fixation durations to the gaze-contingent target (a red circle) and the sequentially presented engaging image (photos of animals) demonstrated that infants made significantly longer fixations to the animal images even though they were only presented for 1.5 s at a time. The authors concluded that the fixation patterns suggest that the infants did not look to the gaze-contingent target simply due to the physical salience of the target, but rather due to their implicitly gained knowledge of the relationship between their gaze and the resulting image.

Recent work has shown that infants as young as 6 months can not only learn contextual information within a laboratory session by detecting the regularities in co-occurring pairs of shapes [101], but also use this knowledge to guide their attentional selection in a visual search task of static images. These results confirm what prior research has suggested, top-down factors begin to influence infant attentional selection around 6 months. However, thus far these results have been limited to static displays which do not fully replicate the amount of visual information available at any one time in the real-world, nor do they contain attention grabbing temporal characteristics inherent in dynamic displays [89, 73]. An important next step is to determine whether the viewing behaviors of infants show the same influence of top-down factors when free viewing dynamic and complex scenes. Given the "attend to me signals" that bottom-up factors tend to exude sawakiluck2010, and the attention-getting quality of temporal characteristics such as motion [73], it is certainly possible that infant attention during free-viewing of dynamic scenes may differ from the attention exhibited while exploring static images.

2.4.1 Current study

In the present study we assessed the influence of rapidly acquired knowledge on the attentional selection of 8, 12, and 18-month-old infants during the free-viewing of dynamic scenes. A gaze-contingent paradigm was used to allow infants to control the visual input presented to them by simply fixating on their predetermined target for 200 ms. Infants were presented with two objects during the initial gaze-contingent trials, one which reacted contingently to the infant's gaze (gaze-contingent object), and one which did not (non-contingent object). Both objects were digitally inserted into 30 s infant-directed "Baby Einstein" video clips that the infants were able to freely watch after the gaze-contingent trials, subsequently referred to as "free-viewing trials". Greater attention to the gaze-contingent object (over the non-contingent object) during free-viewing would indicate attention that is driven by the knowledge gained during the experimental session.

We hypothesize that infants (regardless of age) will show greater attention to the gaze-contingent object during free-viewing. Prior research has found that infants as young as 6 months are able to learn the association from gaze-contingent paradigms in as few as 3 trials [106]. In this manner we can better understand the developmental trajectory of endogenously guided attention in 8, 12, and 18-month-old infants. Additionally, previous research with static images suggests that the influence of top-down features on attentional selection increases with age [60]. We expect these findings to generalize to the current study and our use of dynamic stimuli. As such, we predict age-related increases in attention to the gaze-contingent object during free-viewing of these dynamic stimuli. Such that the oldest infants (18 months) will attend most to the previously gaze-contingent object, followed by the 12-month-olds, and lastly the 8-month-old infants.

2.5 Methods

2.5.1 Participants

The final sample consists of 40 infants (22 female), 14 8-month-olds, 16 12-month-olds and 10 18-month-old infants. The sample was recruited from a shared database of families who had previously indicated their interest in developmental research from the Riverside area. The average age of each group is as follows, 8-month-olds (M = 8.28 months, *SD* = 18.40 days), 12-month-olds (M = 12.6 months, *SD* = 25 days), and 12-month-olds (M = 18.3 months, *SD* = 16.66 days).

Three additional infants were tested but their data were excluded completely due to issues affecting the entire experimental session: one infants failed to complete the gaze-contingent trials due to fussiness/inattention. An additional two participants were excluded for poor calibration (average error $> 1.5°$).

The caregiver of each infant gave informed consent for their child to participate at the beginning of each experimental session. Families were reimbursed 10 USD for their travel and were given a small gift to take home. Each caregiver indicated that they were unaware of any immediate family history of color-blindness and that (to the best of their knowledge) their child had never seen any of the Baby Einstein videos used as stimuli in the present experiment. Caregivers indicated their infant's ethnic and racial identity in a demographic questionnaire and the breakdown of the sample is as follows: 14 caregivers indicated that their infant is of Hispanic or Latino background, 6 chose not to answer, and 20 indicated that their infant is not of Hispanic or Latino background. Furthermore, 1 infant was identified as of American Indian or Alaskan Native heritage, 1 was identified as Asian, 2 were identified as African American, 5 were identified as belonging to multiple ethnic backgrounds, 18 were identified as White, and 13 chose not to answer.

2.5.2 Eye Tracking Apparatus

An Eyelink 1000 Plus remote eye tracker (SR research Ltd.) was affixed to a 43.2 cm (diagonal) wide-screen monitor on an adjustable arm. The free-viewing videos were each presented at 30 Hz, subtending a visual angle of $31^{\circ} \times 19^{\circ}$. Right eye movements were recorded with a temporal resolution of 500 Hz.

2.5.3 Stimuli

Infants were first presented with 8 gaze-contingent trials. These trials were composed of two stationary cartoon objects, a blue flower (visual angle approximately $2^{\circ} \times 2^{\circ}$) and a yellow sun (visual angle approximately $2.5^{\circ} \times 2.5^{\circ}$) on a black background. The overall image subtended a visual angle of approximately $25^\circ \times 14^\circ$. The location of the two objects were presented quasi-randomly, such that no object was ever presented in the center of the screen and the two objects were never present in the same section of the screen (with each section an equal 1/9 of the screen). The order of each trial was presented randomly for each participant and lasted until either the infant made a 200 ms fixation to their pre-determined gaze-contingent target object or 60 s had elapsed with no such fixation. Once the infant fixated the gaze-contingent object, a 5 s video clip appeared on the screen in the previous location of the gaze-contingent object. This 5 s video clip remained the same across all 8 gaze-contingent trials and included bright balls bouncing across the screen on a background that changed color. This lively, animated video clip was paired with a 5 s audio clip of an infant laughing.

Infants were next presented with 8, 30 s free-viewing trials. Each video clip was taken from "Baby Einstein" and was selected so that the entire 30 s video clip did not contain any cuts or scene changes and all contained animated backgrounds and moving puppets. Furthermore, each video was partitioned into its composite frames and still images of the two objects from the gaze-contingent trials were digitally inserted into the scene following the randomization criteria detailed above. Additionally, the two static objects were inserted into the scene in such a way that they did not overlap with any of the puppets who frequently moved around the scene. The content of each video varied such that each video had a differing number of puppets (at least one and as many as three puppets were included in each video clip), the amount each puppet moved varied across each video clip as well (as the actions they performed, and the content of the background varied video to video as well).

The original audio content of each video clip was removed and each video was paired with children's instrumental music.

2.5.4 Procedure

Each infant sat in a highchair secured with a 5-point harness while their parent sat in a chair behind them. Parents were asked to close their eyes and refrain from pointing or speaking for the duration of the experiment. A curtain was used to separate the participant and caregiver from the experimenter in order to reduce any potential distraction. Before the infant was seated a target sticker was placed on their forehead to expedite the detection of their eyes by the eye tracker.

At the beginning of the study, the experimenter adjusted the position of the monitor such that the eye tracking software determined that the participant's eyes were approximately 60 cm from the eye tracker. A 5 point calibration sequence was used for each participant, such that the participants were cued to follow a looming object that randomly appeared in each corner of the screen as well as the center. Immediately following the calibration procedure came a similar 5-point validation check. From this validation data the average error in degrees of visual angle was calculated for each participant. The average error for each age group is as follows, $M = 0.68°$ for the 8-month olds, $M = 0.80°$ for the 12-month olds, and $M = 0.64°$ for the 18-month olds.

After calibration and validation, participants were first shown the 8 gaze-contingent trials in randomized order followed by the 8 free-viewing videos. Each trial was preceded by a gaze-contingent attention-getting object to re-orient the infant's attention to the screen.

2.5.5 Data Processing and Measures

Experiment Builder software (SR Research Ltd) was used to create Areas of Interest (AOIs) for each object to calculate the number of fixations made within each AOI as well as the total dwell time (the total amount of time spent within the boundary of each AOI) for each trial. From these values we were able to calculate the aggregate fixation durations and number of looks made to each AOI as well as which object was looked to first (first looks) and the latency to look to each target. These four dependent variables were calculated as an average for the 8 free-viewing trials.

2.6 Results

Our first hypothesis stated that infants would attend more to the gaze-contingent object compared with the non-contingent object during free-viewing. Additionally, we also hypothesized that older infants were likely to show greater modulation of their attention compared to younger infants. We expect this to be similarly demonstrated as previously outlined.

Infants distributed their attention to the gaze-contingent and non-contingent objects differently during free-viewing. We compared the average fixation duration, the average number of fixations, the total amount of trials each infant looked first to the each object (first looks) and the average latency to look to each object. The following results describe two-way repeated measures analyses of variance (ANOVA) calculated using *R*, Version 2022.02.0+443 [85] to assess the effect of contingency (attention to the gaze-contingent

Table 2.1: Summary of Two-way Repeated Measures ANOVA results demonstrating the effects of target contingency on each variable of interest.

	Contingency			**Age**		
	df	*F*	*p*	*df*	*F*	*p*
Fixation Duration	1	5.97	.02	2	0.12	.88
Fixation Count	1	3.76	.06	2	0.91	.41
Latency	1	0.00	.95	2	0.88	.42
First Looks	1	9.70	.002	2	0.02	.79

* $p < .05$

vs. the non-contingent objects) and age (8, 12, and 18 months) on each of the respective dependent variables.

In support of the first hypothesis, a significant main effect of contingency was found on the average fixation duration, such that infants of all ages spent significantly longer fixating the gaze-contingent object (M = 924 ms, SD = 750) compared to the non-contingent object (M = 583 ms, SD = 474), $F(1,74) = 5.97$, $p = .017$, $\eta_p^2 = 0.07$ (Figure 2.1). One potential outlier was identified as larger than three standard deviations from the mean. We repeated this analysis after replacing the outlier with the mean and found the effect remained significant $F(1, 74) = 5.44$, $p = .022$, $\eta_p^2 = 0.07$.

Additionally, a marginal main effect of contingency was found for the number of fixations to the gaze-contingent and non-contingent objects, such that infants from all age groups made more fixations to the gaze-contingent object ($M = 2.67$, $SD = 1.69$) compared to the non-contingent object ($M = 2.00$, $SD = 1.38$), $F(1,74) = 3.76$, $p = .06$, $\eta_p^2 = 0.05$ (Figure 2.2). One potential outlier was identified and replaced using the same process as outlined above, which resulted in a statistically non-significant difference between groups, $F(1, 74) = 2.72$, $p = .10$, $\eta_p^2 = 0.035$.

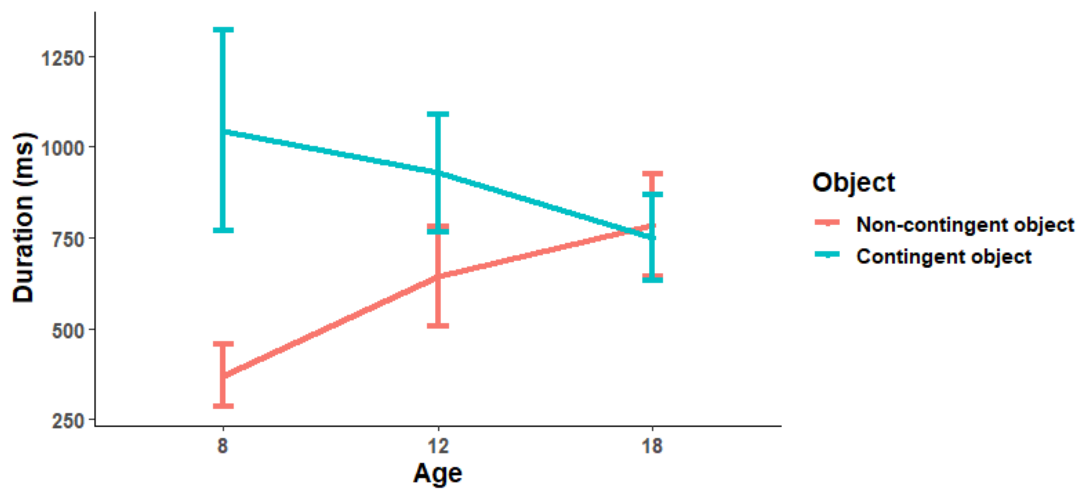

Figure 2.1: Average fixation duration to each target by age group.

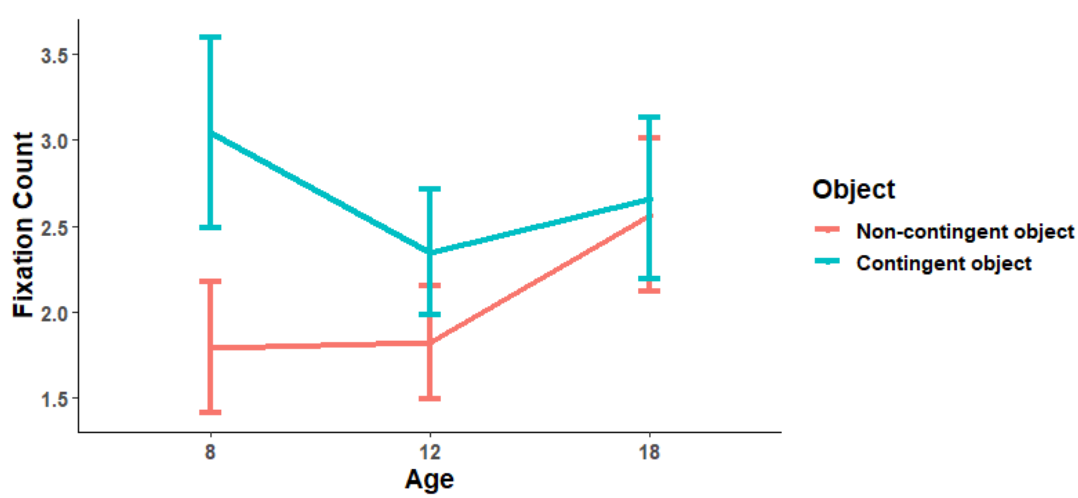

Figure 2.2: Average number of fixations to each target by age group.

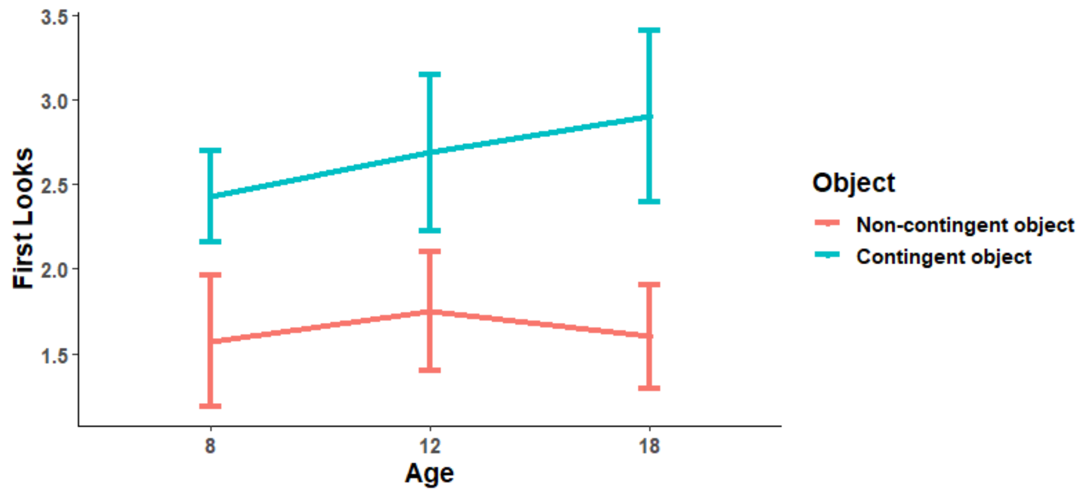

Figure 2.3: Number of first looks to each object by age group.

We next assessed the effects of contingency on which of the two previously seen objects the infant looked to first during the eight free-viewing trials. A significant main effect of contingency was found for the total number of first looks to each object, such that infants from all age groups made significantly more first looks to the gaze-contingent object (M = 2.65, SD = 1.51) compared to the non-contingent object (M = 1.65, SD = 1.29), $F(1, 74) = 9.70$, $p = .002$, $\eta_p^2 = 0.12$ (Figure 2.3). However, no significant main effect of contingency was found for the latency to look to either object, $F(1, 74) = 0.004$, $p = .95$, $\eta_p^2 < 0.001$ (Figure 1.4).

No significant effect of age was found for average dwell time, $F(2, 74) = 1.53$, $p = .22$, the total number of fixations made to each object, $F(2, 74) = 0.91$, $p = .41$, first looks to either object, $F(2, 74) = 0.24$, $p = .79$, or latency to look to either object, $F(2, 74) = 0.88$, $p = .42$. Contrary to our second hypothesis, we found no age by contingency interactions

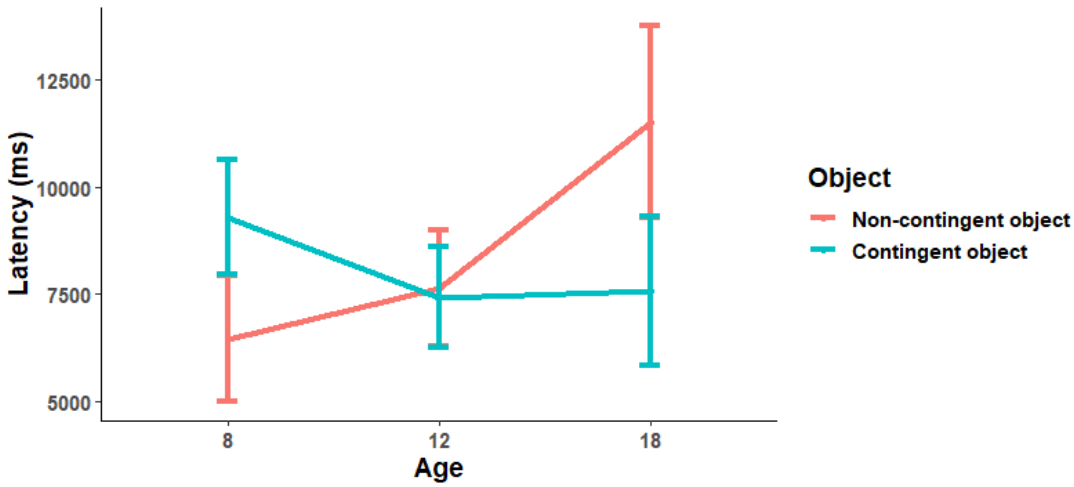

Figure 2.4: Latency to look to each object by age group.

for any of the four variables of interest. However, the age by contingency interaction for fixation duration ($p = .16$) and latency ($p = .12$) both trend towards significance.

2.7 Discussion

The results of this experiment demonstrate evidence of attentional selection driven by top-down processes during free-viewing of dynamic stimuli in 8, 12, and 18-month-old infants. The results described above largely support our first hypothesis in that, for multiple measures of attention (fixation duration, first looks, and the number of fixations), we found evidence of infants differentially attending to the gaze-contingent and non-contingent objects during free-viewing. Importantly, infant attention was significantly more often directed towards the gaze-contingent object over the non-contingent object, providing evidence of successful attention modulation due to knowledge rapidly gained during the experimental session.

While we failed to find any significant age by contingency interactions for any of our variables of interest, we did find two interactions that appear to trend towards significance. Age and contingency appear to interact for both average fixation duration (p = .16) and latency (p = .12) to look to each object. It is possible that we are under powered to detect such an effect. However, in the current sample we failed to find support for our second hypothesis, which stated that there would be age-related increases in attention to the gaze-contingent object.

The present results confirm that infants as young as 8-months can both rapidly gain knowledge during the experimental session, as prior research has demonstrated, [101, 106] and use that knowledge to guide their attention during later free-viewing of dynamic stimuli. These results build off the findings of [101] who used regularities in target presentation to achieve rapid knowledge acquisition to assess infant attentional selection during a static visual search paradigm. These findings not only confirm that infants as young as 8-months are able to use top-down information to guide their attention, but they are able to do so while free-viewing dynamic infant directed media. Prior research has demonstrated that "tot tv" or infant directed media is designed in such a manner to capitalize on infant attention through use of a variety of low level, or bottom-up features including flicker and feature congestion [107]. Thus, these findings not only demonstrate the ability of young infants to demonstrate top-down attention, but to do so in the presence of highly salient content that was designed to capitalize on their attention.

Furthermore, these findings extend the findings of previous work which used infant attention to faces as a metric of top-down attention [60, 33]. Previous findings have demonstrated that between 4 and 8-months the factors that influence infant attention begin to change, such that 6 and 8-month-old infants look first and longest to faces in static arrays, but 4-month-old's would direct their attention more often to the most physically salient item in the array [60]. Thus, in both static and dynamic displays infants as young as 8-months appear able to direct their attention in a top-down manner even when the competing stimuli are designed with infants in mind. Future work should include a wider age range of infants to better assess when this shift may occur in which infants begin to use top-down information to direct their attention.

While these findings do provide evidence of infant attention that is driven in a top-down manner, we are not yet able to discern how changes in scene content influence infant attention. While we found that infants did employ top-down attention during these 30 s video clips, on average infants spent less than 1 s (total) looking to the gaze-contingent object. How these infants deployed their attention across the duration of the video remains an open question. While the gaze-contingent manipulation allowed for a measure of certainty that attention was being deployed in a top-down manner, it was not the only top-down content present in each video clip. These videos clips were chosen because they each contained freely moving puppets with the characteristics of human faces (i.e., top-heavy orientation of eyes and mouth). However, the puppets in these video clips were also the primary source of movement present in each scene and so attention to these puppets cannot reflect only top-down attention. Future work will use windowed analyses to assess infant attention across the duration of each video clip. In this manner we will be able to assess how the content of each scene influences infant attention. Prior research from our lab has used attentional synchrony between adult and infant observers during free-viewing to assess the degree to which infants display adult-like attention during free-viewing [50].

While the current sample size leaves this study under powered to detect the age-related effects that we had originally hypothesized, data collection for this project remains ongoing. Furthermore, this study is an initial foray into this data and planned future directions are to assess the potential of individual differences in infant attention during free-viewing. Given the large degree of variability present in the current sample, we believe that there are likely undetected individual differences present which the sample size is currently under powered to detect. Future directions should assess differences both between and within age groups.

Chapter 3

The Effects of Prior Knowledge and Implicit Learning on Real-World Search Efficiency

3.1 Abstract

Real-world search requires coordinating the *perceptual system* (i.e., eyes nested in the head and head nested in the body [38]) to move about in the environment in pursuit of the search target. Screen-based search limits the perceptual system to looking simply with the eyes and as such cannot fully replicate real-world search. Regardless of the type of search, search efficiency refers to the ability to mitigate unnecessary movements (eye, head, or body) in order to complete the task as quickly as possible. Decades of research have uncovered numerous factors that influence screen-based search efficiency including aspects of the targets themselves (bottom-up factors, e.g., contrast, orientation, etc.) and the prior knowledge of the individual (top-down factors, e.g., semantic knowledge). However, the extent to which these factors influence real-world search efficiency remains largely unknown. In the present study we manipulated two types of top-down factors, environmental regularities and prior knowledge, to assess their effect on real-world search efficiency. 59 participants were randomly assigned to one of three experimental conditions (varied, consistent, or instructions) which differed in the location probability of the targets and the knowledge given to participants before searching. One-way ANOVAs were used to assess between group differences on four measures of search efficiency. These measures were chosen to assess the contribution of the eyes (horizontal spread, or the average standard deviation of horizontal eye position) and body (indirectly through measures related to the search path, e.g., straightness ratio or total path length) to real-world search. While we failed to detect between-group differences with the present analyses, we discuss the possible reasons for the present findings and planned future work related to both between and within group differences.

3.2 Introduction

Real-world and screen-based search differ in meaningful ways. Screen-based search often requires the participant to hold their heads and bodies still to mitigate eye data loss while real-world search often necessitates coordinating the entire *perceptual system* (i.e., eyes nested in the head and head nested in the body [38, 37]) in order to bring more of the visual world into view [65]. Search efficiency refers to how an individual coordinates eye, head, and body movements in the environment in pursuit of their task (e.g., mitigating unnecessary movements). In both screen-based and real-world search, a common metric of search efficiency is the overall duration of the search task; however, other metrics include the physical distance traveled (or scan paths for screen-based search) or the straightness ratio of the search path (real-world only). Decades of research have found numerous factors that influence search efficiency such as the physical salience of the target (e.g., pop-out search [25]), prior knowledge [68] or the ability to detect regularities present in the environment (e.g., regularity detection [111]). However, much of the visual search literature to date has been limited to screen-based paradigms. Thus, the extent to which the factors that influence search efficiency on screens generalizes to visual search in the real world (e.g., search that requires more than eye movements) remains largely unknown. In the present study we manipulated the probabilities associated with target location as well as the instructions given to participants to assess the effect of regularity detection and prior knowledge on real-world search efficiency.

3.3 Search efficiency in screen-based tasks

Studies of screen-based search have found that the eye movements made during search are guided by both bottom-up (exogenous) and top-down (endogenous) factors. Attention driven by bottom-up factors such as the luminence, color, or contrast within an object (sometimes referred to as the physical salience of an object) is exogenous, that is to say that it is driven by factors external to the individual. For example, when engaged in a form of bottom-up search known as "pop-out" search (search for items that are perceptually very distinct from the distractors) target items are often successfully "found" or fixated within milliseconds by both infants and adults [2]. Conversely, top-down attention involves using what you know to inform visual guidance for what you are doing. This knowledge comes in many forms, including contextual or semantic knowledge [68, 15] and previous experience with the demands of a particular task [68, 80]. While bottom-up factors such as the physical salience of the target have been found to influence aspects of screen-based visual search, this appears to be especially true with heavily controlled, simplistic stimuli (e.g., lines oriented in different directions [19]. Search for more complex stimuli, especially those given context (e.g., situated within a visual scene not an unchanging black or grey background), appears to be influenced by top-down factors such as semantic knowledge [68] and prior experience [11].

One potentially powerful top-down mechanisms for guiding visual attention during search is that of detecting regularities in environmental statistics. Regularity detection involves using perceptual information to detect regularities present in the physical environment (e.g., the location of targets). Studies of screen-based search have found that not only can participants use environmental statistics such as target location to inform where they look, but that this resulting spatial bias may be slow to extinguish (e.g., last for over a week) [48].

Repeatedly searching for the same target in the same physical configuration (repeated search trials) leads to faster, more efficient search (known as the contextual cueing effect). Recent work has shown that detecting regularities in environmental statistics acts as a moderator for contextual cueing [113]. The authors reported that while the probability of repeated search trials remained high (80% repeated trials) participants demonstrated contextual cueing significantly faster (compared to baseline values); however, when the probability of repeated trials was low (20%), the participants' failed to demonstrate contextual cueing. Additionally, the authors also reported that when the repeated and nonrepeated search trials were presented in separate streaks (e.g., first a streak of repeated trials followed by a streak of nonrepeated trials) this too expedited the development of contextual cueing. These results demonstrate evidence that adults are able to detect environmental regularities when the probability of repeated search was 80% of all search trials. Additionally, these results demonstrate that the temporal contiguity of those repeated search trials enhances the contextual cueing effect, leading to faster, more efficient search. However, whether these results would generalize to search that requires moving more than just the eyes (e.g., real-world search) remains unknown.

3.4 Real-world search

Beyond screens, the real, three-dimensional world is complex to explore, because walls, furniture, and buildings obstruct our vision, necessitating head and/or body movements to gather visual information from different areas of the environment [64, 65]. Visual exploration is an embodied process, one that does not solely rely on a single sensory organ such as the eyes. Rather, real-world visual exploration uses a perceptual system that includes the eyes which are nested in the head, which is nested in the body [38, 37]. However, visual exploration relies on a continuous perceptual-action loop [38, 32] in which perceptual information is used to guide bodily movements which also alters what is readily available in the field of view (FOV). This perceptual-action loop is not possible during screen-based search as bodily movements are actively discouraged so as to not disrupt eye tracking. Thus, findings from screen-based studies of search may not entirely generalize to real-world search.

Given the nested nature (eyes in head, head in body, body on ground) of the components of this visual exploratory system, there are innumerable ways in which the components of the system could possibly coordinate. However, there are clear biomechanical (e.g., how far the eyes or body can comfortably rotate), physical (e.g., size of the body), environmental (e.g., physical obstacles), and practical (e.g., social expectations) constraints that limit the possible number of combinations. For example, humans are able to engage in *covert* attention in which attention is deployed to a particular spatial region without needing to move the eyes to look at the area in question [82]. However, there is a clear trade-off for this strategy in that any visual information gathered will be constrained by considerably lower resolution because covert attention relies so heavily upon peripheral vision. Indeed, one known way to assess whether or not covert attention is deployed during screen-based tasks is to first assess that no eye or head movements were used *and* then to look for significant changes to measures of search efficiency such as reaction time (search duration). Decades of research have demonstrated that typical screen-based feature or conjunction search (search that is defined by a single feature or a combination of features– e.g., the white circle) are defined by particular patterns in set size and reaction time. That is to say that search reaction times appear to increase as a function of the number of distractors, unless initial attentional orienting was guided by preattentive features (e.g., pop-out search [109, 110]. Whether or not covert attention would influence search efficiency in a similar manner in real-world search remains unknown.

Real-world studies of visual attention have demonstrated that adults can detect regularities present in their environment which in turn can lead to an attentional bias to spatial locations likely to hold their target [111]. One study compared even earlier occurring measures of gross visual attention, the direction of the initial head turn, to overall search reaction times (RT) (the latency to find the target) and found that participant's initial head turns were overwhelmingly biased in the direction of the "rich" quadrant (contained the target in 50% of trials) [111]. This indicates that the first head movement was sensitive to the location probability of the target. Additionally, the authors reported that initial head turns were 10x faster than the search RT *and* were more responsive to changes in environmental statistics (i.e., the location probability of the target). These results extend previous findings from screen-based search [113] and demonstrate the effectiveness of domain-general learning mechanisms in early attentional orienting during real-world search. However, the search area used in this study was relatively small and sterile (participants could see it's entirety without moving their bodies and the area was devoid of any other objects). What remains to be seen is whether similar attentional modulation or search efficiency would be apparent in studies of real-world search that better replicate search in most environments– that is search in a cluttered visual environment that necessitates more than head movements.

While these results are among the first of their kind to demonstrate the effectiveness of top-down factors in guiding attention during visual exploration in the real-world, it is still unknown whether these findings would extend to other, more granular, measures of attention such as the eye movements made in service of the task. Measures such as the *spread* of horizontal eye movements indicate the degree to which attention is spread along the horizontal axis [32]. Greater spread could indicate more looking with the eyes (and less with the head or body). Similarly, less spread could indicate an exploratory strategy that uses more head or body movements and less eye movement.

3.4.1 Current study

In the present study we aim to extend the work of [111] by assessing the effect of implicit learning of underlying environmental statistics (i.e., the location probability of the target) on measures of search efficiency (e.g., the spread of horizontal eye movements, the straightness ratio of the search path, the total search path length, and the search RT). These measures were chosen to give insight into the contribution of the eyes (horizontal spread, or the average standard deviation of horizontal eye position) and body (indirectly through measures related to the search path, e.g., straightness ratio or total path length) to real-world search in an environment large enough that eye or head movements alone are insufficient to complete the task. Lastly, we will also consider the effect of regularity detection on the overall search duration.

To manipulate the location probability of the targets, participants were assigned to one of three possible between-subjects conditions: varied, consistent, and instructions. The targets in the varied condition were dispersed with half of the targets (3) affixed to one trees and the other half (3) affixed to benches. Conversely, all targets (6) in both the consistent and instruction conditions were affixed to one surface (trees or benches). However, participants in the instructions condition also received explicit instructions as to which types of surfaces their targets would be found on (e.g., your targets can be found on trees). Participants were instructed to find their six identical targets as quickly as possible without running. In this manner we planned to assess the effect of prior knowledge and implicit learning of key environmental statistics on measures of search efficiency. Given the extreme disparities in knowledge between the instructions and the variable conditions, we hypothesize that there will be significant differences in search efficiency between these two groups. Furthermore, adult participants should be able to detect the regularities in the target locations and as such participants in the consistent condition should demonstrate significantly more efficient search than participants in the variable condition.

3.5 Methods

Participants

The final sample for this study consists of 59 adult participants (18 - 31 years)(M = 20.81, SD = 3.0, 39 female, 20 male) all of whom had normal or corrected to normal vision. One additional participant was run in the study but ultimately excluded due to poor eye tracking (calibration error $> 5°$). Additional participants were run in the study but excluded due to bystander interference ($n = 2$), technical difficulties (SD card failure, $n = 4$), or because the eye tracking camera slipped during the searching task ($n = 7$).

All participants were undergraduate students at the University of California, Riverside who received course credit as compensation for their participation. Participants gave written informed consent at the beginning and were debriefed at the end of the experimental session. Participants described their ethnicity as White ($N = 20$), Asian ($N = 15$), Black ($N = 2$), more than one race ($N = 8$) or chose not to answer, ($N = 14$).

Design

Participants were randomly assigned to one of three experimental conditions: varied, in which targets were evenly distributed across two surface types ($N = 20$) consistent, in which targets were distributed across one surface type ($N = 20$), and the instruction condition in which targets were distributed across one surface type as in the consistent condition, but participants were additionally given instructions as to where to find their targets ($N = 19$). The two possible surface types in this experiments were either concrete benches or trees. In the varied condition half of all targets (3) were affixed to benches while the other half were affixed to trees. In the consistent and instructions conditions all targets (6) were affixed to either trees *or* benches. Furthermore, half of participants in both the consistent and instructions locations found their targets on trees while the other half found their targets on benches.

Equipment and materials

Each participant was equipped with a Polar V800 Multisport GPS watch which they wore on their right wrists, as well as a belt bag, wide brimmed hat (to reduce data loss from glare associated with the sun) and a Positive Science mobile eye tracker. The eye tracker was used to record the eye movements of each participant. The mobile eye tracker includes a infrared camera pointed toward the right eye of the participant (eye camera) to illuminate the pupil and record eye movements. Similarly, the participant's field of view (FOV) was recorded by a camera that is affixed to the frame of the eye tracker over the participant's right eye (field of view camera). The recorded videos (eye and FOV) were each sent to a recording device that was stored in a belt bag worn by each participant over their right shoulder.

Search task setting

The search arena was a courtyard that measured 45 m wide $\times$ 30 m long for a total area of 1350 m^2. Approximately 823 m^2 was garden space inaccessible to pedestrians. The remaining space was comprised of mature trees, seating areas, and wide cement walkways. The search arena remained open to the campus public during data collection so pedestrians were often present while participants completed the search task.

In the search arena, targets and distractors were fixed to trees and cement benches located throughout the courtyard in a pre-specified set of locations. Targets and distractors were 10 cm $\times$ 10 cm orange fabric squares with a 3.8 cm $\times$ 2.5 cm shape (rectangle or diamond) drawn on the front in black ink. Of the targets affixed to trees (6 total), 2 targets each were each secured 0.25 m from the ground, 1 m from the ground, and 1.5 m from the ground. Of the targets affixed to benches, targets were secured to the frame of the cement benches, never the seats or legs (each bench measured approximately 1.5 m long × 0.3 m long).

Procedure

Participants were met and consented in a private laboratory space at UC Riverside. After the consenting process participants were equipped with the mobile eye tracker, wide brimmed hat, and belt bag before leaving the lab and walking outside with the experimenter. Once outside the building, the experimenter then demonstrated the calibration process for the participant. Details of the methods as well as the data collected in the present study were previously published in Franchak and colleagues (2021) [32].

Eye tracking calibration To calibrate the eye tracker participants were asked to stand in a particular location and hold their heads and bodies still. Next the experimenter positioned themselves approximately 3 m from the participant and held a stick with a brightly colored piece of cardboard affixed to the top. Participants were instructed to look at the center of the bright cardboard wherever the experimenter moved it without moving their heads or bodies. The experimenter chose locations at the bottom, middle, and top of the participant's approximate field of view and paused at each location to allow the participant time to fixate on the calibration stick. This process was repeated after the search task concluded.

The calibration process was captured by the FOV camera and was used offline (after the session) to calibrate the eye tracker using Yarbus software (Positive Science LLC), producing horizontal and vertical time series of gaze locations in field of view video (pixel) coordinates. Calibration accuracy was verified using an additional set of 5 target fixations, independent from those used to calibrate the eye tracker. For each validation point, we calculated the difference between the actual target location in the FOV camera and the gaze location in degrees– calibration error. The calibration error averaged $M = 2.73^{\circ}$ ($SD = 0.69$), ranging from 1.25° to 3.95°.

Searching task Before the search task could begin, the experimenter guided the participants to a location approximately 60 m away from the searching site. This site was chosen so that the participants could not view the targets before the task began. Next participants were read the instructions for the searching task in which the experimenter detailed the boundaries of the searching site and showed the participants examples of their targets and distractors. They were informed that they needed to find all six targets and that the searching task would end when they found the sixth target. Next the experimenter asked the participants to search quickly and efficiently (without running) and then asked the participants to repeat the instructions and ask any clarifying questions they might have. Participants in the instructions conditions were given additional information. They were not only told what their targets were (the same information was given to both the varied and consistent participants), but also where to find the targets. For example, "Your target is the rectangle and those can be found on benches." Next the experimenter assisted the participant in showing the GPS watch face to the scene camera for later data processing (this was used as an indication of the beginning of the search task to offline coders) and the participant walked into the search site and began searching.

During the search task the experimenter walked at a discreet distance behind the participant and kept track of the order that the participant picked up each target. Each participant was instructed to ignore the experimenter to the best of their ability before the search task began. If the experimenter ever thought the participant was looking in their direction they were to look to the ground so as to not accidentally give away the location of either the targets or the distractors. When the participant found and retrieved the sixth target, the experimenter approached the participant and assisted with showing the GPS watch face to the scene camera for the last time (indicating the end of the search task) and turned the GPS watch recording off. The experimenter next guided the participant through the last calibration procedure and walked the participant back to the lab where the equipment was removed. The experimenter asked the participants in the regularity conditions a last question which was, "what did you notice about the location of your targets" and recorded whether they indicated the correct regularity in target location (i.e.,"my targets were on benches").

Data processing

For the dependent variables of this study we exported the time series of horizontal eye rotation and GPS coordinates for the searching task. The total length of the search path of each participant was calculated and used as a measure of search efficiency along with the total search duration. Additionally, we calculated a straightness ratio for each participant by taking the total length of the path used during search and dividing this value by the shortest possible path between the starting and stopping points (1.0 = a perfectly straight path). The spread of horizontal eye position (average standard deviation of horizontal eye position) was calculated to assess the spread of eye movements during search. Finally, the onset and offset of the search task was used to calculate the overall duration of the search task.

3.6 Results

One way analysis of variance (ANOVA) was calculated to assess differences between conditions (varied, consistent, and instructions) on the search duration, straightness ratio, length of search path, and spread of horizontal eye position. The GPS data for a single participant was missing due to technological issues during data collection. These data were assessed for outliers (greater than three standard deviations above or below the mean). One outlier was found for the straightness ratio and the data were subsequently removed from analyses.

Each of the analyses described below failed to reach significance (with an alpha of .05). No significant difference was found in the duration of the search task by condition, $F(2,\ 56) = 2.55$, $p = .09$, $\eta^2 = 0.08$ (Figure 3.1). No significant difference was found between conditions on the straightness ratio of the search path, $F(2,\ 54) = 1.27$, $p = .29$, $\eta^2 = 0.04$ (Figure 3.2). Similarly, no significant difference was found for the average standard deviation of horizontal eye position between conditions, $F(2,\ 56) = 1.44$, $p = .25$, $\eta^2 = 0.05$ (Figure 3.3), nor was any significant difference was found on the total length of the search path by condition, $F(2,\ 54) = 1.60$, $p = .21$, $\eta^2 = 0.06$ (Figure 3.4).

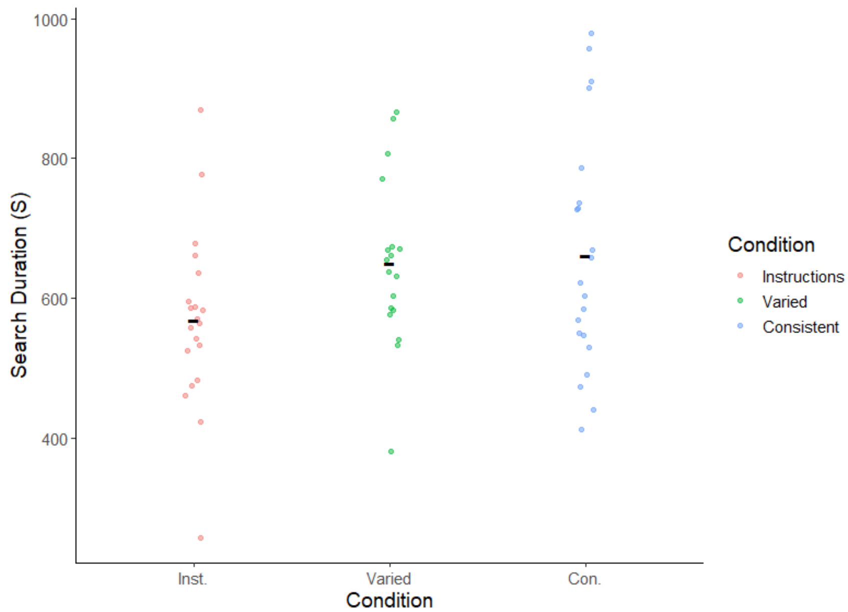

Figure 3.1: The duration of the search task by condition. Black bar represents the mean.

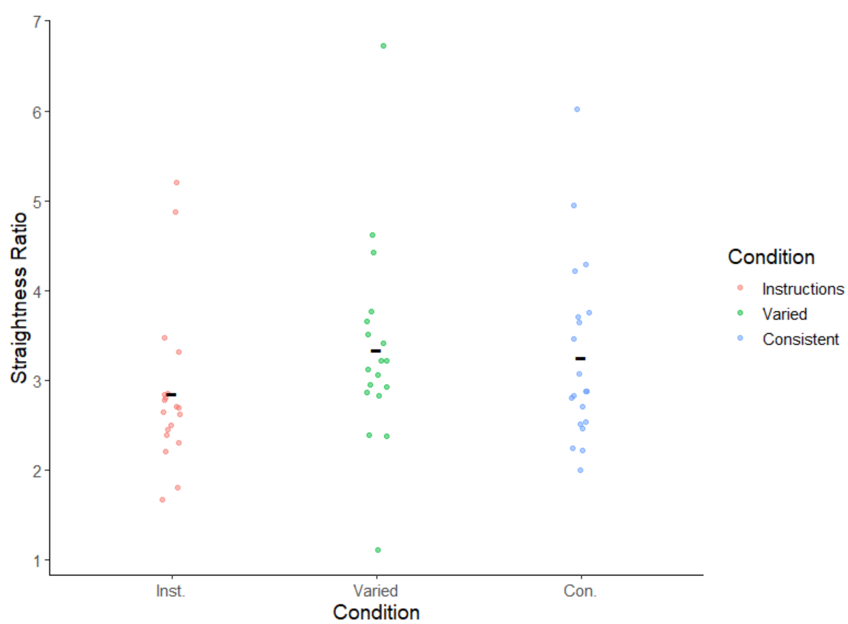

Figure 3.2: The straightness ratio of participant's search path by condition. Black bar represents the mean.

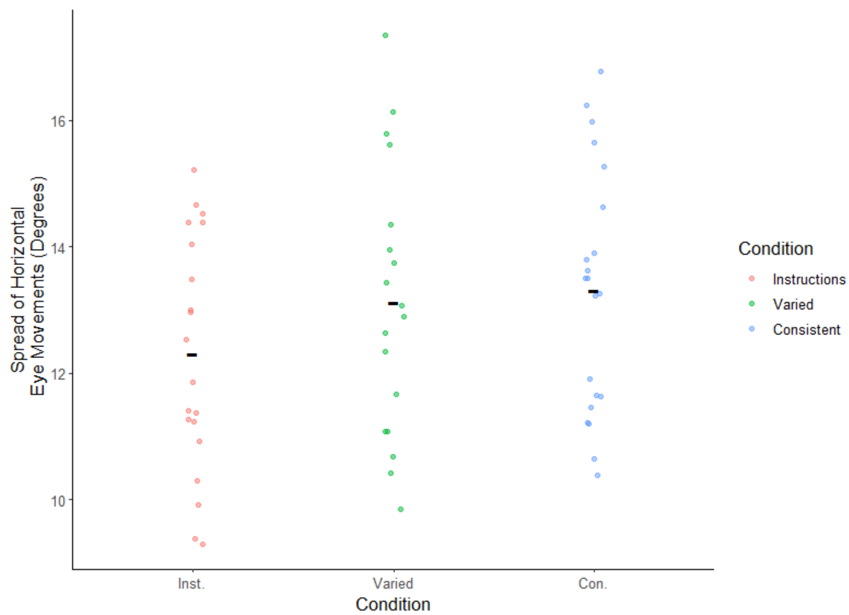

Figure 3.3: The spread of horizontal eye movements by condition. Black bar represents the mean.

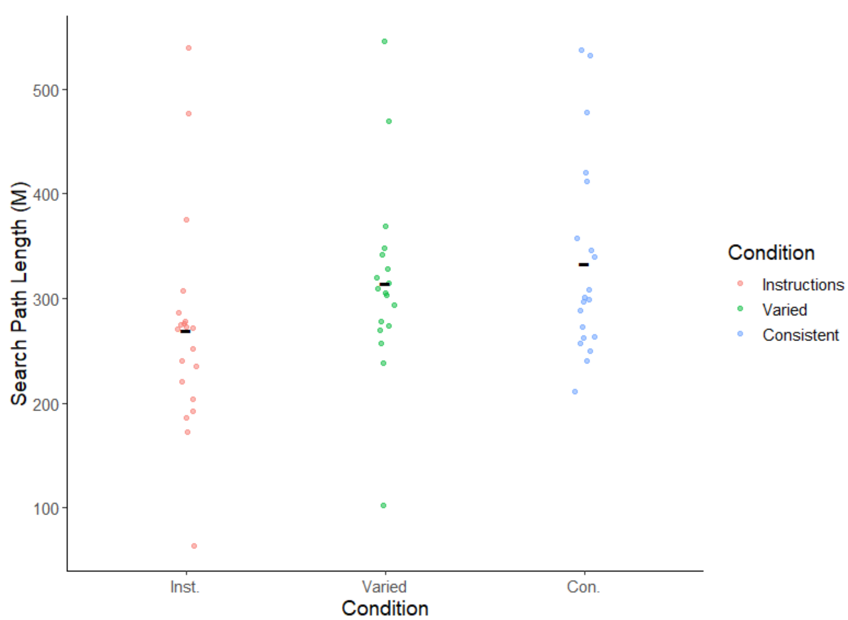

Figure 3.4: The search path length in meters by condition. Black bar represents the mean.

3.7 Discussion

The present study assessed the influence of implicit learning of underlying environmental statistics and prior knowledge on real-world search efficiency. During a real-world search task participants searched for six identical targets that varied between conditions in the probability of the targets being found in the same location types (targets in the varied location were found 50% on trees and 50% on benches while targets in the consistent and instructions conditions were found 100% on either trees *or* benches). To assess the effect of prior knowledge on search efficiency we informed participants in the "instructions" condition where to look for their targets (i.e., "your targets are on benches"). We hypothesized that participants in the instructions condition would demonstrate the most efficient search (as demonstrated by shorter search duration, smaller straightness ratio, etc.) over participants in either the consistent or varied conditions. We further hypothesized that participants in the consistent condition would be able to capitalize on the regularities associated with target location and would demonstrate this with more efficient search metrics compared to those in the varied condition. However, we found no evidence to support our hypothesized differences between conditions.

Prior research has demonstrated that top-down factors such as the ability to detect key environmental regularities or prior knowledge may influence key aspects of search efficiency such as the direction of the first head turn [111] or the overall search times [25]. In the present study we hoped to replicate these findings in a space large and open enough that it would be impossible to see the entirety of the environment while standing still, but also one in which there were both stationary (outdoor tables, garden planters, etc) and moving objects (people) for participants to contend with. In this manner we expected aspects of the search environment to necessitate bodily contributions (walking, bending, head and body movements, etc.) that might otherwise be unnecessary in smaller, more heavily controlled search environments.

While both predictions remain unsupported in the present analyses, it is possible that the current sample is under powered to detect the effect. While each of the four analyses failed to reach significance, participants in the instructions condition demonstrated the most efficient search as reflected in their overall group means. However, given the non-significant findings from the omnibus ANOVAs, we had no reason to further analyse between-group differences. Importantly, neither the current analyses nor the overall group means suggest any differences between the consistent and varied conditions. It is possible that six trials were insufficient for participants to detect the underlying regularities in target location. However, at the conclusion of the experiment participants from the consistent condition were asked "what did you notice about the location of your targets?". 19 participants (out of 20) correctly indicated that they detected the regularity by responding with the appropriate location (e.g., "my targets were all found on trees"). This suggests that either the vast majority of participants were aware of the underlying regularity during the experiment or the question itself prompted participants to reflect on their experience and come to this conclusion after the fact. However, the sheer proportion of participants who verbally indicated that they detected the regularity suggests that the manipulation was successful even if it was not reflected in these particular measures of search efficiency.

If the majority of participants in the consistent condition were able to implicitly learn the regularity associated with their target location, then why wasn't this reflected in any measure of their search efficiency? It is possible that implicit knowledge *alone* was not enough to influence their search, or perhaps that it was gained too late in the search process to sufficiently impact these gross measures of search efficiency. Future analyses should assess these data at the trial level (as opposed to the experiment level) to determine whether there are differences in early opposed to late trials. Differences at the trial level could lend evidence to support this hypothesis, as it may uncover when during the experimental session that the learning occured and whether it was then applied in such a way as to expedite the searching process that was not captured in the present aggregate measures. Significant change to the dependent variables listed above during the search task itself (e.g., a sharp decline in search duration between middle trials) could indicate that the regularity was detected, and if that change persists (e.g., search times remain low) that could indicate that the knowledge was then applied in such a way that enhanced search efficiency.

Further analysis of the existing data set is needed to fully assess possible differences between conditions. Additionally, it is also possible that the differences in this data set are not between conditions, but rather between individuals. That is to say that there are many possible strategies for efficient search and it is unlikely that each participant employed the same one. Given the multiple degrees of freedom for how we move our eyes, head, and bodies [64, 32] during search there are numerous ways to coordinate these elements of the perceptual system that the current analyses are unable to detect. Future analysis of this data set will assess the role of individual differences in regularity detection and search efficiency.

Furthermore, given the ongoing development of the perceptual and attentional systems during childhood, there is a developmental aspect to this question that remains unaddressed. Prior research has indicated that children differ from adults in real-world search in certain key areas such as their use of bottom-up factors (e.g., target conspicuity [87] or in aspects of their visually guided navigation (e.g., looking at obstacles during locomotion [30]. Future work in this area should assess how the role of top-down factors influences visual attention during real-world search across development. In this manner we would be able to better assess how the development of key aspects of the attentional system influences real-world search.

3.8 Acknowledgements

The experimental design and data collected from the present study are previously published [32]. The primary research questions and analyses of both papers are entirely distinct although related in topic.

Chapter 4

Adapting the coordination of eyes and head to differences in task and environment during fully-mobile visual exploration

4.1 Abstract

How are eyes and head adapted to meet the demands of visual exploration in different tasks and environments? In two studies, we measured the horizontal movements of the eyes (using mobile eye tracking in Studies 1 and 2) and the head (using inertial sensors in Study 2) while participants completed a walking task and a search and retrieval task in a large, outdoor environment. We found that the spread of visual exploration was greater while searching compared with walking, and this was primarily driven by increased movement of the head as opposed to the eyes. The contributions of the head to gaze shifts of different eccentricities was greater when searching compared to when walking. Findings are discussed with respect to understanding visual exploration as a motor action with multiple degrees of freedom.

4.2 Introduction

Visual exploration refers to the active process of looking around in the environment. Observers survey the environment by shifting their gaze from one location to another ("scanning") to gather visual information that supports ongoing activities [29, 36, 63]. The predominant paradigm for measuring visual exploration is recording eye movements in observers who look at screens. Although screen-based approaches yield valuable insights about how the eyes scan different types of photographs and videos, they are ill-suited for understanding visual exploration in the context of locomotion because observers must remain stationary. In contrast, mobile eye tracking studies have uncovered how gaze is adapted to different motor tasks, such as walking indoors to search an office mail room [25] or hallway [57, 103], walking outdoors over flat or uneven terrain [27, 22, 99, 71], or even participating in an outdoor geological field expedition [51]. Yet, mobile eye tracking studies, which can measure only the position of the eyes relative to the head, miss a well-appreciated but rarely studied aspect of visual exploration. In everyday life we coordinate the rotations of the body, head, and eyes to scan in all directions [61, 28, 29].

Gaze—where we look in the world—is the culmination of how we rotate the eyes in relation to the head, how we rotate the head in relation to the body, and how we orient the body in space. Combining mobile eye tracking with head tracking from wearable inertial sensors [99, 97, 71, 56] facilitates measuring how gaze depends on nested systems—rotations of the eyes within the head are added to rotations of the head within the body. With multiple degrees of freedom to control (i.e., the eyes, head, and body), how do observers coordinate visual exploration? As we will review in the next section, the eyes and head are subject to different biomechanical constraints and have different energetic costs that shape how they are used. In spite of these constraints, the few existing studies to simultaneously measure eye and head movements suggest that there is considerable flexibility in how observers explore *within* a task [81, 99, 40]. The primary aim of the current study is to ask how exploratory eye and head movements are differentially adapted to varying demands on attention created by different tasks/environments in the context of ongoing locomotion.

4.2.1 The roles of eye and head in visual exploration

The biomechanics of eye and head movements constrain how they can be coordinated to visually explore. The oculomotor range of the eyes is $\pm 55^{\circ}$ along the horizontal axis [39], meaning that shifts of gaze beyond this range require the head to rotate in the same direction as the eyes. Horizontal rotations of the head in combination with eye rotations allow total gaze shifts larger than 160°. Even larger gaze shifts require the trunk to rotate and/or the feet to reorient the body in space [45, 61]. With eyes, head, and body all able to contribute to a single gaze shift, there are multiple degrees of freedom to control. For example, a 20°-amplitude gaze shift can be accomplished in many ways, even when just considering the roles of eyes and head: A 20° eye movement alone with no head movement, a 10° eye movement with a 10° head movement, or a 5° eye movement with a 15° head movement all produce the same gaze result. How, then, does the visual-motor system determine how much the eyes versus head should contribute to a gaze shift?

Laboratory studies that elicit gaze shifts to targets at different amplitudes show that the eyes alone contribute to smaller-amplitude gaze shifts (less than 20°-30°), but for larger amplitude gaze shifts the head increasingly plays a role [39, 35]. It is important to note that the head contributes to gaze shifts smaller than 55°—the limit of the eyes alone—meaning that the head is recruited even when it is not biomechanically required. This allows the eyes to stay within a more comfortable range of $\pm 25^{\circ}$ [90]. Although eye and head contributions appear stereotyped in laboratory tasks that simply ask participants to move the eyes to fixate a target, experimental manipulations show that they are flexibly controlled. When instructed to make two sequential gaze shifts, the head contributes more to the initial gaze shift if the second gaze shift will be in the same direction [76]. In other words, observers are more willing to rotate the head when the head will stay rotated for a while. This speaks to the different *costs* of eye versus head movements. The eyes can move quickly with little effort, whereas the head moves more slowly and requires more energy [61, 40].

The contributions of eyes and head are even more variable when measured during complex tasks. Instead of asking participants to simply fixate targets, Pelz and colleagues (2001) [81] instructed participants to copy a model, placed to the side of the participant, by arranging blocks on a workspace in front of the body. Participants turned their eyes and head to shift gaze between the model and workspace while completing the task. Unlike more controlled studies, the head contributed between 1°-10° for smaller gaze shifts (less than 15° amplitude). Most likely, participants adapted eye and head rotations from moment to moment depending on the demands of looking to the model versus workspace (and scanning back and forth between the two locations). Participants' willingness to visually explore with eyes versus head may reflect the motor costs of each movement. Indeed, a variation of block-copying task that varied the angle of the model found participants looked less frequently at the model when looking required a larger body movement [18]. Similarly, participants comparing two similar-looking cupboards reduced the number of gaze shifts between the cupboards as the distance between the cupboards increased [40], presumably to reduce the number of costly head movements.

4.2.2 How might task demands shape visual exploration with the eyes versus head?

Despite these examples of how changing the motor costs of looking (e.g., placing targets closer or farther) alters the coordination of eyes and head *within a task*, no studies have investigated how eyes versus head are coordinated to meet the informational demands *across different tasks and environments*. Mobile eye tracking studies indicate that observers tend to fixate task-relevant objects when completing tasks such as making a sandwich or cup of tea [42, 67, 94, 62]. However, these examples—which measured eyes only—cannot reveal how both eyes and head are adapted to meet different task demands, given the flexibility and variability inherent in coordinating the eyes and head. Furthermore, locomotion—walking from one place to another—is a common "sub-task" that we must visually guide while completing a primary task, as seen in more natural tasks [31] and everyday life.

Several studies have described the role of the eyes and head in the control of walking over easy versus challenging terrain. Although these examples do not compare different task types, they demonstrate how participants adapt both eyes and head to respond to varying informational demands of locomotor control. Matthis and colleagues [71] found that in the less-demanding task of walking over flat terrain, only half of fixations were directed to the ground surface. *Spread* (or dispersion)—the standard deviation of position over the course of a task—is a commonly-used metric to examine differences in the distribution of visual exploration across tasks. 't Hart and colleagues [91] found that the horizontal spread of eye-plus-head gaze (∼14º) was greater than the vertical spread of gaze (∼7º), reflecting participants' propensity to visually explore targets to the left and right of the body rather than gazing down at the ground. Even though the 14º horizontal spread is well below the oculomotor range of 50º-55º, the head contributed to the horizontal spread of gaze: The horizontal spread of eye position was only 4º-5º, thus, the head accounted for the remaining portion. Similarly, Tomasi and colleagues [99] measured horizontal eye and head movements in walking participants using wearable inertial sensors, and found the head's rotation was responsible for between 37-46% of the total gaze shift amplitude across participants. Other studies of eye movements while walking over flat ground consistently find a larger horizontal than vertical spread of eye position: 14.2º versus 9.7º [104], 7º versus 5º [27], and 11.8 versus 7.2º [57].

Thus, the contributions of eyes and head during simple walking, that is, walking without a secondary task, are well characterized. Observers preferentially spread their gaze horizontally rather than vertically to visually explore the surroundings, but if walking is made more difficult the vertical spread of gaze extends down to better guide foot placement [71, 97, 91]. Moreover, the head contributes more than 35% of the rotation needed to shift gaze, even at amplitudes that are well within the limits of the oculomotor range. Our current studies build on this work to ask how eyes and head adapt to the addition of a non-locomotor task while walking, rather than altering the difficulty of walking. By adding a goal—searching for targets in a complex visual environment—we can compare the role of eye and head movements under different task demands. How might searching while walking alter the roles of eye and head compared with walking alone? Although searching may induce participants to make larger eye movement shifts to scan more broadly within a photograph [72], this may not translate to a fully-mobile searching task. A prior study of whole-body search in virtual reality found that participants primarily looked at mid-height regions rather than searching in areas above and below the body [55], thus, we expect search to primarily impact the horizontal component of gaze (especially with observers walking on flat ground). We predict that gaze will be spread more widely around the observer to successfully search compared to simply walking along a path. However, given the flexibility of coordinating eyes, head, and body, an increase in spread of gaze while searching could be accomplished in different ways: a larger spread of eye position without a change in head position, a larger spread of head position without a change in eye position, or increasing spread of both eyes and head. One possibility is that observers rotate the head more broadly to search in areas to the left and right of the current walking direction beyond the range of the eyes. Another possibility to rule out, however, is whether observers avoid extreme head rotations while searching if it disrupts their ability to guide locomotion. If so, we would observe an increase in the spread of eye movements but not head movements. It is important to note that we make no specific claim about the extent to which changes in the spread of eye or head movements might reflect conscious decision making. Although it is true that observers can consciously choose to employ greater head versus eye movements while exploring, it seems more likely—especially while engaged in a task like searching—that participants are not consciously deciding moment-to-moment how much to move the eyes versus head. Regardless, the current studies were not designed to distinguish between these possibilities.

4.3 Current study

Although previous research has demonstrated the role of task in shaping eye movements, no prior work has considered how observers adapt the coordination of eye and head movements to changing task and environment demands in the context of locomotion. Whereas eye and head movements have different constraints (e.g., speed, range of movement, energetic cost), there is considerable flexibility in how much the eyes versus the head contribute to looking in different directions. We choose to compare two types of naturalistic locomotor tasks, a simple *walking* task in which participants traversed a campus path, and a *search and retrieval* task in which participants walked around a cluttered campus courtyard to find and retrieve 6 hidden targets (referred to as the *search task* for brevity). Whereas the demands on visual exploration in the walking task were minimal—participants simply needed to stay on a flat, paved path—the searching task required participants to simultaneously scan their surroundings to find targets and to start, stop, and turn while walking from one place to the next. The courtyard contained picnic tables, trees, and open concrete areas, creating a more challenging visual scene to search in as participants' view of different areas was occluded. The novel contributions of the current studies are: 1) direct comparisons of visual exploration between walking and searching tasks, and 2) comparing head movements and eye-plus-head gaze shifts, not just eye movements, across tasks.

We report two studies that employed identical procedures but differed in the data recorded. In both studies, participants' eye movements were recorded using a mobile eye tracker, and participants' walking behaviors were recorded with a GPS monitor worn on the wrist to understand the locomotor aspects of the two tasks. Study 2 added wearable inertial sensors that measured participants' head rotations. While wearing the eye tracker, GPS monitor, and (in Study 2) inertial sensors, participants completed the walking task by following a campus path from the Psychology building to an outdoor courtyard. Afterwards, participants completed the search and retrieval task in the courtyard by finding and picking up six targets (fabric squares marked with a particular shape) placed in different locations, while ignoring six distractor targets (similar looking fabric squares with a different shape).

We calculated how the *spread* of visual exploration differed between the the two tasks based on the horizontal rotation (in degrees) of the eyes (Studies 1 and 2) and head (Study 2). As in past work [104, 27, 57], spread was defined as the standard deviation of the horizontal rotation of the eyes/head and represented the degree to which participants distributed their visual exploration narrowly versus broadly over the duration of each task. As in other studies [99], we focused on horizontal eye and head movements because horizontal gaze movements are more common than vertical gaze movements when walking over flat terrain [91, 27]. Study 2 also provided an opportunity to extend laboratory studies that measured the contribution of the head to gaze shifts of varying eccentricity to a more naturalistic task. By calculating the total amplitude of each gaze shift (adding the rotations of eyes and head together), we could determine the *head contribution* (in percentage) of each gaze shift and whether that varied according to task. We predicted that the head would increasingly contribute to larger amplitude gaze shifts regardless of task, consistent with previous laboratory studies [39, 35]. Moreover, we predicted that the head would contribute more to gaze shifts in the searching task to facilitate a wider spread of gaze in the environment.

4.4 Study 1: How are eye movements adapted to explore in different tasks/environments?

4.4.1 Method

The study's procedures were designed in accordance with the Declaration of Helsinki. The UC Riverside Institutional Review Board approved the project (HS-14-137 "Eye movements during everyday activities") before data collection began. Participants gave written informed consent before the study began.

Participants

The final sample consisted of $N = 59$ adult participants between the ages of 18 and 31 years ($M = 20.81$ years, $SD = 3.0$, 39 female, 20 male). One additional participant was run in the study, but their data were excluded from the final sample after their eye-tracking error was found to be unusually large ($> 5^{o}$). To be included in the study, participants were required to have normal vision or corrected-to-normal-vision with contact lenses (eye glasses could not be worn with the eye tracking headgear) and to have no motor impairments that would prevent them from engaging in the tasks. Additional participants were run in the study but excluded before data processing due to bystander interference ($n = 2$), technical difficulties (e.g., battery or SD card failure) ($n = 4$), or because the camera slipped during the searching task ($n = 7$).

Participants were undergraduate students at the University of California, Riverside who received course credit as compensation for their participation. Written informed consent was obtained at the beginning of the experimental session. Participants described their race as: White ($N = 20$), Asian ($N = 15$), Black ($N = 2$), more than one race ($N = 8$), or chose not to answer ($N = 14$). Participants described their ethnicity as: Hispanic or Latinx ($N = 27$), Not Hispanic or Latinx ($N = 29$), or chose not to answer ($N = 3$).

Walk and search task settings

The walking task took place along a 311-m path in the University of California, Riverside campus. Participants walked East for approximately 26 m, North for 150 m, then East for 135 m on paved sidewalks. This path took participants in between closely spaced buildings and also through a wide, open field. The walking path ended 60 m away from the courtyard, ensuring that participants could not see search target locations before they began the search task. The search arena was a courtyard that measured 45 m wide × 30 m long for a total area of 1350 m^2. Approximately 823 m^2 was garden space inaccessible to pedestrians. The remaining space was comprised of mature trees, seating areas, and wide cement walkways. Both the walking path and search arena were open to the campus public, so pedestrians were often present while participants walked through both areas. Examples of one participant's GPS location overlaid on a campus map is shown for the walking and searching tasks in Figure 4.1A. An example video available at `https://nyu.databrary.org/volume/1147` shows excerpts from the walking and searching tasks.

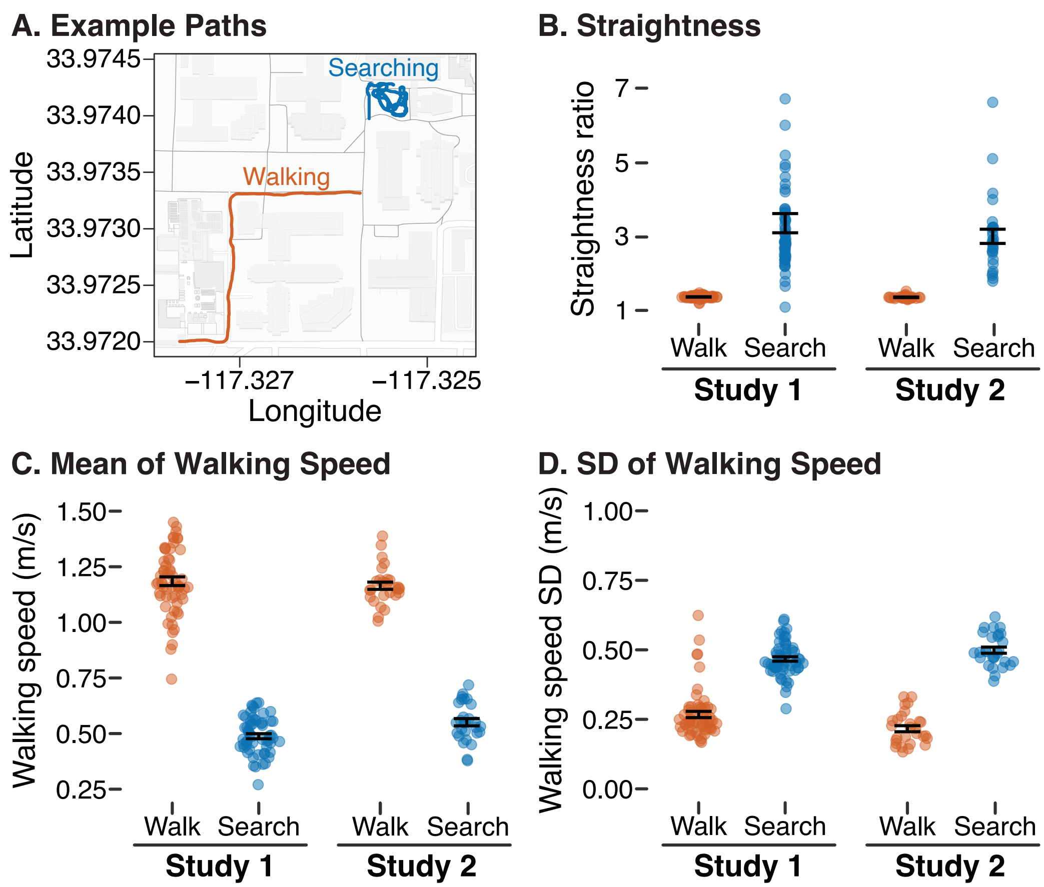

Figure 4.1: Characteristics of locomotion derived from GPS data in the walking and searching tasks (orange = walking task, blue = search and retrieval task). A) Example GPS recording of a participant's path in the walking and searching tasks overlaid on a campus map. Graphs show differences in B) straightness ratio, C) mean walking speed, and D) SD of walking speed for Studies 1 and 2 according to task. Each symbol represents a single participant's data; points are horizontally offset for visibility. Black error bars are centered on the mean and show ±1 standard error.

In the search arena, targets and distractors were fixed to trees and cement benches located throughout the courtyard in a pre-specified set of locations. Targets and distractors were 10 cm × 10 cm orange fabric squares with a 3.8 cm × 2.5 cm shape (rectangle or diamond) drawn on the front in black ink. Of the targets affixed to trees (6 total), 2 targets each were each secured 0.25 m from the ground, 1 m from the ground, and 1.5 m from the ground. Of the targets affixed to benches, targets were secured to the frame of the cement benches, never the seats or legs (each bench measured approximately 1.5 m long × 0.3 m wide).

Eye movement and GPS recording

A Positive Science head-mounted eye tracker was used to record the eye movements of each participant. An infrared camera that pointed towards the participant's right eye (eye camera) recorded eye movements, and the field of view of each participant was recorded by a camera that sits above the right eye and points out (field of view camera). Both eye and field of view (FOV) cameras were affixed to a modified eye glass frame that was securely hooked over each ear and held onto the participant's head with a strap. Each camera's video was fed to a recording device that was stored in a belt bag that participants wore over their right shoulder for the duration of the study. Participants wore a wide brimmed hat to reduce eye tracker data loss from sunlight [71] and a Polar V800 Multisport GPS watch on their right wrist. The example video (`https://nyu.databrary.org/volume/1147`) shows real-time eye position and GPS data for an example participant.

Before the start of each task and at the end of the study, participants completed a calibration procedure that maps participant's eye position from the eye camera to their gaze location in the FOV camera. During the calibration procedure, the experimenter stood approximately 3 m from the participant and asked the participant to hold their heads as still as possible while moving only their eyes to look at locations that the experimenter indicated. The experimenter cued the participants to look at a walking stick with a brightly colored piece of cardboard at one end. The experimenter moved the colored calibration target in different locations within the FOV camera's field of view: along the central, vertical axis (top to bottom), along the horizontal axis (left to right), and along both diagonals (from corner to corner). The experimenter periodically stopped the target to allow the participant time to fixate on the calibration target without blinking or moving their head.

These video recordings were used offline (after the session) to calibrate the eye tracker using Yarbus software (Positive Science LLC), producing horizontal and vertical time series of gaze locations in field of view video (pixel) coordinates. Calibration accuracy was verified using an additional set of 5 target looks, independent from those used to calibrate the eye tracker. Calibration validation was done at the end of the walking task and at the end of the search task. For each validation point, we calculated the difference between the actual target location in the FOV camera and the gaze location in degrees—calibration error. In Study 1, participants' calibration error averaged $M = 2.73^{\circ}$ ($SD = 0.69$), ranging from 1.25° to 3.95°.

Procedure

Participants were fitted with the head-mounted eye tracker, hat, belt bag and GPS watch in the laboratory. Afterwards, the experimenter led them to a flat, shady, area outdoors for the first eye tracker calibration. The GPS watch was turned on after the calibration; this event was recorded in the eye tracker's FOV camera to allow synchronization. Next, participants completed the walking task along the prescribed path. The experimenter walked alongside the participant, providing verbal directions about where to go. At the conclusion of the walking task, the participant completed the second eye tracker calibration to account for any potential movement of the eye tracking equipment that may have occurred during the walking task.

Before the start of the search and retrieval task, the experimenter read instructions that detailed the boundaries of the search arena, explained how to identify the assigned targets versus the distractors, and how many targets were hidden (6 targets and 6 distractors). Participants were instructed to pick up each of the six targets with their hands and to leave the distractors in place. Participants were told to retrieve their targets as quickly and efficiently as possible, without running. After hearing the instructions, the search and retrieval task began. A final calibration check after the search task ensured the accuracy of the eye tracking data throughout the task.

Data processing

The first step in data processing was to synchronize the eye tracking and GPS time series data. The FOV camera frames that corresponded to the the GPS watch turning on/off were recorded from the FOV camera video. Using those synchronization points, we offset, scaled, and upsampled (from 1 Hz to 30 Hz) the GPS time series to match the eye tracker's time series. FOV camera videos from the eye tracker were also used to find and record the beginning and end times of each task. After synchronization, time series were extracted for horizontal eye rotation and GPS coordinates during each task to be used in subsequent analyses.

GPS coordinates were used to calculate three measures to characterize how participants walked during each task. *Walking speed* was calculated based on the length of each participant's total walking path in each task divided by the task time. *Walking speed SD* measured the amount that participants changed their speed during each task (e.g., stopped and started walking) by calculating their instantaneous speed for each video frame, and then calculating the standard deviation of instantaneous speed across the task. Finally, the degree to which participants walked a straight path versus a circuitous path was expressed by the *straightness ratio*: the total length of the walking path divided by the shortest path between the starting and stopping points (1.0 = a perfectly straight path). Although it is expected that paths while walking will be straighter compared with paths while searching, we report these values as a way to characterize the degree of straightness to compare with future work.

Horizontal eye gaze coordinates represented how much participants rotated their eyes from left to right within the FOV camera image, measured in pixels. In order to measure eye-in-head rotations in degrees, we converted pixels to degrees based on the camera's horizontal field of view, 111º. However, the wide-angle fisheye lens meant that the pixel-to-degrees calculation could not be performed without first correcting for lens distortion [99]. We used the Matlab "Camera Calibration Toolbox" to correct the points for lens distortion before converting to degrees of visual angle. A checkerboard test image was recorded with the FOV view camera, which allowed the toolbox to create a model of the lens. The `undistortFisheyePoints` function was then used to transform each participant's raw eye movement data to remove the lens distortion. After this transformation, the eye movement data were then converted from pixels into degrees.

Using the corrected horizontal eye movement data (in degrees of rotation), we determined how much participants distributed horizontal eye movements widely versus nar-

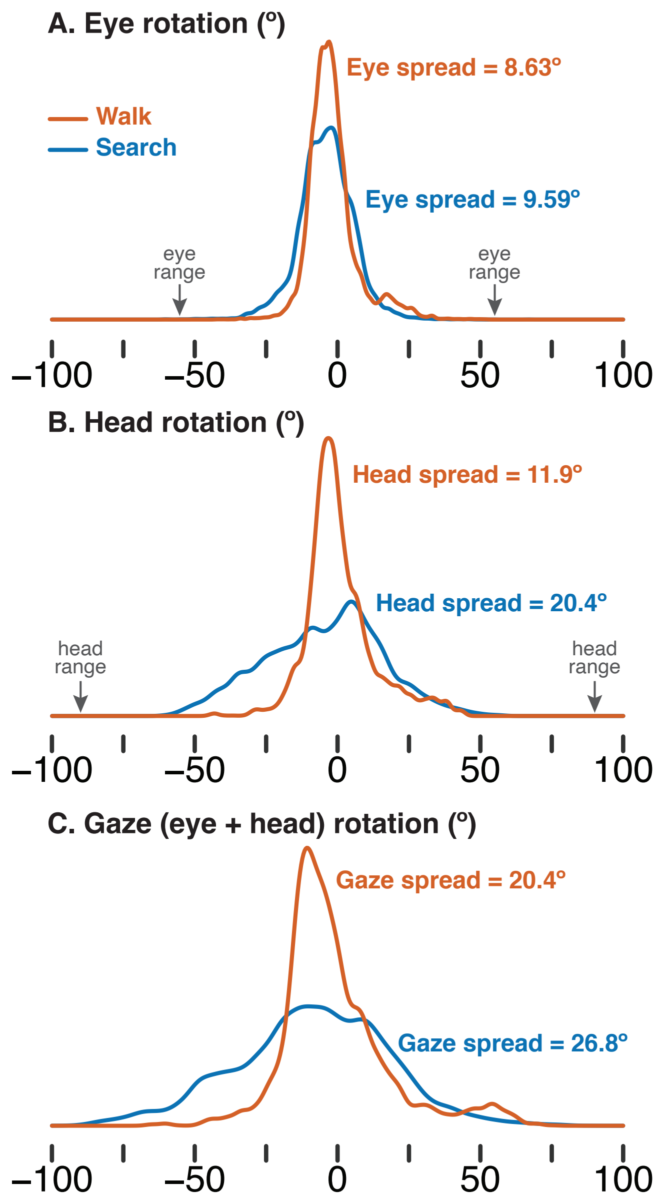

Figure 4.2: Example density plots of (A) eye rotation, (B) head rotation, and (C) gaze rotation (eyes-plus-head) for one participant. Orange lines show the distribution for the walking task and blue lines show the distribution for the search and retrieval task. Arrows indicate the approximate biological limits on (A) eye rotation and (B) head rotation for reference. Text labels show the spread of visual exploration (*SD*) based on the rotation data from each task.

rowly by calculating *spread*: The standard deviation of horizontal eye position (in degrees) across each task. Figure 4.2A shows one participant's eye rotation distributions and corresponding spread measures in the walking and searching tasks.

4.4.2 Results and discussion

Analyses were conducted in *R* [85]. Paired t-tests were used to calculate the difference in each measure between walking and searching tasks. We checked for outliers based on a threshold of 3 SD around the mean within a condition, but no outliers were found. The dataset and analysis code are shared in a reproducible "capsule" on CodeOcean (`https://doi.org/10.24433/CO.8767371.v2`).

Locomotion differed between the tasks/environments

The walking task time averaged $M = 268.1$ s ($SD = 36.5$) with participants walking a total distance of $M = 313.0$ m ($SD = 23.8$). The searching task time averaged $M = 625.4$ s ($SD = 145.7$) with participants walking a total distance of $M = 305.7$ m ($SD = 99.2$). Analysis of locomotion from GPS data illustrated the differences in behavior between the walking and searching tasks (Figure 4.1B-D. In the walking task, participants' paths were straighter (straightness ratios closer to 1.0, $M = 1.38$, $SD = 0.05$), they walked more quickly (speed $M = 1.18$ m/s, $SD = 0.15$), and they walked at a more regular pace (speed SD $M = 0.28$ m/s, $SD = 0.09$). In the search and retrieval task, participants walked a more circuitous path (straightness ratio farther from 1.0, $M = 3.37$, $SD = 1.99$) at a slower average speed (speed $M = 0.49$ m/s, $SD = 0.09$), and their speed varied considerably from moment-to-moment while switching between searching for targets and stopping to retrieve them (speed SD $M = 0.47$ m/s, $SD = 0.06$). Significant paired t-tests were found comparing straightness ratios ($t(57)$ = -7.63, $p < .0001$, d = -1.00), average walking speed ($t(57)$ = 33.6, $p < .0001$, $d = 3.92$), and walking speed SDs ($t(57)$ = -15.0, $p < .0001$, d = -1.97) between the two tasks.

Visual exploration differed across tasks/environments

Figure 4.3 (Study 1) shows that the horizontal spread of eye movements was greater in the search and retrieval task ($M = 12.9^{o}$, $SD = 2.00$) compared with the walking task ($M = 11.7^{o}$, $SD = 2.59$). When searching for targets, participants spent longer periods of time with their eyes rotated far to the left/right, whereas participants kept their eyes in a more narrow range within their orbits when walking without searching. This difference was confirmed by a significant paired-samples t-test between walking spread and searching spread, $t(58)$ = -4.18, $p = .0001$, d = -0.54. Thus, participants adapted their eye movements to fit each task. With little demand on visual attention in the walking task, participants kept their eyes in a narrow window centered within the head. In contrast, participants who searched and retrieved targets broadened the scope of their eye movements to spread their gaze while looking for targets.

4.5 Study 2: How are eye and head movements adapted to explore in different tasks/environments?

Study 1 indicated that participants adapted the spread of eye movements to fit the demands of the task. When walking along a straight, uniform path with no other demands

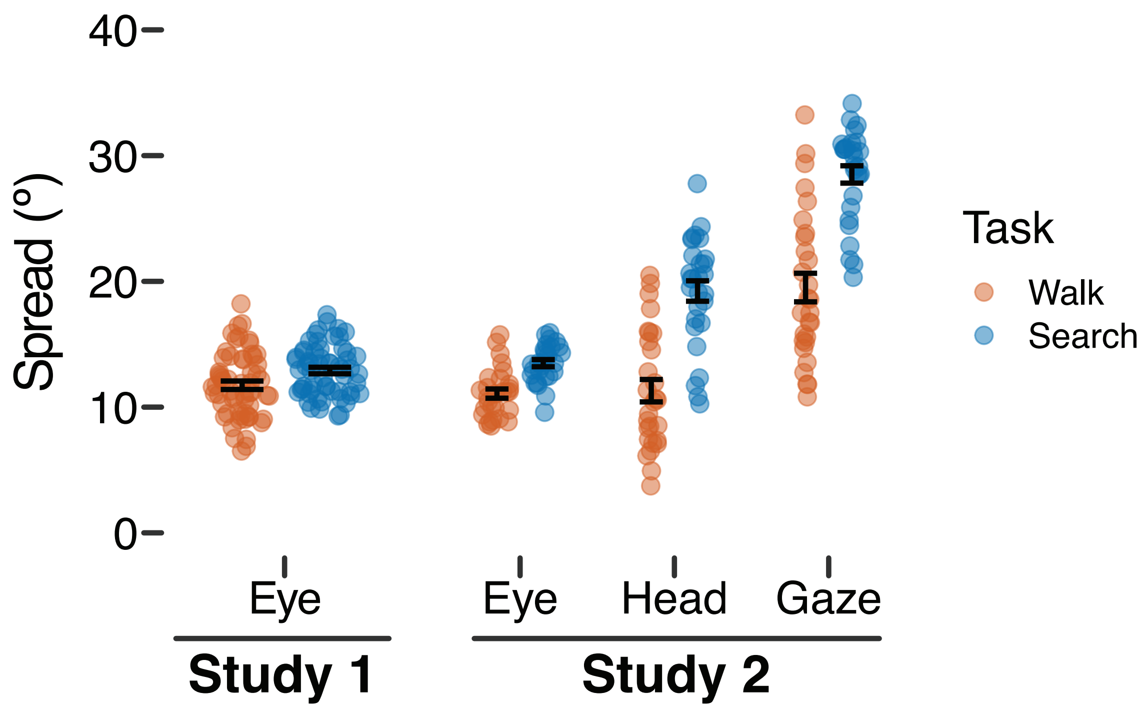

Figure 4.3: Horizontal spread (standard deviation of rotational position in degrees) in the walking task (orange symbols) versus search and retrieval task (blue symbols). Study 1 shows spread for horizontal eye movements, and Study 2 shows spread for eye movements, head movements, and gaze-in-body (eye-plus-head rotation). Each symbol represents a single participant's data; points are horizontally offset for visibility. Black error bars are centered on the mean and show ±1 standard error.

on attention, participants moved their eyes within a small area. In contrast, when searching and retrieving targets participants' eyes were often rotated in different directions (within the head). However, because gaze direction in the world, relative to the body, depends on both eye and head rotation, Study 1 could not measure how much gaze was spread in different directions. It is possible that the more extreme rotations of the eyes during the search task were oppositional movements to compensate for head rotation. If so, the observer would not truly be spreading gaze more in the searching task compared with the walking task. Alternatively, if participants in the search task rotated their eyes and heads more in the same direction at the same time, then the spread of gaze when searching would truly be greater. Thus, Study 2 was designed to extend Study 1 by measuring head rotation.

4.5.1 Method

The study's procedures were designed in accordance with the Declaration of Helsinki. The UC Riverside Institutional Review Board approved the project (HS-14-137 "Eye movements during everyday activities") before data collection began.

Participants

This study included $N = 28$ undergraduate students at the University of California, Riverside between the ages of 18 and 24 years old ($M = 20.29$ years, $SD = 1.43$, 16 male, 12 female). Participants were recruited from the psychology department participant pool at the University of California, Riverside and received course credit as compensation for their participation in this study. To be included in the study, participants needed to have normal or corrected-to-normal vision without wearing eyeglasses and were required to have no motor impairments that would prevent them from engaging in the tasks. Each participant gave informed consent at the beginning of the experimental session. Participants described their race as: Asian ($N = 13$), White ($N = 7$), Black ($N = 2$), Native Hawaiian or other pacific islander ($N = 1$), more than one race ($N = 1$), or chose not to answer ($N = 4$). Participants described their ethnicity as: Hispanic or Latinx ($N = 8$) or Not Hispanic or Latinx ($N = 20$). Five additional participants completed the study, but their data were ultimately excluded from the final sample due to technical difficulties ($n = 3$), the camera slipping out of place during the searching task ($n = 1$), and bystander interference ($n = 1$).

As in Study 1, calibration validation was performed at the end of the walking task and at the end of the search task. Calibration errors for the 28 participants averaged $M = 3.50^{o}$ ($SD = 0.81$), ranging from 1.70^{o} to 4.57^{o}.

Head movement recording

All procedural aspects of Study 2 were equivalent to Study 1, with the only change being the addition of wearable inertial motion sensors that recorded head position. Two STT systems (STT-IWS) motion sensors were worn throughout the duration of the entire study. One sensor was placed on the seventh cervical vertebra (C7) using a Velcro chest harness and the other was secured on top of the participant's head (underneath the wide-brimmed hat) with a Velcro headband. To facilitate synchronization of the motion sensors with the eye tracking data, participants were instructed before each eye tracking calibration to hold their heads still and look straight ahead and then to make a quick head rotation to the left then right.

Data processing

Measures of walking from GPS data and measures of eye movement spread were processed as in Study 1. To integrate head rotation measures with eye movement and GPS data, we extracted head rotation time series from the STT systems using their proprietary iSen software. The software calculated time series of head position (400 Hz) from the acceleration and gyroscope data collected by the head sensor, using the C7 sensor as a reference point. To synchronize the head movement time series to the eye-tracking time series, we identified the head-turn synchronization events in the eye tracker's FOV camera video (moment that the field of view changed during the rapid head rotation) and the matching timestamp from the head rotation time series data from a plot. Based on the synchronization event times at the beginning and end of the session, we offset, scaled, and downsampled the head rotation data to match the eye movement and GPS time series. Eye movement data were undistorted and converted into degrees as in Study 1, resulting in synchronized time series of horizontal eye and head rotation in the same measurement units. The example video (`https://nyu.databrary.org/volume/1147`) shows head rotation data synchronized with eye rotation and GPS.

Head rotation spread was calculated in the same way as eye movement spread. In addition, we calculated a gaze-in-body time series by adding eye and head rotations together (negative rotations corresponding to left, 0 corresponding to center, and positive rotations corresponding to right). We calculated gaze-in-body spread from this time series (the standard deviation of gaze position) to determine the overall distribution of gaze relative to the observer's body. Figure 4.2B-C shows one participant's head and gaze rotation distributions and corresponding spread measures in the walking and searching tasks.

4.5.2 Results and discussion

We compared locomotion (straightness, walking speed, and walking speed SDs), visual exploration (spread of eye/head movements), and the head contribution to gaze shifts across tasks. With the additional factor of eye versus head movements, we employed linear mixed-effect models (LMMs) in R using the *lme4* package [5] with participant as a random effect. Maximal models that included random slopes of fixed factors failed to converge, so only random intercepts of participant were included. Significance tests for LMMs were calculated using the *lmerTest* package [59] implementation of the Satterthwaite correction. Pairwise follow-up tests were corrected for multiple comparisons using the Holm-Bonferroni correction. All measures were checked for outliers according to a 3-*SD* criterion, but none were found. The data and analysis code are available in the same CodeOcean capsule as Study 1 (`https://doi.org/10.24433/CO.8767371.v2`).

Locomotion differed across tasks/environments

The walking task time averaged $M = 279.7$ s ($SD = 20.5$) with participants walking a total distance of $M = 324.2$ m ($SD = 12.0$). The searching task time averaged $M = 731.1$ s ($SD = 157.0$) with participants walking a total distance of $M = 407.6$ m ($SD = 125.2$). The three GPS-derived measures of locomotion differed according to task, mirroring the results of Study 1 (Figure 4.1B-D). When completing the walking task, paths were straighter ($M =$ 1.37, $SD = 0.04$), walking speed was greater ($M = 1.16$ m/s, $SD = 0.09$), and they walked at a more regular pace (speed SD $M = 0.22$ m/s, $SD = 0.06$). When searching, paths were less straight ($M = 3.02$, $SD = 1.01$), average walking speeds were slower ($M = 0.55$ m/s, $SD = 0.09$), and speed varied more (speed SD $M = 0.50$ m/s, $SD = 0.06$). Significant paired t-tests were found comparing straightness ratios ($t(27) = -8.57$, $p < .0001$, $d =$ -1.62), average walking speed ($t(27) = 24.3$, $p < .0001$, $d = 4.6$), and walking speed SDs ($t(27) = -18.6$, $p < .0001$, $d = -3.51$) between the two tasks.

Visual exploration differed across tasks/environments

Figure 4.3 shows the spread of visual exploration for the eyes, head, and gaze (eyes-plus-head) for Study 2. Consistent with our prediction, gaze was spread more broadly during the search task (M = 28.5º, SD = 3.65) compared with the walking task (M = 19.5º, SD = 6.01; $t(27)$ = -8.82, $p < .0001$, d = -1.67).

How were eyes and head adapted between the walking and searching tasks to spread gaze-in-body more broadly when walking and searching? We used a 2 task (walking vs searching) × 2 effector (eyes vs head) LMM to model spread based on task and effector as fixed factors and participant as a random intercept. Replicating Study 1, and consistent with the gaze result in the previous paragraph, a significant main effect of task, $F(1, 81)$ = 80.18, $p < .0001$, indicated that spread was greater when searching compared with walking. A significant main effect of effector, $F(1, 81)$ = 26.68, $p < .0001$, and a significant task × effector interaction, $F(1, 81)$ = 22.70, $p < .0001$, reveal that the increase in gaze spread from walking to searching was more dependent on the head compared with the eyes. When walking, the spread in head position (M = 11.3º, SD = 4.72) and eye position (M = 11.1º, SD = 1.94) were similar, and spread did not significantly differ in a pairwise comparison between eyes and head (p = .77). In contrast, head position spread in the searching task (M = 19.2º, SD = 4.30) was significantly greater than the spread in eye position (M = 13.5º, SD = 1.5; $p < .0001$).

Thus, the spread of both eye and head movements increased from walking to searching, allowing gaze to be distributed more broadly in the environment when looking for and retrieving hidden targets. However, the adaptation of spread was more pronounced in head movements compared with eye movements.

Head contribution to gaze shifts differed across tasks/environments

The final set of analyses examined the *head contribution* to gaze shifts to different eccentricities relative to the body in the two tasks. Using the gaze-in-body time series, we identified local minima (shifts to the left of the body) and maxima (shifts to the right of the body) using Matlab's *findpeaks* function. Peaks were required to be a minimum of 10 video frames (333 ms) apart and were only recorded during times that both eyes and head were rotated in the same direction. For each peak, we calculated the head contribution as the percentage of the gaze shift accomplished by the head. For example, if the eyes rotated 20º to the left and the head rotated 20º to the left for a combined eccentricity of 40º, the head contribution would be half (50%) of the total eccentricity. Figure 4.4 shows three examples of the head's contribution to gaze shifts of different eccentricities (the black arrow indicates the total eccentricity of the shift, the green shaded region indicates the amount the head rotated, and the gray region represents the additional rotations of the eyes). In order to analyze the relative contribution of the head as a function of the total eccentricity of the gaze shift, we found each participant's average head contribution by aggregating over peaks in eight 10º-wide bins (i.e., total shifts 10º-20º, 20º-30º, 30º-40º, 40º-50º, 50º-60º, 60º-70º, 70º-80º, and 80º+). In Figure 4.4, each bin is labelled by the lower bound of the bin (e.g., 10º-20º is labelled 10º).

Figure 4.4 shows that more eccentric gaze shifts recruited a greater head contribution in both walking and searching tasks, suggesting that previous laboratory results [39, 35] generalize to a naturalistic locomotor task. Unlike laboratory tasks, the head contributed to even the smallest shifts of gaze (10º). Visual inspection of Figure 4.4 suggests that eyes

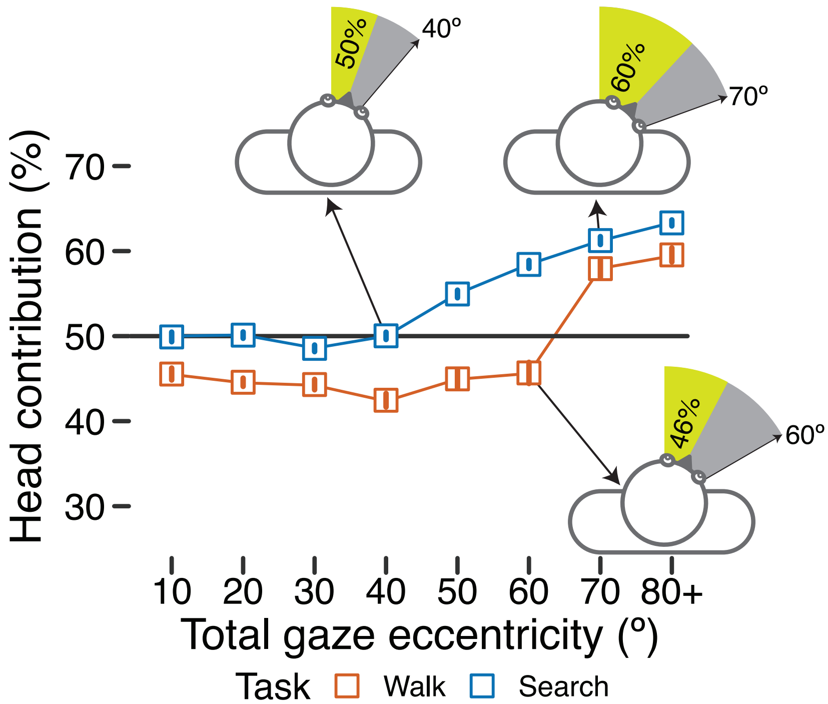

Figure 4.4: Head contribution to gaze shifts of varying eccentricity (x-axis). Each symbol shows the mean head contribution to a gaze shift—the percentage of the gaze shift accomplished through head rotation as opposed to eye rotation. Symbols above the black horizontal line at 50% indicate that the head contributed more than the eyes; symbols below 50% indicate that the eyes contributed more to the gaze shift compared to the head. Orange symbols represent the walking task and blue symbols represent the search and retrieval task. Error bars (within the symbols) indicate ± 1 standard error. Three top-down drawings of an observer depict the eye contribution (gray shading) versus head contribution (green shading) for gaze shifts at 40º in the search task, 60º in the walking task, and 70º in the search task.

and head played consistent, near-equal roles at smaller eccentricities (less than 50º), but the head increasingly contributed at larger eccentricities. However, the head contributed more in the searching task compared with the walking task at every eccentricity. These results were confirmed by a 2 task × 8 eccentricity LMM on head contribution with random intercepts by participant, which revealed a significant main effect of task, $F(1, 425.12) = 93.58$, $p < .0001$, and a significant main effect of bin, $F(7, 424.97) = 8.02$, $p < .0001$. Although it appeared that the increase in head contribution occurred at different eccentricities for each task (between 40º-50º for the searching task but between 60º-70º for the walking task), the task × eccentricity interaction was non-significant ($p = .14$). Pairwise comparisons between tasks at every eccentricity were statistically significant (ps $< .047$), confirming that the head contributed more when searching regardless of the eccentricity of the gaze shift.

4.6 General discussion

To summarize, the current study investigated how task and environment affect the spread of eye and head visual exploration during outdoor locomotion. We found that that eye and head movements are adapted differently when walking along a path (walking task) compared with walking around a cluttered courtyard while searching for and retrieving targets (searching task). More specifically, individuals spread their gaze (relative to the body) more broadly during the search and retrieval task compared to the walking task through a large increase in the spread of head movements paired with a modest increase in the spread of eye movements. We also extended a laboratory effect—that the head contribution to a gaze shift increases as a function of the amplitude of a gaze shift—to show that it holds in walking observers, and additionally showed that the degree of head contribution changes depending on the task/environment. The head's contribution to gaze shifts was greater while searching compared to when walking for gaze shifts of every amplitude.

There is abundant research from both screen-based [89, 21, 9, 72] and mobile eye tracking studies [42, 81, 94] showing that eye gaze is adapted to the observer's task. As expected, we found in Study 1 that the spread of eye movements increased modestly when searching compared with walking (12.9º versus 11.7º). Given that the horizontal eye spread in previous walking studies ranged from 5º-14º [104, 91, 57, 27], a task difference of 1.2º appears quite small, even though it was statistically significant. Yet, measuring the eyes alone tells only part of the story. As expected, the degree to which gaze-in-body changed between tasks was large (28.5º for searching versus 19.5º for walking in Study 2), demonstrating that the two tasks placed very different demands on visual exploration that were not apparent from examining the movements of the eyes alone. Indeed, the largest adaption was evident in movements of the head, with a spread of 19.2º in head position observed while searching compared to only 11.3º while walking. The differential contributions of eyes and head show the value of measuring head position during visual exploration. Research using eyes-only measures of visual exploration should be especially cautious in the treatment of null effects if the head's contribution is not characterized.

Given the winding, circuitous paths participants took through the courtyard when searching (Figure 4.1), it was expected that participants would distribute gaze more broadly around the environment to explore while searching. However, the flexibility in how the eyes, head, and body can contribute to gaze shifts means that the eyes alone, the head alone, or eyes and head in different combinations could have been adapted to meet the demands of the searching task. Indeed, the gaze density plot in Figure 4.2 (bottom) shows that most shifts of gaze were well within the biomechanical range of the eyes and head. But despite the multiple degrees of freedom afforded to participants, they arrived at a similar solution: increasing the spread of *both* eyes and head when searching, but increasing the spread of the head by a greater degree. Whether this is the most optimal or efficient strategy remains to be tested. Indeed, we cannot claim from the present work that energetic cost is the critical factor in shaping how eyes versus head contribute. Although head movements are more energetically costly, they also generate vestibular and proprioceptive information that eye movements do not. Future work could experimentally restrict head movement or increase the energetic cost of head movements to determine: 1) whether the eyes compensate by increasing their spread when head movement is reduced, and 2) whether a diminished contribution of the head to visual exploration degrades search performance.

Finally, measuring concurrent eye and head movements afforded us an opportunity to ask how the eyes and head contribute to gaze shifts of varying amplitude. Whereas the comparisons of head versus eye speed/spread were temporally coarse (aggregating across the entire task), measuring the the eye and head contributions to each gaze shift showed how they were coordinated in the moment. Like Tomasi and colleagues [99], who studied eye and head rotations in natural outdoor locomotion, we replicated the laboratory finding that the head contribution to gaze shifts increases as the total amplitude increases [90, 39, 35]. Our investigation extends those prior studies to show that this is true both while walking and searching in more naturalistic situations. Moreover, our study adds a novel finding: The relative contributions of eyes and head change as a function of task/environment, not merely amplitude, as evidenced by an overall greater head contribution in the search task. This suggests that the overall strategy of visual exploration changed in the searching task—the head was not just recruited to look at extreme locations, but contributed more to visual exploration in all locations. Perhaps, the head contributed more to smaller shifts of gaze in the searching task in anticipation of subsequent, larger shifts in the same direction, as in previous laboratory work [76]. How much this strategy is a conscious choice of the participant remains to be tested. Although participants might introspectively recognize that they “look around” more in the searching task, it seems unlikely that they are aware of precisely how much they adapted movements of the eyes versus head. Since visual exploration is over-learned—we continually shift gaze from moment to moment—observers may automatically adjust their exploration to suit the task. Developmental studies of visual exploration in infants and children may shed light on how exploratory control is acquired.

We acknowledge several limitations in our study that can be addressed in future research. First, we designed the study to use two different environments, each paired with a different task, to create unique demands on visual attention. Although this was helpful for using locations that fit with each activity (e.g., the walking path did not contain locations that would have been suitable for hiding targets), it also makes it more difficult to interpret what differences between the conditions were most important for changing visual exploration. In future work, we can compare walking with walking and searching in the same environments to better tease apart how the demands of the task and the visual features of the environment may have contributed to visual exploration. We also note that aggregating visual exploration across the entire walking task and entire searching task is an oversimplification. Although it was a useful way to broadly characterize how the spread of visual exploration differs across the two tasks, we are unable to address how moment-to-moment changes in actions and goals within each task (i.e., searching, retrieving, navigating during the search task) may have changed visual exploration over time. Finally, we acknowledge that the current studies cannot address the degree to which the selection of eye and head movements reflect conscious versus automatic processes.

In conclusion, the current studies show the importance of measuring both eyes and head to understand gaze behavior in complex, real-life tasks. Although differences were apparent in eye movements alone (Study 1), studying eye and head movements together uncovered that each effector contributed differently to visual exploration (Study 2). Adaptations to eyes-plus-head gaze were evident both in aggregate across the task as well as at the level of moment-to-moment gaze shifts, showing that the entire visual exploratory system was adapted to meet task demands. Our study shows the feasibility of using wearable, wireless eye and head tracking to characterize behavior "in the wild"; this method can be used profitably to investigate eye-head adaptation in a wider range of tasks across different environments. In doing so, we may better understand how visual exploration meets the various demands of daily life.

4.7 Acknowledgments

The authors would like to acknowledge that Chapter 4 of this dissertation was published in PLoS One in August of 2021.

Chapter 5

Conclusions

Across three studies this dissertation offers insight into the influence of top-down factors on screen-based (Chapter 2) and real-world (Chapters 3 and 4) visual attention. In Chapter 2 we explored the relationship between top-down information and infant free-viewing of dynamic stimuli, while Chapters 3 and 4 gave us insight into how top-down factors interact with visual and motor systems during real-world visual exploration (Chapter 3) and insight into how differences between tasks influence the coordination of elements of the visual and motor systems (Chapter 4). These findings extend both our understanding of the developmental trajectory of visual attention as well as the influence of environmental regularities and motor factors on adult attention and search efficiency in real-world visual exploration. Furthermore, this collection of work offers additional insight into the complexities of translating screen-based assessments of cognitive processes into the real-world.

The study described in Chapter 2 used a gaze-contingent paradigm to allow the infant sample to control their own stimuli presentation. In this manner infants were able to learn the association between their gaze and the subsequent visual stimuli. This rapidly acquired knowledge was found to later influence infant attention during their subsequent free-viewing of dynamic stimuli. Prior research has demonstrated that infants as young as 6-months are able to rapidly detect regularities in visual stimuli and use that information to guide their attention when viewing static images [101]. However, static stimuli do not contain the attention-grabbing characteristics of dynamic stimuli that are related to how visual features change over time (e.g., flicker, [73]). The results of Chapter 2 replicate and extend the findings of Tummeltshammer and Amso (2018), such that we now have conclusive evidence of 8-month-old infants employing top-down attentional selection in both static and dynamic displays. These results demonstrate an important milestone in the developmental progression of attentional selection. Future work may assess whether this development occurs earlier than 8-months as prior research has determined with screen-based stimuli [60, 101].

While this approach allowed us to assess the influence of top-down factors on infant attention, the variables of interest in the present study (attention to the static gaze-contingent and non-contingent objects) are only one of the possible forms of top-down information present in this study. Further inquiry into the present data set should assess the role of both bottom-up and other top-down factors in the moment to moment prioritization of visual information [50]. This will add to our existing understanding of how both bottom-up (e.g., feature congestion or flicker) and top-down (e.g., agentic characters or content) factors interact within infant directed media to influence infant attention [107].

Our findings from Chapter 3 suggest that past findings from screen-based assessments of visual attention may not seamlessly generalize to the real-world. That is to say that while findings from screen-based studies have indicated the influence of certain top-down factors such as the ability to detect environmental regularities [48] and prior knowledge [68] on search efficiency, the particulars of the present study combined in such a way that this influence was not detected in the present aggregate analyses. It is certainly possible that such an effect is present in the present data set and either 1) the aggregate approach of these analyses are unable to detect the effect, or 2) that the effect is present but the current sample is under powered to detect the effect.

However, it is also possible that these findings demonstrate that the relationship between these variables is actually not so straightforward in the real-world. That is to say that elements of real-world search (e.g., complex three dimensional visual information, the relative increase in motor effort, etc.) combine in such a way as to influence aspects of attention, object processing, or higher order cognition such as planning and executive functioning to the degree that hinders search efficiency. Furthermore, screen-based studies tend to use relatively simplistic stimuli that are likely faster and easier to process. This, combined with the relative ease of moving the eyes compared with moving larger body parts such as the head or trunk could all interact to decrease the influence of said top-down information on real-world search efficiency.

Finally, our findings from Chapter 4 demonstrate robust differences in how eye and head movements are adapted to meet task demands across locomotor tasks (walking vs. walking while searching). Furthermore, these findings extend prior research [98] which demonstrate that the contribution of the head to a gaze shift increases as a function of the amplitude of the gaze shift. Furthermore, we extended these findings by demonstrating that the head's contribution to gaze shifts is greater with the inclusion of additional task demands when searching compared to simply walking without explicit instructions.

Collectively, the findings from Chapters 3 and 4 demonstrate the influence of both task demands and the contributions of the perceptual system (eyes, head, and body) in real-world visual exploration. These studies demonstrate the viability of using real-world paradigms to assess even basic cognitive processes such as attentional selection. Furthermore, these studies reinforce the need for continued use of ecologically improved real-world paradigms. Finally, these findings also reinforce the need for continued assessment of even the most basic elements of research design including convenience sampling and the social aspects of the research environment (e.g., how even relatively basic adult behaviors may be influenced by the presence of observers).

The findings from each of these three studies enhance our understanding of how particular aspects of top-down information (regularities and task) influence attention and interact with motor factors. From these findings we have a better understanding of when top-down orienting begins to develop, and the aspects of real-world search that is influenced by top-down information. While prior research has indicated that top-down factors influence search efficiency in screen-based visual search tasks [48], this finding did not extend to this real-world search paradigm. However, consistent with screen-based research we did demonstrate the efficacy of another type of top-down information (task demands) on the coordination of eye and head movements.

The inconsistent pattern of results from Chapters 3 and 4 demonstrate how real-world behavior is dependent upon the coordination of multiple systems and interacting factors. The nature of viewing stimuli on a screen minimizes these complex interactions in such a way that screen-based findings cannot all generalize to real-world behaviors. The advancement of traditional screen-based tasks into real-world paradigms affords us the opportunity to manipulate and understand the interplay of visual and motor factors in the pursuit of understanding human behavior in real-world contexts.

Bibliography

[1] Alpher Acik, Adjmal Sarwary, Rafael Schultze-Kraft, Selim Onat, and Peter Konig. Developmental changes in natural viewing behavior: bottom-up and top-down differences between children, young adults and older adults. *Frontiers in Psychology*, 1(207):1–14, 2010.

[2] Scott Adler and Jazmine Orprecio. The eyes have it: visual pop-out in infants and adults. *Developmental Science*, 9(2):189–206, 2006.

[3] Nadja Althaus and Denis Mareschal. Using saliency maps to separate competing processed in infant visual cognition. *Child Development*, 83(4):1122–1128, 2012.

[4] Dima Amso, Sara Haas, and Julie Markant. An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PloS one*, 9(1):e85701, 2014.

[5] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

[6] Hermann Bulf, Scott P Johnson, and Eloisa Valenza. Visual statistical learning in the newborn infant. *Cognition*, 121(1):127–132, 2011.

[7] Richard L. Canfield and Marshall M. Haith. Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology*, 27(2):198–208, 1991.

[8] Marisa. Carrasco. Visual attention: The past 25 years. *Vision Research*, 51(13):1484–1525, 2011.

[9] Monica S. Castelhano, Michael L. Mack, and John M. Henderson. Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9:1–15, 2009.

[10] Kyle R. Cave and Jeremy M. Wolfe. Modeling the role of parallel processing in visual search. *Cognitive Psychology*, 22(2):225–27, 1990.

[11] Xin Chen and Gregory J Zelinsky. Real-world visual search is dominated by top-down guidance. *Vision research*, 46(24):4118–4133, 2006.

[12] Marvin M. Chun and Wolfe. Jeremy M. Visual attention. In E.B. Goldstein, editor, *Blackwell Handbook of Perception*, chapter 9, pages 272–310. Blackwell Publishing Ltd, Oxford, 2000.

[13] Leslie B. Cohen. Attention-getting and attention-holding processes of infant visual preferences. *Child Development*, 43(3):869–879, 1972.

[14] John Colombo. The development of visual attention in infancy. *Annual Review of Psychology*, 52:337–367, 2001.

[15] Floor de Groot, Falk Huettig, and Christian N.L. Olivers. When meaning matters: The temporal dynamics of semantic influences on visual attention. *Journal of Experimental Psychology*, 42(2):180–196, 2016.

[16] Scania de Schonen and Eric Mathivet. Hemispheric asymmetry in a face discrimination task in infants. *Child Development*, 61(4):1192–1205, 1990.

[17] Robert Desimone and John Duncan. Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1):193–222, 1995.

[18] Dejan Draschkow, Melvin Kallmayer, and Anna C Nobre. When natural behavior engages working memory. *Current Biology*, 31(4):869–874, 2021.

[19] Dejan Draschkow, Jeremy M Wolfe, and Melissa L-H Vo. Seek and you shall remember: Scene semantics interact with visual search to build better memories. *Journal of Vision*, 14(8):1–18, 2014.

[20] John Duncan and Glyn W. Humphreys. Visual search and stimulus similiarity. *Psychological review*, 96(3):433–458, 1989.

[21] W. Einhauser, U. Rutishauser, and C. Koch. Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, 8:1–19, 2008.

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

[23] Gennady Erlikhman, Sion Gutentag, Christopher D. Blair, and Gideon P. Caplovitz. Interactions of flicker and motion. *Vision Research*, 155:24–34, 2019.

[24] Jozsef Fiser and Richard N. Aslin. Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99(24):15822–15826, 2002.

[25] Tom Foulsham, Craig Chapman, Eleni Nasiopoulos, and Alan Kingstone. Top-down and bottom-up aspects of active search in a real-world environment. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie experimentale*, 68(1):8–19, 2014.

[26] Tom Foulsham, Esther Walker, and Alan Kingstone. The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, 51(17):1920–1931, 2011.

[27] Tom Foulsham, Esther Walker, and Alan Kingstone. The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, 51:1920–1931, 2011.

[28] J. M. Franchak. Looking with the eyes and head. In Jeffrey B Wagman and Julia J. C. Blau, editors, *Perception as Information Detection: Reflections on Gibson's Ecological Approach to Visual Perception*, pages 205–221. Routledge, 2020.

[29] John M. Franchak. Visual exploratory behavior and its development. In K. D Federmeier and E. R. Schotter, editors, *Psychology of Learning and Motivation*, pages 59–94. Elsevier, 2020.

[30] John M. Franchak and Karen E. Adolph. Visually guided navigation: head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research*, 50:2766–2774, 2010.

[31] John M. Franchak, David J. Heeger, Uri Hasson, and Karen E. Adolph. Free viewing gaze behavior in infants and adults. *Infancy*, 21(3):262–287, 2015.

[32] John M Franchak, Brianna McGee, and Gabrielle Blanch. Adapting the coordination of eyes and head to differences in task and environment during fully-mobile visual exploration. *PLoS one*, 16(8):e0256463, 2021.

[33] Michael C. Frank, Edward Vul, and Scott P. Johnson. Development of infants' attention to faces during the first year. *Cognition*, 110(2):160–170, 2009.

[34] Michael C. Frank, Edward Vul, and Rebecca Saxe. Measuring the development of social attention using free-viewing. *Infancy*, 17(4):355–375, 2012.

[35] Edward G Freedman. Coordination of the eyes and head during visual orienting. *Experimental Brain Research*, 190(4):369, 2008.

[36] J. J. Gibson. *The ecological approach to visual perception.* Houghton Mifflin Company, Boston, MA, 1979.

[37] James J Gibson. *The ecological approach to visual perception: classic edition.* Psychology Press, 2014.

[38] James Jerome Gibson and Leonard Carmichael. *The senses considered as perceptual systems*, volume 2. Houghton Mifflin Boston, 1966.

[39] D. Guitton and M. Volle. Gaze control in humans: Eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of Neurophysiology*, 58:427–459, 1987.

[40] Gregor Hardiess, Sabine Gillner, and Hanspeter A. Mallot. Head and eye movements and the role of memory limitations in a visual search paradigm. *Journal of Vision*, 8(1):7, 2008.

[41] Taylor R. Hayes and John M. Henderson. Center bias outperforms image salience but not semantics in accounting for attention during scene viewing. *Attention, Perception, Psychophysics*, 82:985–994, 2020.

[42] Mary M Hayhoe, Anurag Shrivastava, Ryan E B Mruczek, and Jeff B Pelz. Visual memory and motor planning in a natural task. *Journal of Vision*, 3:49–63, 2003.

[43] Andrea Helo, Sebastian Pannasch, Louah Sirri, and Pia Rama. The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research*, 103:83–91, 2014.

[44] John M Henderson, James R Brockmole, Monica S Castelhano, and Michael Mack. Visual saliency does not account for eye movements during visual search in real-world scenes. In R van Gompel, M Fischer, W Murray, and R Hill, editors, *Eye movements: A window on mind and brain*, pages 537–562. Elsevier, Oxford, 2007.

[45] Mark A. Hollands, Nausica V. Ziavra, and Adolfo M. Bronstein. A new paradigm to investigate the roles of head and eye movements in the coordination of whole-body movements. *Experimental Brain Research*, 154(2):261–266, 2004.

[46] Laurent Itti and Christof Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–1506, 2000.

[47] Laurent Itti and Christof Koch. Computational modeling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001.

[48] Yuhong V. Jiang, Khena M. Swallow, Gail M. Rosenbaum, and Chelsey Herzig. Rapid acquisition but slow extinction of an attentional bias in space. *Journal of Experimental Psychology*, 39(1):87–99, 2013.

[49] Mark H Johnson, Suzanne Dziurawiec, Hadyn Ellis, and John Morton. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1-2):1–19, 1991.

[50] Kellan Kadooka and John M. Franchak. Developmental changes in infants' and children's attention to faces and salient regions vary across and within video stimuli. *Developmental Psychology*, 56(11):2065–2079, 2020.

[51] Tommy P Keane, Nathan D Cahill, John A Tarduno, Robert A Jacobs, and Jeff B Pelz. Computer vision enhances mobile eye-tracking to expose expert cognition in natural-scene visual-search tasks. *Human Vision and Electronic Imaging XIX*, page 90140F, 2014.

[52] David J Kelly, Sofia Duarte, David Meary, Markus Bindermann, and Oliver Pascalis. Infants rapidly detect human faces in complex naturalistic visual scenes. *Developmental Science*, 22(6):e12829, 2019.

[53] Natasha Z. Kirkham, Jonathan A. Slemmer, and Scott P. Johnson. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, 83(2):B35–B42, 2002.

[54] Natasha Z. Kirkham, Jonathan A. Slemmer, C. Richardson, Daniel, and Scott P. Johnson. Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, 78(5):1559–1571, 2007.

[55] Dmitry Kit, Leor Katz, Brian Sullivan, Kat Snyder, Dana Ballard, and Mary Hayhoe. Eye movements, visual search and scene memory, in an immersive virtual environment. *PLoS ONE*, 9(4):e94362, apr 2014.

[56] Rakshit Kothari, Zhizhuo Yang, Christopher Kanan, Reynold Bailey, Jeff B. Pelz, and Gabriel J. Diaz. Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities. *Scientific Reports*, 10(1), feb 2020.

[57] K. S. Kretch and K. E. Adolph. Active vision in passive locomotion: Real-world free viewing in infants and adults. *Developmental Science*, 18:736–750, 2015.

[58] Kari S. Kretch and Karen E. Adolph. The organization of exploratory behaviors in infant locomotor planning. *Developmental science*, 20(4):e12421, 2017.

[59] Alexandra Kuznetsova, Per B Brockhoff, and Rune Haubo Bojesen Christensen. lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 2017.

[60] Mee-Kyoung Kwon, Mielle Setoodehnia, Jongsoo Baek, Steven J Luck, and Lisa M Oakes. The development of visual search in infancy: Attention to faces versus salience. *Developmental Psychology*, 52(4):537–555, 2016.

[61] M. F. Land. The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research*, 159:151–160, 2004.

[62] M. F. Land and M. M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision Research*, 41:3559–3565, 2001.

[63] Michael Land and Benjamin Tatler. *Looking and acting: Vision and eye movements in natural behaviour*. Oxford University Press, 2009.

[64] Michael F. Land. Eye movements and the control of actions in everyday life. *Progress in retinal and eye research*, 25:696–324, 2006.

[65] Michael F. Land. Vision, eye movements, and natural behavior. *Visual Neuroscience*, 26(1):51–62, 2009.

[66] Michael F. Land and Sophie Furneaux. The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1358):1231–1239, 1997.

[67] Michael F. Land, Neil Mennie, and Jennifer Rusted. The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11):1311–1328, 1999.

[68] Chia-Ling Li, M.P. Aivar, Dmitry M. Kit, Matthew H. Tong, and Mary M. Hayhoe. Memory and visual search in naturalistic 2d and 3d environments. *Journal of Vision*, 16(8):1–20, 2016.

[69] Lijia Liu and Dana H. Ballard. Humans use minimum cost movements in a whole body task. *Scientific Reports*, 11(1):1–15, 2021.

[70] George L. Malcolm and John M. Henderson. The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, 9(11):1–13, 2009.

[71] Jonathan S. Matthis, L. Yates, Jacob, and Mary M. Hayhoe. Gaze and the control of foot placement when walking in natural terrain. *Current Biology*, 28:1224–1233, 2018.

[72] M. Mills, A. Hollingworth, S. Van der Stigchel, L. Hoffman, and M. D. Dodd. Examining the influence of task set on eye movements and fixations. *Journal of Vision*, 11(8):17–17, 2011.

[73] Parag K. Mital, Tim J. Smith, Robin L. Hill, and John M. Henderson. Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, 3(1):5–24, 2011.

[74] John Morton and Mark H Johnson. Conspec and conlern: a two-process theory of infant face recognition. *Psychological review*, 98(2):164, 1991.

[75] Rebecca Nako, Rachel Wu, and Martin. Eimer. Rapid guidance of visual search by object categories. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1):50–60, 2014.

[76] B. S. Oomen, R. M. Smith, and J. S. Stahl. The influence of future gaze orientation upon eye-head coupling during saccades. *Experimental Brain Research*, 155:9–18, 2004.

[77] Oliver Pascalis, Scania de Schonen, John Morton, Christine Deruelle, and Marie Fabre-Grenet. Mother's Face Recognition by Neonates: A replication and an extension. *Infant Behavior and Development*, 18:79–85, 1995.

[78] Aftab E. Patla, Allan Adkin, Christine Martin, Rachel Holden, and Sandra Prentice. Characteristics of voluntary visual sampling of the environment for safe locomotion over different terrains. *Experimental Brain Research*, 112:513–522, 1996.

[79] Aftab E Patla and Joan N. Vickers. How far ahead do we look when required to step on specific locations in the travel path during locomotion? *NeuroReport*, 148:133–138, 2003.

[80] Jeff B. Pelz and Roxanne Canosa. Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, 41(25-26):3587–3596, 2001.

[81] Jeff B. Pelz, Mary Hayhoe, and Russ Loeber. The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139:266–277, 2001.

[82] Michael I Posner. Orienting of attention. *Quarterly journal of experimental psychology*, 32(1):3–25, 1980.

[83] Johnathan E. Prunty, Jolie R. Keemink, and David J. Kelly. Infants scan static and dynamic facial expression differently. *Infancy*, 26(6):831–856, 2021.

[84] Paul C. Quinn, Lesley Uttley, Lee Kang, Alan Gibson, Michael Smith, Alan M. Slater, and Oliver Pascalis. Infant preference for female faces occurs for same-but not other-race faces. *Journal of Neuropsychology*, 2(1):715–26, 2008.

[85] R Core Team. R: A language and environment for statistical computing; 2015, 2018.

[86] Greg D. Reyolds and Kelly C. Roth. The development of attentional biases for faces in infancy: A developmental systems perspective. *Frontiers in Psychology*, 9(222), 2018.

[87] Marcos Francisco Rosetti, Luis Pacheco-Cobos, Hernán Larralde, and Robyn Hudson. An experimental and theoretical model of children's search behavior in relation to target conspicuity and spatial distribution. *Physica A: Statistical Mechanics and its Applications*, 389(22):5163–5172, 2010.

[88] Risa Sawaki and Steven J. Luck. Capture versus suppression of attention by salient singletons: Electrophysiological evidence for an automatic attend-to-me signal. *Attention, Perception, Psychophysics*, 72(6):1455–1470, 2010.

[89] Tim J Smith and Parag K. Mital. Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, 13(8):1–24, 2013.

[90] J. S. Stahl. Amplitude of human head movements associated with horizontal saccades. *Experimental Brain Research*, 126(1):41–54, 1999.

[91] B. M. 't Hart and W. Einhauser. 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, 2012.

[92] Bernard M. 't Hart and Wolfgang Einhäuser. Mind the step: complementary effects of an implicit task on eye and head movements in real-life gaze allocation. *Experimental Brain Research*, 223(2):233–249, 2012.

[93] Michael K. Tanenhouse, Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217):1632–1634, 1995.

[94] B. W. Tatler, Y. Hirose, S. K. Finnegan, R. Pievilainen, C. Kirtley, and A. Kennedy. Priorities for selection and representation in natural tasks. *Philosophical Transactions of the Royal Society B*, 368:20130066, 2013.

[95] Benjamin W. Tatler, Roland J. Baddeley, and Iain D. Gilchrist. Visual correlates of fixation selection: effects of scale and time. *Vision Research*, 45:643–659, 2005.

[96] Benjamin W Tatler, Mary M Hayhoe, Michael F. Land, and Dana H. Ballard. Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11(5):1–23, 2011.

[97] Nicholas DA Thomas, James D Gardiner, Robin H Crompton, and Rebecca Lawson. Look out: an exploratory study assessing how gaze (eye angle and head angle) and gait speed are influenced by surface complexity. *PeerJ*, 8:e8838, 2020.

[98] Matteo Tomasi, Shrinivas Pundlik, Alex R. Bowers, Eli Peli, and Gang Luo. Mobile gaze tracking system for outdoor walking behavioral studies. *Journal of Vision*, 16(3):1–15, 2016.

[99] Matteo Tomasi, Shrinivas Pundlik, Alex R Bowers, Eli Peli, and Gang Luo. Mobile gaze tracking system for outdoor walking behavioral studies. *Journal of vision*, 16(3):27–27, 2016.

[100] Anne M Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.

[101] Kirsten S. Tummeltshammer and Dima Amso. Top-down contextual knowledge guides visual attention in infancy. *Developmental Science*, 21(4):1–9, 2018.

[102] Kirsten S. Tummeltshammer and Natasha Z. Kirkham. Learning to look: Probabilistic variation and noise guide infants' eye movements. *Developmental Science*, 16(5):760–771, 2013.

[103] Kathleen A Turano, Duane R Geruschat, and Frank H Baker. Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research*, 43:333–346, 2003.

[104] Fernando Vargas-Martin and Eli Peli. Eye movements of patients with tunnel vision while walking. *Investigative Opthalmology & Visual Science*, 47(12):5295, 2006.

[105] Melissa L.H. Vo and Jeremy M. Wolfe. When does repeated search in scenes involve memory? looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1):23–41, 2012.

[106] Quan Wang, Jantina Bolhuis, Constantin A Rothkopf, Thorsten Kolling, Monika Knopf, and Jochen Triesch. Infants in control: rapid anticipation of action outcomes in a gaze-contingent paradigm. *PloS one*, 7(2):e30884, 2012.

[107] Smith Wass and Tim J. Smith. Visual Motherese? Signal-to-noise ratios in toddler-directed television. *Developmental Science*, 18(1):24–37, 2015.

[108] Robert L. Waters and Sara Mulroy. The energy expenditure of normal and pathological gait. *Gait and Posture*, 9:207–231, 1999.

[109] Jeremy M. Wolfe. Visual search in continuous, naturalistic stimuli. *Vision Research*, 34(9):1187–2295, 1994.

[110] Jeremy M Wolfe. Guidance of visual search by preattentive information. *Neurobiology of attention*, pages 101–104, 2005.

[111] Bo-Yeong Won, Hyjin J. Lee, and Yuhong V. Jiang. Statistical learning modulates the direction of the first head movement in a large-scale search task. *Attention, Perception, Psychophysics*, 77(7):2229–2239, 2015.

[112] Gregory J. Zelinsky, Wei Zhang, Bing Yu, Xin Chen, and Dimitris Samaras. The role of top-down and bottom-up processes in guiding eye movements during visual search. *Advances in neural information processing systems*, pages 1569–1576, 2006.

[113] Artyom Zinchenko, Markus Conci, Hermann J Müller, and Thomas Geyer. Predictive visual search: Role of environmental regularities in the learning of context cues. *Attention, Perception, & Psychophysics*, 80(5):1096–1109, 2018.