Looking with the head and eyes

John M. Franchak

University of California, Riverside

franchak@ucr.edu

Accepted chapter to appear in *Perception as Information Detection: Reflections on Gibson's Ecological Approach to Visual Perception,* edited by Jeffrey Wagman and Julia Blau

Looking with the head and eyes

Although *The Ecological Approach to Visual Perception* (Gibson, 1979) is perhaps best known for the theories of affordances and direct perception, Gibson's theory of visual exploration with the eye-head-body system was no less revolutionary. The evolution of his thought on visual exploratory behavior can be traced through his major works. In *Perception of the Visual World,* Gibson (1950) considered the consequences of head and body movements on visual experience, and emphasized the role of the body in stabilizing vision. In *The Senses Considered as Perceptual Systems,* Gibson (1966) defined the visual perception system as extending beyond the eyes to involve active exploration with the eyes, head, and body. The culmination of these ideas appears in *The Ecological Approach to Visual Perception*: Gibson argued that the purpose of a whole-body visual system is for exploring and interacting with one's surroundings. Gibson dissociated what people are asked to do in typical laboratory studies, what he termed "looking at a page or a picture", from what people actually do in real life, that is, "looking around":

> The perceptual capacities of the organism do not lie in discrete anatomical parts of the body but lie in systems with nested functions. Even so, it might be argued, one surely *looks* with the eyes even if one does not *see* with the eyes. But looking with the eyes alone is mere looking *at*, not looking *around*. It is the scanning of an object, a page of print, or a picture. One also looks with the head, not just the eyes, more exactly with the head-eye system… (p. 205, emphasis in original)

This characterization of visual exploration was at odds with the dominant paradigm in vision science, in which the eyes scan restricted, two-dimensional stimuli while the head and body remain still. Today, *screen-based tasks* that present shapes, photographs, or movies to passive observers are still far more common compared to *natural tasks* in which active observers explore the environment with unfettered body movement. Gibson dismissed studies of “looking at” for three reasons, each of which has been influential in guiding contemporary research on naturalistic looking with the eyes and head.

First, Gibson claimed that studies using such simplistic, two-dimensional displays lead researchers to overweight the influence of stimulus features on gaze behavior: “It is certainly a fallacy to assume that a saccadic movement is a response to a ‘stimulus’ on the periphery of the retina that brings it to the fovea.” (p. 212). Gibson rejected the notion of stimulus-driven, involuntary eye movements in favor of a functional approach in which eye movements support the perception of ecologically-relevant aspects of the environment and facilitate action performance. Indeed, contemporary work using both screen-based and natural tasks have consistently found that observer’s tasks and goals are more influential in determining where people look compared to the appearance of visual targets (Tatler, Hayhoe, Land, & Ballard, 2011).

Second, Gibson claimed that “the awareness of the observer’s own body in the world is a part of the experience” (p. 207) of perceiving the visual world. An observer’s experience of the environment is *active*: Visual exploration supports interacting with the environment and, in turn, the observer’s movements affect what is seen. In contrast, an observer’s experience of looking at photographs or screen-based displays is *mediated* (Stoffregen, Chapter 15; Blau, Chapter 16). When visually exploring an image, the observer directly perceives the representational medium (the image on a screen) but only indirectly perceives what is represented by the image. The observer is aware of being seated looking at an image, not of being present in the visual world that depicted. Moving the head does not change the observer’s perspective within the photograph, but merely changes where the photograph is within the field of view.

Third, Gibson claimed that studies of “looking at” give undue prominence to eye movements, which are only one part of an integrated perceptual system: “One sees the environment not with the eyes but with the eyes-in-the-head-on-the-body-resting-on-the-ground” (p. 205). Studying the nested movements that permit looking around is not possible with a seated observer who looks at a computer screen. Naturalistic studies of adults’ visual exploration show that eye, head, and body movements are coordinated to pick up information in the environment (Land, 2004). Moreover, infant studies show that developmental changes in the motor skills underlying visual exploration fundamentally alter what infants see (Franchak, Kretch, & Adolph, 2018; Kretch, Franchak, & Adolph, 2014).

In this chapter, I review empirical work that has expanded on these three main points. Each section describes work with adults as well as infant and child research to highlight how differences in the developing eye-head-body system affect visual exploration. I begin with a discussion of the methodological innovations that were necessary to study naturalistic visual exploration.

Measuring visual exploration

Whereas research in affordances followed shortly after the 1979 publication of *The Ecological Approach to Visual Perception* (e.g., Warren, 1984), the first studies of naturalistic visual exploration did not appear until the late 1990s (Land & Furneaux, 1997; Land, Mennie, & Rusted, 1999). One source of delay was the need for technology for studying looking around rather than mere looking at—mobile eye tracking. Why does visual exploration require specialized technology to study when other ecological aspects of visual perception, such as affordances, do not? The relevant behaviors for studying affordance perception are gross motor movements—reaching for objects, stepping on risers, leaping over gaps—that can be scored by human observers or recorded with a basic video camera. In contrast, accurately scoring where a participant is looking varies from extremely difficult to utterly impossible depending on the complexity of the environment and the required precision of the measurement when relying on a basic 3rd person video recording. A few observational studies have successfully scored gaze behavior in naturalistic (but constrained) tasks from a 3rd-person perspective, such as when adults glance down while descending stairs (Rosenbaum, 2010).

Because the eyes are small and move extremely quickly—with speeds as great as 700°/s (Land, 2006)—high-speed eye tracking devices with specialized cameras are required to record their movements, and thus to determine gaze direction. Various eye tracking systems have been devised dating back to the start of the 20th century; however, most systems—which I will refer to as *stationary eye trackers*—require a seated observer, a fixed display, and (often) immobilization of the head (for a review, see Tatler & Land, 2015). Gibson noted the contribution of high speed stationary eye trackers in discovering microsaccades, extremely quick movements of the eye during fixations. However, Gibson recognized that stationary eye movement recording was insufficient to study whole-body visual exploration. Although modern stationary systems can record eye movements with incredible spatial accuracy (0.02°-0.5°) and temporal resolution (500-2000 Hz), the observer is required to sit in place and eye movements can only be measured within the boundaries of a computer display.

Studying whole-body visual exploration requires *mobile eye trackers* that record eye movements in an observer who is free to move the head and body. The first mobile eye trackers were developed in the 1940s-1960s but did not allow for the study of natural behavior because the large devices occluded parts of observer’s view and were too cumbersome to allow natural movement (Tatler & Land, 2015). In the early 1990s, Land (1992) developed a portable, head-mounted eye tracker that simultaneously recorded videos of the observer's eye and first-person field of view (Figure 1). Eye position is calibrated with gaze locations within the first-person

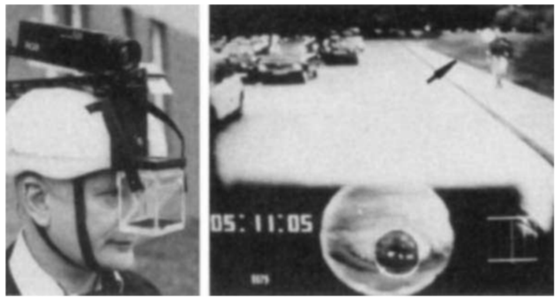


Figure 1. Left: Mobile eye tracker devised by Land (1992). Right: Simultaneous eye and field of view recording for tracking eye gaze.

(i.e., head-centered) field of view video. The field of view video records the changes in viewpoint that result from shifts of head and body position; thus, mobile eye trackers can truly measure looking around in natural tasks. Land and his colleagues used this new technique to study visual exploration during driving (Land, 1992; Land & Lee, 1994), playing sports (Land & Furneaux, 1997), and while completing everyday activities such as making a cup of tea (Land et al., 1999).

Technological advancements have led to several improvements in mobile eye tracking. Whereas early mobile eye tracking required manual coding of eye position for every video frame, later systems automate this process using computer vision algorithms to detect the eye (D. Li, Babcock, & Parkhurst, 2006). Although the spatial (.5°-2°) and temporal (30-200 Hz) resolution of mobile eye trackers is modest compared to stationary eye trackers, it is sufficient for most natural tasks. Miniaturization of video and electronics components continues to make eye tracking devices more portable and less obtrusive to conduct studies "in the wild" as opposed to in the lab (Figure 2A). Moreover, the development of lightweight mobile eye trackers (Figure

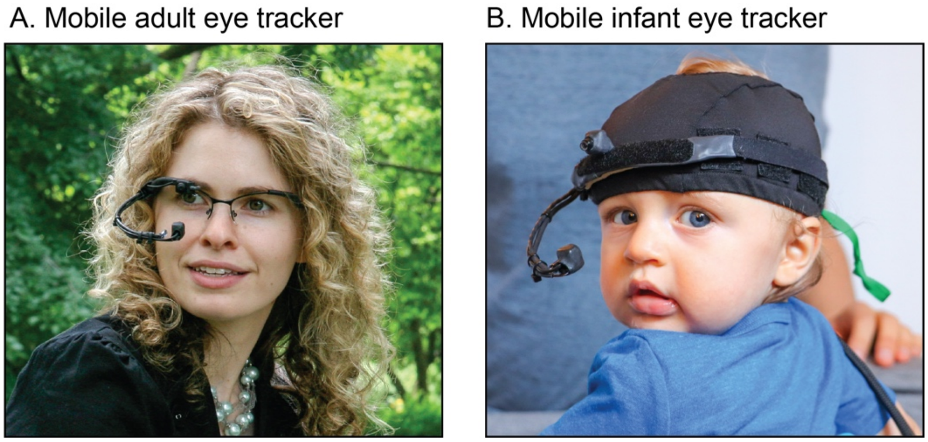


Figure 2. (A) Lightweight, mobile adult eye tracker (photo courtesy of Jason Babcock). (B) Mobile infant eye tracker.

2B) allowed for the first recording of real-world visual exploration in freely-moving infants (Franchak, Kretch, Soska, & Adolph, 2011; Franchak, Kretch, Soska, Babcock, & Adolph, 2010) and children (Franchak & Adolph, 2010).

Integration of mobile eye tracking with other technologies has opened new avenues of research. Simultaneous eye and motion tracking allows researchers to precisely measure how eye, head, hand, and body movements are coordinated in everyday tasks. For example, Franchak and Yu (2015) studied infants' and caregivers' naturalistic play to measure eye and head alignment to toys while reaching. Adults adjusted the speed of their reaches depending visual alignment—when they aligned their eyes directly on toys they reached more quickly but slowed their reaches when toys were in the periphery. In contrast, infants did not systematically coordinate eye movements with reaching speed. Matthis and colleagues (Matthis, Yates, & Hayhoe, 2018) combined mobile eye tracking with full-body inertial sensing to examine visual guidance of locomotion while hiking outdoors and found that walkers altered their gaze allocation to account for the different task demands associated with walking over different terrains (e.g., rocky trail vs. uniform dirt path).

Scoring where observers look is challenging for mobile eye tracking research compared with stationary eye tracking research. With stationary eye trackers, the locations of gaze targets in are the same for every participant because a computer display presents them uniformly and consistently. For example, calculating how often participants look at a face in an image means defining which pixels comprise the face, and then calculating how frequently participants' gaze coordinates fall on those pixels. In contrast, in mobile eye tracking the locations of gaze targets in the field of view are unique to each participant depending on how each participant chose to orient the field of view by moving the body and head (Franchak, 2017). For example, scoring how often infants look at caregivers' faces in a naturalistic interaction means manually scoring where faces are in each infant's field of view at every moment—for each infant, the caregiver's face will be in view in different locations at different times. The time-intensive, frame-by-frame scoring means that sample sizes in mobile eye tracking studies are typically smaller than in comparable stationary eye tracking studies and might also contribute to why the size of the mobile eye tracking literature is still modest.

One promising solution is automated computer vision detection of targets of interest. For example, researchers have developed systems to detect where faces (Frank, Simmons, Yurovsky, & Pusiol, 2013), hands (Bambach, Franchak, Crandall, & Yu, 2014), and objects (Yu & Smith, 2013) are located in the field of view for each participant to automatically score looking behavior. However, accuracy varies and is typically worse compared with human coding because detection depends on visual clutter in the scene and the discriminability of targets. Another solution is to use eye trackers embedded in virtual reality headgears; manual coding of looking is unnecessary because the location of gaze targets must be known to render the virtual environment (Diaz, Cooper, Kit, & Hayhoe, 2013). Yet another promising method is linking the mobile eye tracking reference frame to that of a motion capture system to automatically register eye gaze with respect to body movements (Matthis et al., 2018).

Task-driven visual exploration

Gibson dismissed the idea that eye movements are directed towards targets based on stimulus features. Instead, Gibson stressed the importance of visual exploration for supporting the observer's behavioral agenda. Moving the eyes and head to look around is an exploratory behavior that should depend on what information is relevant to the observer's task. Indeed, early studies by Buswell and Yarbus demonstrated that different instructions when looking at artwork influenced how observers directed their eye movements in both space and time (Buswell, 1935; Yarbus, 1967).

Despite early work highlighting the role of observers' tasks and goals, stimulus-driven theories of visual attention have been prominent in modeling eye movements in screen-based tasks (Borji & Itti, 2013; Itti, Koch, & Niebur, 1998; Itti & Koch, 2001). So-called saliency models predict that observers make eye movements to targets that "pop out" from the surrounding scene based on their appearance—brightness, color, and movement. One reason that stimulus-driven approaches have continued to thrive is that many screen-based studies fail to provide the observer with a clearly-defined task. Typically, observers are instructed to simply view a photograph for a few seconds or watch a video. Assigning a task, such identifying the location where a photograph was taken, reduces the influence of stimulus appearance on eye movements (Smith & Mital, 2013).

Task influences on visual exploration are even more striking when observers are free to engage in everyday actions that involve movement. For example, when making a cup of tea (Land et al., 1999) or a peanut butter and jelly sandwich (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003), most eye movements are directed toward task-relevant objects (e.g., the tea kettle or the jelly jar). Irrelevant objects and the “background” of a scene are rarely looked at even when salient. Even when visually-salient, irrelevant objects are inserted in the background they are rarely fixated (Rothkopf, Ballard, & Hayhoe, 2007). Instead, varying instructions about which tasks to treat as obstacles and which to approach systematically altered looking behavior. Even in a real-world search task—where target “pop out” should be more likely to drive eye movements—target appearance had little influence unless observers were explicitly given instructions to expect that the search target would look different compared with distractors (Foulsham, Chapman, Nasiopoulos, & Kingstone, 2014).

The pickup of visual information about task-relevant targets shows not just spatial but also temporal coordination with actions. The eyes shift to objects just before the hand reaches, such as looking at a knife before reaching towards it, or the eyes may monitor ongoing events, such as when watching a tea kettle as it is being filled with water (Hayhoe et al., 2003; Land et al., 1999). Less frequently, observers may break from looking at the current, in-use object to “look ahead” at an object that will be relevant a few seconds in future, such as glancing at the soap dispenser while approaching the sink (Pelz & Canosa, 2001).

Task-specificity and timeliness of visual exploration is also evident in infants’ naturalistic behavior. While infants and caregivers played together with a set of toys, infants most often look at toys they are playing to with to control manual actions and spend less time looking at caregivers (Franchak et al., 2018; Yu & Smith, 2013). Infants coordinate the timing of eye and hand movements during natural reaching: Infants keep their eyes fixed on toys as they guide reaching movements and look away either just before or just after the hand makes contact (Franchak et al., 2011). From 9-24 months of age, infants increasingly orient their heads to center toys in view, possibly because looking with aligned eyes and head facilitates visual guidance of reaching and object manipulation (Franchak, Smith, & Yu, under review).

Awareness of the self and others

The second problem in laboratory tasks that Gibson recognized is the mediated experience of looking at a scene in a photograph (in modern tasks, a computer screen). Even when looking at a "natural image", such as a photograph of an outdoor scene, the observer's experience is not of being present in that scene but of being seated in a chair in a laboratory room looking at a photographic representation. Gibson stressed that the visual consequences of observer movement are important for self-perception. However, when viewing a photograph in a laboratory study, moving the head changes the photograph's location in the observer's visual field rather than shifting the observer's viewpoint within the photograph. Additionally, when perception is mediated, the observer cannot act on the environment but can only passively view the represented scene.

Indeed, agency—having the ability to act on the environment—has consequences on how people visually explore. Foulsham and colleagues (2011) tested this directly by comparing *active observers* and *passive observers.* Active observers' gaze was recorded with a mobile eye tracker while they walked across campus to get a cup of coffee. Passive observers sat in the lab and watched a screen that displayed scenes recorded from the active observers' point of view. Whereas active observers had agency—they could choose how to move their bodies and heads to interact with and explore their surroundings—passive observers could only watch the scenes presented to them. As a consequence, active and passive observers' eye movements differed: Active observers clustered their gaze tightly within the head-centered field of view, presumably because they were able to use their heads to center targets of interest. In contrast, passive observers (who could only explore within the scene by moving their eyes) distributed gaze more broadly around the image. Active observers more often looked at or slightly below the horizon and towards the path as they walked, but passive observers who did not need to guide locomotion looked less often at areas relevant for locomotion. In a similar comparison between active and passive observers that used a tea-making task, both groups of observers looked more at task-relevant objects compared to task-irrelevant objects (Tatler et al., 2013). However, active observers looked even more at task-relevant objects compared to passive observers because they needed to manipulate those objects to make the cup of tea. A second condition that removed the active observer's need to manipulate objects (walking through the room to memorize the objects) eliminated the difference; both active and passive observers looked equally at task-relevant objects.

Social looking, that is, looking at the hands, faces, and bodies of people, differs between natural and screen-based tasks because looking at people who can engage in social interaction versus images of people who cannot interact with the observer are fundamentally different experiences. When watching videos of people, adult observers frequently look at faces, specifically at the eyes to monitor others' gaze (Birmingham, Bischof, & Kingstone, 2008). However, looking at a video recording of a stranger is different than looking at a stranger in real life because the real-life person can recognize eye contact as a social gesture that signals a desire to interact. Consequently, observers are less likely to look at a real person in a laboratory room compared to a video of another person displayed on a computer screen (Laidlaw, Foulsham, Kuhn, & Kingstone, 2011). In the study comparing active and passive observers walking to get a cup of coffee, passive observers were more likely than active observers to look at nearby people in the scene (Foulsham et al., 2011).

Similar parallels are seen when comparing screen-based versus natural tasks of infant social looking. Infants frequently gaze at faces in screen-based tasks, with 12-month-old infants spending as much as 50% of the time looking at faces while watching videos (Franchak, Heeger, Hasson, & Adolph, 2016; Frank, Vul, & Johnson, 2009). However, in real-life locomotor play infants spend only 4% of the time looking at caregivers' faces and 15% of the time looking at caregivers' hands and bodies but spend 38% of the time looking at toys (Franchak et al., 2018). Even when caregivers spoke to infants, infants responded only 8% of the time by looking at the caregiver's face (Franchak et al., 2011). When infants and caregivers sat across from one another at a table, infants spent 12% looking at caregivers' faces but 60% of the time looking at toys (Yu & Smith, 2013).

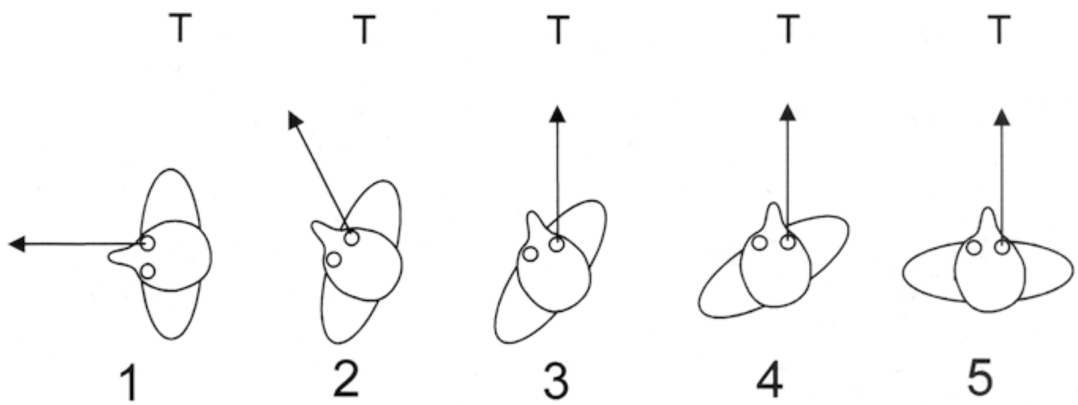
Low rates of face-looking are likely due to two factors. First, adult's faces are often outside infants' field of view, especially when infants tilt their heads down to look at toys (Franchak et al., under review), unlike in screen-based tasks in which faces are placed in easily viewed locations in a display. Indeed, when playing on the floor, infants look more often at sitting caregivers' faces compared with standing caregivers' faces (Franchak et al., 2018; Franchak et al., 2011). Second, infants' preoccupation with guiding locomotion and playing with toys means that faces are competing with other task-relevant locations (toys, furniture, distant locations). No such competition exists in screen-based tasks when infants passively view photographs or videos. A real-world study of infants being carried through a hallway created a passive viewing situation more similar to a screen-based tasks, and found high rates of infants looking at people and faces (significantly more than caregivers whose eye movements were also recorded) (Kretch & Adolph, 2015).

A nested perceptual system for looking around

Gibson emphasized that the nested visual system—the eyes within the head within the body—is coordinated to visually explore one's surroundings. Eye, head, and body movements serve two complementary functions in visual exploration: shifting gaze from one location in the world to another and stabilizing gaze on a desired target while compensating for movement of the target, the observer, or both.

Saccades and head movements are tightly coordinated to shift gaze from one location in the environment to another. Saccades are rapid movements of the eyes to shift the fovea—the location of greatest visual acuity—from one place to another within the head-centered field of view. The eyes can rotate approximately 50° along the horizontal axis within the head to look at extremely eccentric targets (Land, 2006). However, such extreme eye rotations are seldom observed in natural tasks. Both infants and adults keep the eyes relatively centered in their orbits to look in the direction that the head is pointed (Bambach et al., 2014; Bambach, Smith, Crandall, & Yu, 2016; Einhauser et al., 2007; Foulsham et al., 2011). That observers rarely make extreme eye movements indicates that much of visual exploration is accomplished with head movements. Indeed, laboratory studies find that head movements accompany eye movements to prevent extreme rotations of the eye, even when the eye can fixate the target without moving the head (Guitton & Volle, 1987; Oomen, Smith, & Stahl, 2004). One reason that observers may choose to minimize eye rotation is that it is efficient for visual exploration (Tatler, 2007). Maximum rotation of the eyes in a direction precludes them from rotating further in that direction, which could be detrimental when tracking targets (especially ones that move unpredictably). Continual, compensatory head movements that re-center the eyes means that the eyes are free to quickly move in any direction at a moment's notice.

Eye and head movements are also coordinated to stabilize gaze. The vestibular-ocular reflex (VOR) is a rapid counter-rotation of the eyes that compensates for rotations of the head. VOR happens concurrently with head movements; feed-forward information about future head movement informs on when and how to counter-rotate the eye (Land, 2004). Gibson noted the importance of VOR, "It compensates for the turning of the head. Thus, it is a *nonmotion* of the eyes relative to the *environment*, a posture, like fixation." (p. 210, emphasis in original). Just as body postures like sitting and standing involve compensatory movements to keep the body stable on a support surface, compensatory eye movements keep gaze stable with respect to the environment and keep the observer from experiencing the visual world "swinging" around as the



head and body move. Gibson recognized that the term "reflex" is inaccurate because compensatory eye movements are in fact voluntary. VOR is automatic but it can be suspended, such as when the eyes and head move in synchrony to make a large (> 50°) shift of gaze (Guitton & Volle, 1987).

Studies of whole-body visual exploration show that eye, head, and body movements are tightly coordinated to achieve both gaze shifts and stabilization from moment to moment, but coordination patterns are flexibly assembled to meet the demands of various tasks. Land (2004) measured the timing between eye, head, and trunk rotations during the tea-making task. Figure 3 shows a typical 90° shift of gaze from one location in the world to another. The eyes, head, and trunk all begin to move together in step 1. At step 2, the eyes have stopped moving relative to the head while the head and trunk continue to rotate. At step 3, the eyes have reached the desired target, so they counter-rotate to compensate for continued rotations of the trunk and head. At step 4, the head has reached its desired position, so it must counter-rotate along with the eyes to

Figure 3. Different stages of eye, head, and body rotations during a 90° shift of gaze from one location in the world to the target "T" (Land, 2004).

stabilize gaze on the target while the trunk finishes its rotation at step 5. Other studies have investigated coordination between eyes, head, and hands in different tasks. In a block-copying task that required looking back and forth between a model and a workspace, Pelz and colleagues (Pelz, Hayhoe, & Loeber, 2001) discovered task-dependent synergies between the timing of eye, head, and hand movements as participants repetitively checked the model while manipulating objects in the workspace. Unlike laboratory investigations of eye-head coupling that only found head movements accompanying large shifts of gaze (20°), in the block-copying task head rotation varied widely during gaze shifts (from 1-10°) meaning that eye-in-head position also varied. Other work using a similar task found that head movements were sometimes dissociated from gaze direction to orient towards future reaching targets that the eyes had not yet fixated (Smeets, Hayhoe, & Ballard, 1996).

Moving the head is important for minimizing extreme eye movements, however, head movements themselves should also minimized to make visual exploration efficient because head movements are effortful. Observers must choose whether to move the head to align the eyes towards a target or whether to view a target in the periphery without moving the eyes and head. During naturalistic locomotion, observers avoid unnecessary head movements because moving the head to look down at the ground surface is energetically costly and shifts the visual field away from the destination ahead. When walking on a flat, uniform surface, observers look down less often compared to when traversing a more difficult, uneven surface (’t Hart & Einhauser, 2012). When observers do look down towards the terrain, they do so in a way to maximize the efficiency of gait planning (Matthis et al., 2018). The difficulty of the terrain informs visual-motor planning; walkers can manage to look longer ahead (4 steps in advance) when traversing easy terrain, but on difficult terrain observers look closer to the moment of stepping (2 steps ahead). Watching the current step or looking only 1 step ahead is seldom done because corrections at that stage disrupt the gait cycle. Comparing 4- to 8-year-old children to adults in an obstacle navigation task revealed that children timed fixations to obstacles in a similar way as adults—both groups fixated obstacles 2-3 steps ahead (Franchak & Adolph, 2010). This suggests that children, like adults, avoid looking down at their feet while traversing obstacles in favor of looking ahead. Even 14-month-old infants—while walking over bridges spanning a precipice—only glanced at bridges at the start of each trial and seldom looked down at the bridge while traversing it (Kretch & Adolph, 2017).

Decisions about whether to look at targets in the environment at may depend on the interaction between motor effort and informational costs of looking. In laboratory investigations of obstacle navigation, adult participants fixated obstacles in advance 78% of the time while repeatedly walking up to and over a single obstacle trial after trial (Patla & Vickers, 1997). However, when obstacle navigation was embedded in another task—searching for objects around a room in a scavenger hunt—adults fixated obstacles in advance only 32% of the time (Franchak & Adolph, 2010). Turning the head down to look at obstacles would interfere with searching around the room, so participants chose instead to rely on peripheral information about obstacles. Children fixated obstacles significantly more often (59% of the time), possibly because children's shorter stature meant that obstacles on the ground were easier to view without tilting the head down. Figure 4A-B shows the visual fields for a typical 8-year-old girl compared to a

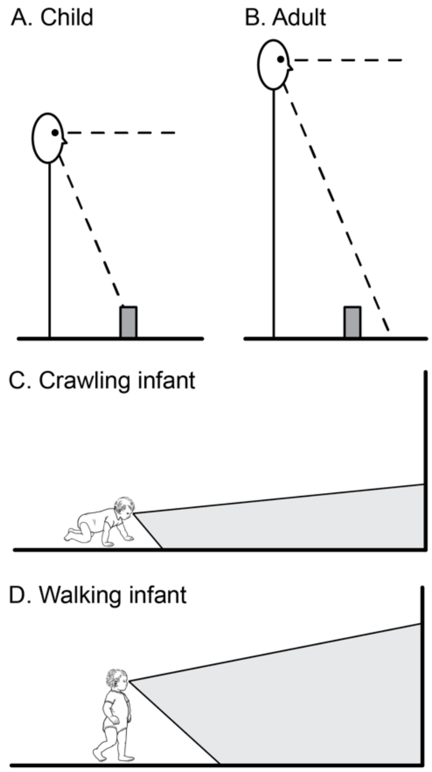


Figure 4. Field of view for an average-height 8-year-old girl (A) compared with an average-height 20-year-old woman (B), adapted from Franchak and Adolph (2010). The obstacle (grey rectangle) is in the field of view for the shorter child but not for the taller adult. Field of view measurements for a crawling infant (C) compared to a walking infant (D) show that crawlers see more of the ground and less of the world ahead compared with walkers, adapted from Kretch and colleagues (2014).

20-year-old woman. Because of the difference in height, the lower bound of the field of view is closer to the feet for the child compared to the adult, meaning that the child has greater opportunity to view obstacles in her path.

Just as differences in height affect visual access to the environment, different body postures affect how infants see the world. Kretch and colleagues (Kretch et al., 2014) measured the visual fields of infants while crawling versus walking (Figure 3C-D). The differences in viewpoint are striking: Walking infants can see more of the world ahead, whereas crawlers see more of the floor. A quarter of the time crawling infants see only the floor and nothing of the world ahead. These postural effects on visual access to the world have functional consequences. When navigating obstacles, 14-month-old infants fixated 72% of obstacles while walking compared with 90% of obstacles while crawling (Franchak et al., 2011), presumably because obstacles are more likely to be in view while crawling than walking. When caregivers showed infants an attractive toy, walking infants were more likely than crawling infants to look at the toy, suggesting that crawling infants were unable to compensate for their restricted view (Kretch et al., 2014).

Social looking also depends on infants' body posture. Infants are less likely to look at caregivers' faces while crawling compared to while sitting or standing upright (Franchak et al., 2018). The effect is specific to looking at caregivers' faces; looks to caregivers' bodies did not differ by posture because bodies were accessible even in crawlers' view. Changes in infants' posture over development might alter what infants see in everyday life. In-home studies using wearable head cameras find that the presence of faces in infants' field of view declines over the first two years of life (Jayaraman, Fausey, & Smith, 2015). Indeed, young infants spend nearly 50% of the time held up off the ground (Franchak, under review), which provides a good vantage point for face-looking (Kretch & Adolph, 2015). As infants learn to sit, crawl, and walk, they spend more time down on the floor in positions that are less conducive to face looking.

Differences between postures place constraints on what infants can see, but within each posture infants can choose how to orient their heads to determine what is in view. When sitting across from caregivers and playing with toys, 12-month-old infants try to keep both toys and faces in view simultaneously, which results in neither type of target being well-centered in view (Franchak et al., under review). By 24 months, infants center toys in view at the expense of caregivers' faces. This is true of face and toy locations both while those targets are fixated as well as when they are not; Figure 5 shows the differences in the locations of faces and toys in

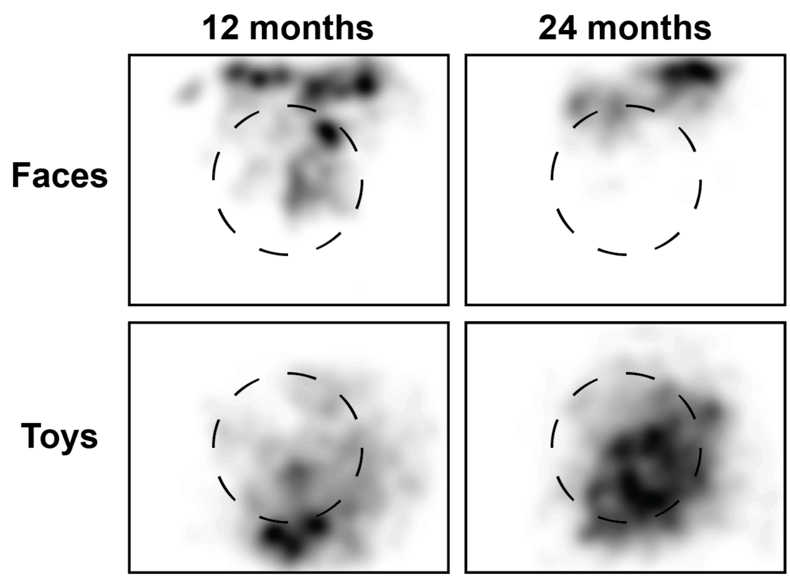


Figure 5. Heat maps showing the frequency of face (top row) and toy (bottom row) locations within the head-centered field of view for 12- and 24-month-old infants during fixations. White indicates low frequency while increasingly dark shades of gray indicate higher frequency. Dashed black circles indicate a 15° radius around the center of view. For 12-month-olds, neither toys nor faces were often within the middle of the field of view. For 24-month-olds, toys were frequently in the center of view but faces were not.

head-centered field of view of 12- and 24-month-old infants during fixations. Biasing toys in view might help infants control manual actions with toys and deal with unpredictable movement of toys in the field of view. These effects may be task dependent. Whereas 12-month-olds did not bias their view to favor toys over faces in the seated play task, they did bias toys over faces in a locomotor play task (Luo & Franchak, in preparation). Viewpoint bias may depend on the relative locations of different types of targets in the environment and the observer's task.

Looking ahead

To summarize, Gibson argued that studies of looking at photographs in the laboratory cannot inform on the real-life behavior of looking around the world. The advent of mobile eye tracking allowed researchers to collect empirical data that support this claim. First, naturalistic studies of eye tracking across a variety of tasks show that people use the eyes and head to select visual targets that support their goals and actions within a task; physical appearance of stimuli bears little on where people look. Second, observers' agency—being able to actively engages with the environment and other social agents—leads to different decisions about where to look in natural tasks compared with mediated studies of looking at images by passive observers. Third, the coordination of eyes, head, and body to select where to look is flexible, task-specific, and a ubiquitous part of natural behavior that cannot be studied in screen-based tasks. What adults see depends on how they choose to orient the nested visual system; what infants see is constrained by how they are able to orient the nested visual system. Efficient visual exploration depends on deciding how to coordinate visual exploration amidst the informational and motor demands of other ongoing tasks. It is important to note that the studies cited in each of the above sections could easily have been applied support any of these three points; nearly every study of naturalistic visual exploration shows the importance of task, agency, and whole-body exploration.

Given these limitations of screen-based tasks, which have been consistently highlighted by 20 years of mobile eye-tracking research, it is important to consider the role that screen-based tasks should play in visual perception research as Gibson did 40 years ago. The advantage of screen-based tasks is the ability to show multiple observers the identical stimulus and to manipulate the qualities of those stimuli. Psychophysics and physiological studies of vision would be impossible without the level of control provided by screen-based tasks. However, this comes at the expense of generalization because observers cannot move their bodies or interact with their surroundings. Undoubtedly, the results of screen-based tasks generalize to sedentary activities like watching television and, to a lesser extent, using a computer (even interacting with computer software with a mouse or touchscreen is an active task). Beyond this, generalization is not guaranteed. At the broadest level of analysis, screen-based studies of are concordant with natural tasks in showing that task factors outweigh target appearance in influencing where people look (e.g., Smith & Mital, 2013). But at a more detailed level, comparisons between active and passive observers find differences in visual exploration even when watching the same “stimulus” (e.g., Foulsham et al., 2011).

More work is needed to test which aspects of screen-based tasks generalize to natural tasks. Eye tracking integrated with virtual reality provides a means by which researchers can study whole-body visual exploration while maintaining similar experimental control (and the ability to replicate) screen-based tasks in three dimensions. For example, visual search time is comparable in both 2D (screen) and 3D (virtual reality) search tasks, but those in 3D learn spatial associations better and use fewer fixations to complete the task (C.-L. Li, Pilar Aivar, Kit, Tong, & Hayhoe, 2016)

However, the laboratory and naturalistic paradigms are about more than methodological differences. Widespread acceptance of the screen-based paradigm is rooted in the theoretical commitment to what Gibson termed the “sequence theory” of visual perception. In this view, visual perception is based on a series of discrete retinal images, so studying observers’ perception of images on a screen is an appropriate methodological choice. But Gibson argues that this theory cannot account for our perception of a visual world that is persistent and stable in both space and time:

> The error was to suppose in the first place that perception of the environment is based on a sequence of discrete images. If it is based instead on invariance in a flow of stimulation, the problem of integration does not arise. There is no need to unify or combine different pictures if the scene is *in* the sequence, specified by the invariant structure that underlies the samples of the ambient array. The problem of explaining the experience of…what I would now call the surrounding environment is a false problem. The retinal image is bounded, to be sure, and the foveal image has even smaller bounds, but the ambient array is unbounded. (p. 221-222, emphasis in original).

In this view, attempting to study visual perception by providing the retina with a stimulus image is misguided because awareness of the world comes through the observers' active exploration of the environment over time. This exploration—scanning with the eyes within the head to bring the fovea across different areas on the environment or moving the head to bring objects in and out of the visual field—is a series of complex transformations of optical information. Invariance in such transformations, Gibson claims, carries the information for the stable, unbounded visual world that we do perceive.

Looking ahead, we must test how observers' active visual exploration supports perception of a stable visual world. Rather than asking how an image or video compels the observer to look, we should ask how the observer actively looks with the eyes, body, and head bring objects in and out of view and how observers choose to shift the visual field to maintain contact with their surroundings. Maintaining awareness of the visual world can be thought of as the most basic "task" of visual exploration. The coordination of eyes and head for this "super-ordinate task" might provide the basis from which observers adapt visual exploration to fit the many specific "sub-tasks" of everyday life (e.g., domestic tasks, social interactions, locomotion) that have already been studied.

References

’t Hart, B. M., & Einhauser, W. (2012). Mind the step: complementary effects of an implicit task on eye and head movements in real-life gaze allocation. *Experimental Brain Research, 223*, 233-249.

Bambach, S., Franchak, J. M., Crandall, D. J., & Yu, C. (2014). Detecting hands in children's egocentric views to understand embodied attention during social interaction. *Proceedings of the 36th annual meeting of the Cognitive Science Society*.

Bambach, S., Smith, L. B., Crandall, D. J., & Yu, C. (2016). *Objects in the center: How the infant's body constrains infant scenes.* Paper presented at the IEEE 6th Joint International Conference on Development and Learning and Epigenetic Robotics.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition, 16*, 341-355.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*, 185-207.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art.* Chicago: Chicago University Press.

Diaz, G., Cooper, J., Kit, D., & Hayhoe, M. M. (2013). Real-time recording and classification of eye movements in an immersive virtual environment. *Journal of Vision, 13*, 1-14.

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

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

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what, and when of gaze allocation in the lab and the natural environment. *Vision Research, 51*, 1920-1931.

Franchak, J. M. (2017). Using head-mounted eye tracking to study development. In B. Hopkins, E. Geangu, & S. Linkenauger (Eds.), *The Cambridge Encyclopedia of Child Development* (2nd ed., pp. 113-116). Cambridge: Cambridge University Press.

Franchak, J. M. (under review). Changing opportunities for learning in everyday life: Infant body position over the first year.

Franchak, J. M., & Adolph, K. E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research, 50*, 2766-2774. doi:10.1016/j.visres.2010.09.024

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

Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant-caregiver social looking during locomotor free play. *Developmental Science, 21*, e12626.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development, 82*, 1738-1750. doi:10.1111/j.1467-8624.2011.01670.x

Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., & Adolph, K. E. (2010). *Head-mounted eye-tracking in infants' natural interactions: A new method.* Paper presented at the Proceedings of the 2010 Symposium on Eye Tracking Research and Applications, Austin, TX.

Franchak, J. M., Smith, L. B., & Yu, C. (under review). Developmental changes in how head orientation structures infants' visual experiences.

Franchak, J. M., & Yu, C. (2015). Visual-motor coordination in natural reaching of young children and adults. *Proceedings of the 37th annual meeting of the Cognitive Science Society*.

Frank, M. C., Simmons, K., Yurovsky, D., & Pusiol, G. (2013). Developmental and postural changes in children's visual access to faces. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 454-459). Austin, TX: Cognitive Science Society.

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

Gibson, J. J. (1950). *The perception of the visual world*. Oxford: Houghton Mifflin.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin Company.

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

Hayhoe, M. M., Shrivastava, A., Mruczek, R. E. B., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*, 49-63.

Itti, L., Koch, B., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 1254-1259.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*, 194-203.

Jayaraman, W., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PloS One*. doi:10.1371/journal.pone.0123780

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

Kretch, K. S., & Adolph, K. E. (2017). The organization of exploratory behaviors in infant locomotor planning. *Developmental Science, 20*, e12421.

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*. doi:10.1111/cdev.12206

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential socail interactions are important to social attention. *Proceedings of the National Academy of Sciences, 108*, 5548-5553.

Land, M. F. (1992). Predictable eye-head coordination during driving. *Nature, 359*, 318-320.

Land, M. F. (2004). The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research, 159*, 151-160. doi:10.1007/s00221-004-1951-9

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research, 25*, 296-324.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London B, 352*, 1231-1239.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature, 369*, 742-744.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*, 1311-1328.

Li, C.-L., Pilar Aivar, M., Kit, D. M., Tong, M. H., & Hayhoe, M. M. (2016). Memory and visual search in naturalistic 2D and 3D environments. *Journal of Vision, 16*, 1-20.

Li, D., Babcock, J., & Parkhurst, D. J. (2006). openEyes: A low-cost head-moutned eye-tracking solution. *Proceedings of the 2006 Symposium on Eye Tracking Research and Applications*.

Luo, C., & Franchak, J. M. (in preparation). Infant eye-head alignment to objects and faces during free play.

Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology, 28*, 1224-1233.

Oomen, B. S., Smith, R. M., & Stahl, J. S. (2004). The influence of future gaze orientation upon eye-head coupling during saccades. *Experimental Brain Research, 155*, 9-18.

Patla, A. E., & Vickers, J. N. (1997). Where and when do we look as we approach and step over an obstacle in the travel path. *Neuroreport, 8*, 3661-3665.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research, 41*, 3587-3596.

Pelz, J. B., Hayhoe, M. M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research, 139*(3), 266-277. doi:10.1007/s002210100745

Rosenbaum, D. A. (2010). Walking down memory lane: where walkers look as they descend stairs provides hints about how they control their walking behavior. *The American Journal of Psychology, 122*(4), 425-430.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7*, 1-20.

Smeets, J. B. J., Hayhoe, M. M., & Ballard, D. H. (1996). Goal-directed arm movements change eye-head coordination. *Experimental Brain Research, 109*, 434-440.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision, 13*, 1-24.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*, 1-17. doi:10.1167/7.14.4

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*, 1-23.

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

Tatler, B. W., & Land, M. F. (2015). Everyday visual attention. In J. M. Fawcett, E. F. Risko, & A. Kingstone (Eds.), *The Handbook of Attention*. Boston, MA: MIT Press.

Warren, W. H. (1984). Perceiving affordances: Visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 683-703.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE, 8*, e79659. doi:10.1371/journal.pone.0079659