

Head-Mounted Eye Tracking: A New Method to Describe Infant Looking

John M. Franchak, Kari S. Kretch, Kasey C. Soska, and Karen E. Adolph

New York University

Despite hundreds of studies describing infants' visual exploration of experimental stimuli, researchers know little about where infants look during everyday interactions. The current study describes the first method for studying visual behavior during natural interactions in mobile infants. Six 14-month-old infants wore a head-mounted eye-tracker that recorded gaze during free play with mothers. Results revealed that infants' visual exploration is opportunistic and depends on the availability of information and the constraints of infants' own bodies. Looks to mothers' faces were rare following infant-directed utterances but more likely if mothers were sitting at infants' eye level. Gaze toward the destination of infants' hand movements was common during manual actions and crawling, but looks toward obstacles during leg movements were less frequent.

Each hour, the average toddler hears over 300 words (Hart & Risley, 1999), takes 1200 steps, falls 16 times (Adolph, Badaly, Garciaguirre, & Sotsky, 2010), and spends 30 min playing with objects (Karasik, Tamis-LeMonda, & Adolph, 2011). What is the visual information that accompanies this riot of activity? James (1890) famously suggested that infants' visual experiences are a "blooming, buzzing confusion." But no research has described the visual input in infants' natural interactions: What do infants look at?

A casual scan of the everyday environment makes the possibilities seem endless. But infants do not see everything that surrounds them. Structural and physiological characteristics of the visual system filter out some of the possible stimuli. By 12 months of age, infants' binocular field approximates that of adults—180° wide (Mayer & Fulton, 1993). Input beyond the visual field is temporarily invisible. Attentional processes further filter the input. Infants (like adults) do not process information across all areas of the retina equally (Colombo, 2001). Most often, attention is closely tied to foveal vision—the central 2° of the retina where visual

acuity is the greatest. Moving the eyes shifts the point of visual fixation, and infants make an estimated 50,000 eye movements per day (Bronson, 1994; Johnson, Amso, & Slemmer, 2003).

The Meaning of a Look

Looking—stable gaze during visual fixation or tracking a target during smooth pursuit—is pointing the eyes toward a location. Anything beyond a description of where and when the eyes move is interpretation. Most frequently, researchers interpret gaze direction in terms of the focus of visual attention (Colombo, 2001). Longer durations of looking to one display over another indicate discrimination of stimuli, as in preferential looking paradigms with infants (Gibson & Olum, 1960). Increased looking to a test display is interpreted as interest or surprise, as in infant habituation (Fantz, 1964) and violation of expectation paradigms (Kellman & Arterberry, 2000). Pointing gaze toward certain parts of a display provides evidence about visual exploration, information pick up, and anticipation of events (Gibson & Pick, 2000).

Researchers also use infants' looking behaviors to support inferences about the function of vision. In contrast to the standard infant looking paradigms—preferential looking, habituation, and violation of expectation—in everyday life, looking typically entails more than watching areas of interest. For example, looking may function to coordinate visual attention between an object of interest

This research was supported by National Institute of Health and Human Development Grant R37-HD33486 to Karen E. Adolph. Portions of this work were presented at the 2009 meeting of the International Society for Developmental Psychobiology, Chicago, IL, and the 2010 International Conference on Infant Studies, Baltimore, MD. We gratefully acknowledge Jason Babcock of Positive Science for devising the infant head-mounted eye-tracker. We thank Scott Johnson, Daniel Richardson, and Jon Slemmer for providing calibration videos and the members of the NYU Infant Action Lab for assistance collecting and coding data.

Correspondence concerning this article should be addressed to Karen E. Adolph, Department of Psychology, New York University, 6 Washington Place, Room 410, New York, NY 10003. Electronic mail may be sent to karen.adolph@nyu.edu.

© 2011 The Authors

Child Development © 2011 Society for Research in Child Development, Inc.

All rights reserved. 0009-3920/2011/8206-0003

DOI: 10.1111/j.1467-8624.2011.01670.x

and infants' caregivers. In such joint attention episodes, infants look back and forth between a model's face and the object, monitoring the model's gaze direction and affect (Moore & Corkum, 1994; Tomasello, 1995). In studies of visual control of action, looking reveals how observers collect visual information about objects and obstacles to plan future actions. Infants must coordinate vision and action to move adaptively. Neonates extend their arms toward objects that they are looking at (von Hofsten, 1982). Infants navigate slopes, gaps, cliffs, and bridges more accurately after looking at the obstacles (Adolph, 1997, 2000; Berger & Adolph, 2003; Witherington, Campos, Anderson, Lejeune, & Seah, 2005). Generally speaking, the functional meaning of a look depends on the context.

Measuring a Look

The least technologically demanding method for measuring a look is recording infants' faces with a camera (a "third-person" view). Coders score gaze direction by judging the orientation of infants' eyes and head. However, inferring gaze location reliably depends on the complexity of the environment. Third-person video is sufficient for scoring whether infants look at a toy in their hands when no other objects are present (e.g., Soska, Adolph, & Johnson, 2010). But with multiple objects in close proximity, third-person video typically cannot distinguish which object infants look at or which part of an object they fixate. Consequently, researchers constrain their experimental tasks to insure that gaze location is obvious from the video record. For example, preferential looking capitalizes on large visual displays placed far apart to facilitate scoring from video.

Desk-mounted eye-trackers provide precise measurements of infants' eye movements (Aslin & McMurray, 2004). With desk-mounted systems, the tracker is remote; infants do not wear any special equipment. The eye-tracker sits on the desktop with cameras that point at the infant's face, and software calculates where infants look on the visual display or object within a small, calibrated space. Infants sit on their caregiver's lap or in a high chair and look at displays on a computer monitor or sometimes a live demonstration. Desk-mounted eye-trackers are well suited for measuring visual exploration of experimental stimuli and allow automatic scoring of infants' gaze patterns based on predefined regions of interest. But they are not designed to address questions of where infants look in everyday contexts and interactions. Since the trackers are stationary,

infants must also be relatively stationary, precluding study of visual control of locomotor actions. Use of desk-mounted eye-trackers for studying manual actions is limited because infants' hands and objects frequently occlude the camera's view of their eyes. Moreover, stationary eye-trackers can only measure gaze in a fixed reference frame, so experimenters choose the visual displays to present to infants.

"Headcams"—lipstick-sized cameras attached to elastic headbands—record infants' first-person perspective and circumvent some of the limitations of third-person video and desk-mounted eye tracking for studying natural vision (Aslin, 2009; Smith, Yu, & Pereira, 2011; Yoshida & Smith, 2008). Infants' view of the world differs starkly from an adult's perspective. Infants' hands dominate the visual field and objects fill a larger proportion of infants' visual fields compared to adults (Smith et al., 2011; Yoshida & Smith, 2008). However, the headcam technique comes with a cost. Rather than tracking infants' eye movements, headcams provide only a video of infants' field of view. Researchers sacrifice the measurements of eye gaze provided with desk-mounted eye-trackers (the cross-hairs shown in Figure 1B). Aslin (2009) has addressed this limitation by collecting headcam videos from one group of infants and then using a remote eye-tracker to record gaze from a second set of infants as they watched the headcam videos. However, this approach divorces eye movements from the actions they support.

Headcams do provide sufficient input for researchers to calculate the relative saliency of objects available in the visual field for infants to look at (Yu, 2007). Image processing algorithms detect infants' and mothers' hands and faces and objects on the tabletop. Contrast, motion, and size are computed for faces, hands, and objects, and a weighted combination of these values indicates relative saliency. Nonetheless, saliency computations have important limitations. First, the complexity of the computations requires strict control over the experimental task and environment. In Yu's studies, interactions were limited to tabletop play with three objects. Dyads were dressed in white costumes and the tabletop was ringed with white curtains to facilitate image processing. Although infants could reach for toys on the table, they had only their caregiver and the three toys provided by the experimenter to look at. Second, eye gaze is not necessarily coincident with the most salient location in the visual field. Findings from adults wearing head-mounted eye-trackers in natural tasks demonstrate that gaze is closely linked to the task at hand rather

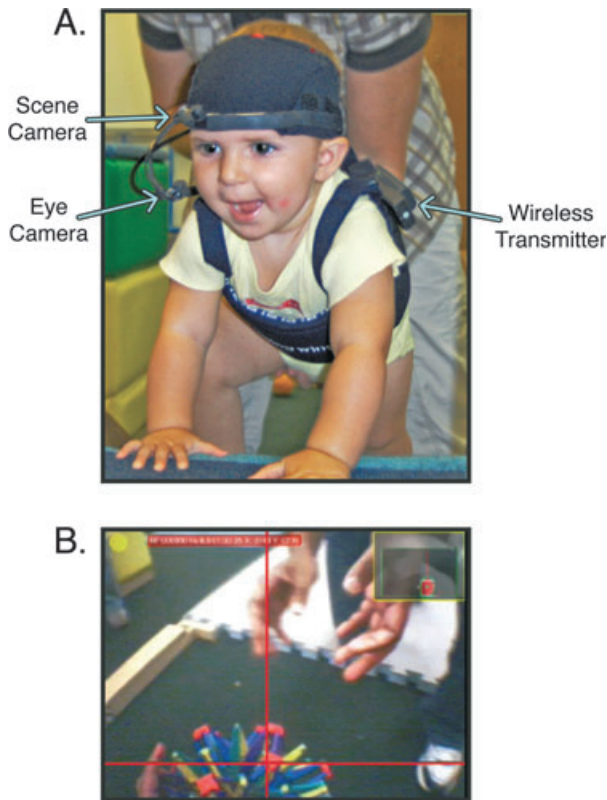


Figure 1. (A) Infant wearing head-mounted eye-tracker, wireless transmitter, and battery pack. (B) Gaze video exported with Yarbus software.

Note. Red crosshairs indicate the infant's point of gaze. Inset shows picture-in-picture video from the eye camera.

than captured by visual saliency (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land, Mennie, & Rusted, 1999). Yu (2007) points out that infants create saliency through their own actions: Infants' movements make some objects more salient than others. Although low-level features (motion, contrast, size) may help to predict gaze location, higher level parameters like the actor's intentions and actions are just as important and, in many cases, are the source of the low-level features.

A few brave souls have attempted to record infants' gaze in reaching tasks using head-mounted eye-trackers (Corbetta, Williams, & Snapp-Childs, 2007). As with the desk-mounted systems and in Yu and colleagues' headcam setups, infants sat in one place, limiting possibilities for engaging with the environment. The experimental arrangement did not allow infants to spontaneously play with toys of their own choosing. Moreover, limitations in the available headgear and software resulted in high subject attrition.

To summarize previous work, existing methods have suited research where limited numbers of

visual stimuli are needed, infants require little range of movement, and when automated gaze processing is beneficial. Yet, technological and procedural considerations have limited researchers' abilities to observe infants during unrestrained movement with multiple visual stimuli available to view. As such, no study has observed where and when infants look while they do the things infants normally do—interact with caregivers, locomote through the environment, and play with objects.

Mobile Eye Tracking in Infants

In the current study, we report a novel methodology for studying infants' visual exploration during spontaneous, unconstrained, natural interactions with caregivers, objects, and obstacles. Our primary aim was to test the feasibility of tracking mobile infants' eye gaze in a complex environment. Our method involves two major innovations. One innovation is the recording technology: This study is the first to use a wireless, head-mounted eye-tracker with freely locomoting infants. The resulting data presented a problem—how to relate eye movements to the actions they support. Accordingly, the second innovation is a method of manually scoring eye movement data based on specific, target action events.

Infant eye-trackers must be small, lightweight, and comfortable. The smallest and lightest adult headgears are modified safety goggles or eyeglasses (Babcock & Pelz, 2004; Pelz, Canosa, & Babcock, 2000). But infants' noses and ears are so small that they cannot wear glasses. Heavy equipment can compromise infants' movements and their precarious ability to keep balance. Additional challenges stem from infants' limited compliance. The equipment must be easy to place on infants and comfortable enough that infants will tolerate wearing it while engaged in normal activities.

The current technology redresses the problems that had previously prevented eye tracking in mobile infants. In collaboration with Positive Science (<http://www.positivescience.com>), we developed a small, light, and comfortable infant headgear. Of the first 44 infants we tested in various experiments, only 4 refused to wear the headgear (9–14 months of age). Of the remaining 40, all yielded usable eye movement data. Here, we report data from the first 6 infants. The eye-tracker transmits wirelessly, so infants could move unfettered through a large room cluttered with toys and obstacles. Thus, we could observe spontaneous patterns of visual exploration during activities of infants'

own choosing. And, in contrast to headcams, the tracker recorded the location of gaze, providing a more in-depth, comprehensive picture of infants' visual world.

A major challenge in eye-tracking research is data management. With desk-mounted trackers, automatic processing of visual fixations is straightforward because the tracker operates in a fixed reference frame. But with head-mounted eye tracking, the reference frame is dynamic. There is no way to designate fixed regions of interest because every head turn reveals a different view of the world. Rather than tailor the task and environment to fit the constraints of automatic coding, we devised a strategy for manually scoring data from eye-tracking videos. Although it is possible to code the entire video frame by frame (e.g., Hayhoe et al., 2003; Land et al., 1999), we opted to code eye movements based on target actions so as to investigate the relations between looking and action. We focused on two types of events: infants' looks to mothers following mothers' vocalizations and infants' looks to objects and obstacles before reaching for objects and navigating obstacles. Our rationale was that these events could illustrate the feasibility and benefits of collecting mobile eye-tracking data across a range of research applications.

We employed a similar coding strategy for each type of event. First, we defined key encounters—each time mothers spoke, infants touched an object, and infants locomoted over a surface of a different elevation. In a second pass through the data, we coded eye movements in a 5-s window before (reaching and locomotion) or after (mothers' vocalizations) the encounter. This method linked visual fixations to actions to provide contextual information for interpreting eye movements.

Method

Head-Mounted Eye Tracker

As shown in Figure 1A, infants wore an ultralight, wireless, head-mounted eye-tracker (Positive Science). The headgear consisted of two miniature cameras mounted on a flexible, padded band that rested above infants' eyebrows. Since the eye-tracker transmitted wirelessly, infants were completely mobile and their behavior was unconstrained. The headgear stayed firmly in place with Velcro straps attached to an adjustable spandex cap. The entire headgear and cap weighed 46 g. An

infrared LED attached to the headgear illuminated infants' right eye for tracking of the dark pupil and creating a corneal reflection. An infrared *eye camera* at the bottom right of the visual field recorded the eye's movements (bottom left arrow in Figure 1A) and a second *scene camera* attached at eyebrow level faced out and recorded the world from the infants' perspective (top left arrow in Figure 1A). The scene camera's field of view was 54.4° horizontal by 42.2° vertical.

Infants also wore a small, wireless transmitter and battery pack on a fitted vest (right arrow in Figure 1A), totaling 271 g. Videos from each camera were transmitted to a computer running Yabus software (Positive Science), which calculated gaze direction in real time. The software superimposed crosshairs that indicated point of gaze on the scene camera video based on the locations of the corneal reflection and the center of the pupil. Pupil data were smoothed across two video frames to insure robust measurement. The *gaze video* (scene video with superimposed point of gaze) and eye-camera video were digitally captured for later coding (Figure 1B).

Participants

Six 14-month-old infants (± 1 week) and their mothers participated. Three infants were girls and three were boys. Families were recruited through commercially available mailing lists and from hospitals in the New York City area. All infants could walk independently and were reported to understand several words. Walking experience ranged from 3.4 to 6.7 weeks ($M = 4.5$ weeks). Two additional infants did not contribute data because they refused to wear the eye-tracker. Unfortunately, we did not assess infants' individual levels of language development, thus limiting our ability to examine links between language knowledge and visual responses to maternal vocalizations.

Calibrating the Eye Tracker

An experimenter placed the eye-tracking equipment on infants gradually, piece by piece (vest, transmitter, hat, and finally headgear), while an assistant and the mother distracted infants with toys until they adjusted to the equipment. To calibrate the tracker (link eye movements with locations in the field of view video), we used a procedure similar to that for remote eye-tracking systems (Falck-Ytter & von Hofsten, 2006; Johnson et al., 2003). Infants sat on their mothers' lap in

front of a computer monitor. Sounding animations drew infants' attention to the monitor. Then a sounding target appeared at a single location within a 3×3 matrix on the monitor to induce an eye movement. Crucially, the calibration software allowed the experimenter to temporarily pause the video feed during calibration so that the experimenter could carefully register the points in the eye-tracking software before the infant made a head movement. Calibration involved as few as three and as many as nine points spread across visual space, allowing the experimenter to tailor the calibration procedure to match infants' compliance. Then, the experimenter displayed attention getters in various locations to verify the accuracy of the calibration. If the calibration was not accurate within $\sim 2^\circ$ (the experimenter judged whether fixations deviated from targets subtending 2°), the experimenter adjusted the location of the eye camera and repeated the calibration process until obtaining an accurate calibration. The entire process of placing the equipment and calibrating the infant took about 15 min. (In subsequent studies, we found it simpler to calibrate using a large posterboard with a grid of windows—an assistant rattles and squeaks small toys in the windows to elicit eye movements.) The eye-tracker was previously measured (when worn by adults) to have a spatial accuracy of $2\text{--}3^\circ$ (maximum radius of error in any direction), and the sampling frequency was 30 Hz (as determined by the digital capture of the gaze video). Although the spatial accuracy is lower than that of typical desk-mounted systems ($0.5\text{--}1^\circ$), the resolution was sufficient for determining the target of fixations in natural settings.

Procedure

After calibration, infants played with their mothers for 10 min in a large laboratory playroom (6.2×8.6 m). Mothers were instructed to act naturally with their infants. Attractive toys (balls, dolls, wheeled toys, etc.) were scattered throughout to encourage infants to manually explore objects, interact with their mothers, and visit different areas of the room. Numerous barriers and obstacles were scattered throughout the room to provide challenges for locomotion: a small slide, a staircase (9-cm risers), 12- to 40-cm-high pedestals, 4-cm-high platforms, and 7-cm-high planks of wood. The experimenter followed behind infants holding "walking straps" to insure their safety when they fell forward, a precautionary measure to prevent the eye camera from poking them in the eye. After the 10 min, the

experimenter verified the calibration by eliciting eye movements to various toys to check the accuracy of the gaze calculation; calibrations for all infants remained accurate throughout the session.

An assistant followed infants and recorded their behaviors on a handheld video camera. A second camera on the ceiling recorded infants' and mothers' location in the room. After the session, the gaze video (field of view with superimposed crosshairs), handheld video, and room video were synchronized and mixed into a single digital movie using Final Cut Pro. Audio was processed using Sonic Visualizer to create a video spectrogram that was synchronized with the session video to facilitate precise coding of mothers' speech onsets.

Data Coding and Analyses

First, coders identified social, crawling, walking, and manual encounters, and then they coded visual behavior in relation to each encounter. Fixations were defined as ≥ 3 consecutive frames (99.9 ms) of stable gaze, following the conventions of previous studies of head-mounted eye tracking (Patla & Vickers, 1997). Because of limits on the spatial resolution of the eye-tracker, saccades smaller than $2\text{--}3^\circ$ may not have been detected.

A primary coder scored all of the data using a computerized video coding system, OpenSHAPA (<http://www.openshapa.org>) that records the frequencies and durations of behaviors (Sanderson et al., 1994). Data were scored manually: Coders watched videos and recorded behaviors in the software. Figure 2 shows a sample timeline of 15 s of one infant's eye movements during her locomotor, manual, and social interactions. A reliability coder scored $\geq 25\%$ of each child's data for each behavior. Interrater agreement on categorical behaviors ranged from 90% to 100% (Kappas ranged from .79 to 1). Correlations for duration variables scored by the two coders ranged from $r_s = .90\text{--}.99$, $p < .05$. All disagreements were resolved through discussion.

The results are primarily descriptive. For inferential statistics, we used analyses of variance (ANOVAs) to compare durations and latencies of eye movements. To compare how looking varied across contexts, we performed chi-square analyses pooling infants' fixations across the sample, following the conventions used in studies of infant reaching in which a small number of participants produced a large number of reaches and variability of individual infants was not of central interest (e.g., Clifton, Rochat, Robin, & Berthier, 1994; Goldfield & Michel, 1986; von Hofsten & Rönnqvist, 1988).

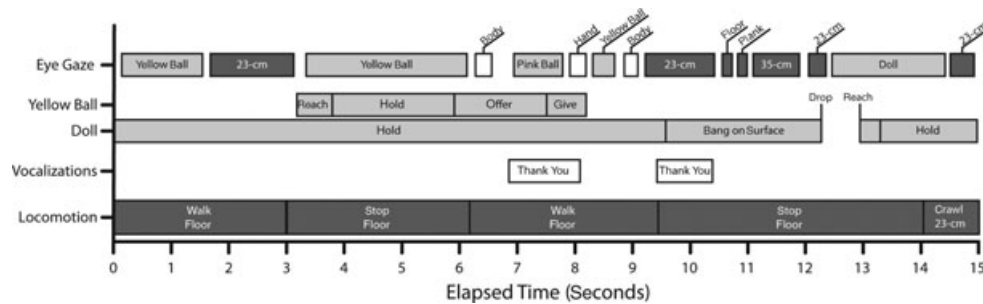


Figure 2. Timeline of 15 s of one infant's interactions with her mother.

Note. The top row shows the infant's eye gaze—white bars are fixations of people (mother, experimenter), light gray bars are fixations of objects, and dark gray bars are fixations of obstacles. The second and third rows mark the infant's manual interactions with two different objects. The fourth row shows each of the mother's vocalizations. The fifth row displays the infant's locomotor activity and the ground surface on which the infant is moving (floor or 23-cm pedestal).

Results and Discussion

Eye Movements in Response to Maternal Speech

Mothers' speech carries a great deal of information: Mothers point out interesting things in the environment, name novel objects, and warn infants of potential danger. But to make use of this information, infants need to know the referent of their mothers' vocalizations. Information from the mother's face is crucial for determining the referent (from eye gaze) and her appraisal of the referent (from facial expressions). Coordinated looking between the mother's face and an object establishes shared reference (Bakeman & Adamson, 1984); previous work has shown that infants benefit from joint attention in word-learning situations (Tomasello & Todd, 1983). Furthermore, if infants look to the mother to learn about her appraisal of an ambiguous situation, infants can use that information to guide their own behavior (Moore & Corkum, 1994). These examples argue that infants should look to mothers in response to their vocalizations to optimally exploit information gathered in social interactions.

Although infants can clearly glean information from their caregivers' faces (Tomasello, 1995), do they choose to do so during unconstrained, natural interactions? Although infants and parents do interact while seated across from one another in everyday life, they also spend a significant portion of their day in less constrained situations. Head-mounted eye tracking provides a way to determine if infants look to their caregivers' faces in a natural setting, even when infants and mothers are far apart from one another or interacting from different heights.

Coding visual fixations relative to mothers' vocalizations. To identify infants' looking behavior follow-

ing mothers' speech, we first identified each vocalization by mothers that was directed to infants. Vocalization onsets were coded from spectrograms and were verified by listening to the audio track; we counted utterances as unique if they were separated by at least 0.5 s. All speech sounds were transcribed and scored, including nonwords (e.g., "ah" and "ooh") that might direct or capture infants' attention. Coders determined which vocalizations referred to objects or places in the room, whether mothers were visible in infants' field of view camera at the onset of the vocalization, and finally whether mothers were standing up or sitting down.

Infants heard vocalizations at a rate of $M = 8.1$ per minute ($SD = 2.8$). However, the range of utterances varied widely among mother-infant dyads: At one extreme, two infants heard 3–4 utterances per minute, and at the other extreme, two infants heard 10–11 utterances per minute. Unlike word-learning scenarios presented in laboratory studies, few of mothers' vocalizations referred to objects or locations in the room. Of the 371 utterances that we transcribed across all infants, 64.7% had no physical referent. These nonreferential utterances included phrases like "Good job," "Ooh," and "Yeah, yeah." Occasionally, mothers named specific objects or places (19.6%, e.g., "Where's the ball?" "Wanna do the slide again?"), and other times they referred to objects or locations without explicitly naming the referent (15.8%, e.g., "What's that?").

Coders scored infants' visual exploration in 5 s following mothers' vocalizations to determine whether infants fixated their mothers after she spoke. Since mothers might be moving, coders did not distinguish between smooth pursuit and fixation of mothers. Smooth pursuit movements and visual fixations were not separated in analysis, and we refer to all visual behaviors collectively as

“fixations.” Coders scored the onset and offset of the first fixation following mothers’ vocalizations. For each fixation, the coder determined what part of the body was fixated: mother’s face, hand, hand holding an object, or any other body part.

Rate of fixations to mothers. Most of infants’ fixations of mothers were in response to rather than in advance of mothers’ speech. For 75.1% of mothers’ vocalizations, infants were looking somewhere else when mothers began speaking; in response to mothers’ speech, infants could look at their mothers or continue to look somewhere else. For the remaining 24.9% of vocalizations, mothers began speaking while infants were already fixating them. We omitted advance fixations from analyses since we were interested in infants’ visual responses to mothers’ vocalizations.

In principle, mothers could use speech to capture infants’ attention so as to provide information about objects or places in the room. Accordingly, 53.9% of mothers’ utterances elicited a shift in gaze to the mother. For the remainder of mothers’ vocalizations, infants maintained their gaze in the same location or shifted gaze to another object or surface.

What determined whether infants shifted their gaze in response to mothers’ vocalizations? Infants looked to mothers marginally more often in response to referential compared to nonreferential speech (63.8% vs. 52.5%, respectively), $\chi^2(1, N = 277) = 3.26, p = .07$, indicating that infants’ decisions to look may reflect whether utterances are potentially informative. Physical proximity is another factor that may influence infants’ visual responses to mothers’ speech—when dyads move freely, infants’ looks to mothers might depend on the mothers’ location relative to infants’ current head direction. On 40.4% of language encounters, mothers were visible in the field of view camera (and not fixated) at the start of the vocalization, indicating that infants’ heads were pointed in the general direction of the mother. Thus, on the remaining 59.6% of language encounters, infants had to turn their heads away from wherever they were looking to bring the mother into the field of view. Decisions to fixate mothers followed an opportunistic strategy: Infants fixated mothers 66.9% of the times mothers were present in the infants’ field of view but only 48.9% of times that mothers were not already in the field of view, $\chi^2(1, N = 292) = 9.36, p = .002$.

Looks to mothers’ faces, bodies, and hands. The high spatial resolution of eye tracking allowed us to determine precisely what region of the mothers’ body infants fixated. Out of the response fixations

to mothers, 50% were directed at mothers’ body, 33.8% at their hands, and only 16.2% at mothers’ face. Given researchers’ emphasis on the importance of joint attention and monitoring of adults’ eye gaze for language learning (Baldwin et al., 1996; Tomasello & Todd, 1983) this proportion is surprisingly low (Feinman, Roberts, Hsieh, Sawyer, & Swanson, 1992). Although referential speech increased the probability of looking to mothers, referential speech did not increase infants’ fixations to mothers’ face ($p = .47$).

Why did infants fixate mothers’ faces so infrequently? The physical context of the playroom may have shaped social interactions: Infants’ small stature and the allure of attractive objects on the floor might explain infrequent fixations of mothers’ faces. If mothers are standing, fixating mothers’ faces requires infants to crane their necks upward, shifting attention from objects and obstacles on the ground. Mothers’ posture varied from one vocalization to the next—they stood upright during 49.4% of vocalizations and sat, squatted, or knelt during the other 50.6%. The location of infants’ fixations varied according to mothers’ posture, $\chi^2(2, N = 156) = 7.32, p = .03$. Infants fixated mothers’ faces during only 10.4% of language encounters while mothers stood upright; the rate of fixations to the face doubled—21.5%—when mothers sat. In fact, mothers sitting accounted for 68.0% of all fixations to mothers’ faces.

Mothers attracted infants’ attention by holding objects and showing objects to infants. Infants were more likely to look at mothers’ hands when mothers were holding objects (72.5% of fixations to hands) compared to when mothers’ hands were empty (binomial test against 50%, $p < .001$), consistent with findings from headcam studies that demonstrate that hands and objects are often in infants’ field of view (Aslin, 2009; Smith et al., 2011; Yoshida & Smith, 2008).

Timing of fixations to mothers. On average, infants responded quickly with visual fixations to mothers, $M = 1.80$ s ($SD = 1.44$), and fixations were relatively short, $M = 0.53$ s ($SD = 0.51$). However, response times ranged widely within and between infants. Infants looked to mothers as early as 0.2 s or as late as 5 s after the start of utterances. Infants fixated mothers more quickly when mothers were already present in the field of view ($M = 1.18, SD = 0.42$) compared to when fixating mothers required a head turn ($M = 2.2, SD = 0.24$), $t(4) = 3.97, p = .02$. Fixations ended $M = 2.30$ s ($SD = 1.53$) after the start of mothers’ utterances, frequently while mothers were still speaking. These brief glances

following the onset of mothers' vocalizations suggest that infants' looks to mothers may function to check in, something like, "Mom is talking and she is over there."

Even though fixations to mothers' faces were infrequent, we considered the possibility that infants might fixate faces for a longer duration if they are processing more information from facial expressions compared to when they look to mothers' hands or bodies. However, a repeated measures ANOVA on the duration of fixations to the three parts of mothers' bodies did not reveal an effect of gaze location on fixation duration, $F(2, 8) = 0.31, p = .61$. Fixations to mothers' faces lasted $M = 0.49$ s ($SD = 0.16$), similar to the durations of fixations to mothers' bodies ($M = 0.41$ s, $SD = 0.06$) and hands ($M = 0.57$ s, $SD = 0.06$). Objects in mothers' hands did not increase duration of fixations to their hands ($p = .48$).

Summary. The key benefit of using head-mounted eye tracking to study social interactions was that we could observe vision during unconstrained, natural interactions. Occasionally, mothers sat on the floor at eye height with infants, but more typically mothers stood and watched as their children explored toys and obstacles. Studying infants' responses to mothers' speech across varying physical contexts proved critical—fixations to mothers depended on convenience. Infants readily looked to their mothers when they did not have to turn to find her, and were more likely to look at her face if she was low to the ground. How might we reconcile our finding that free-moving infants rarely look at their mothers' face with the large body of work that touts joint attention as a primary facilitator of learning? One possibility is that the accumulated volume of input from mothers makes up for the relatively low proportion of responses from infants. Another possibility is that infants might not need to look to mothers to know what she is referring to. Most of the time mothers' speech did not refer to anything in particular, so infants had little reason to look to her. When mothers did name a specific referent, they usually referred to the object or location that infants were already attending to (Feinman et al., 1992).

Visual Guidance of Crawling, Walking, and Reaching

How might infants use vision to guide manual and locomotor actions? Laboratory tasks with older children and adults suggest that prospective fixations of key locations contribute to successful action: Participants reliably fixate objects while

reaching for them (Jeannerod, 1984) and often fixate obstacles before stepping onto or over them (Di Fabio, Zampieri, & Greany, 2003; Franchak & Adolph, 2010; Patla & Vickers, 1997). Restricting visual information disrupts the trajectories of the hands and feet (Cowie, Atkinson, & Braddick, 2010; Kuhtz-Bushbeck, Stolze, Johnk, Boczek-Funcke, & Illert, 1998; Sivak & MacKenzie, 1990). The role of visual information in motor control depends on the timing of fixations of objects and obstacles. When reaching, object fixations are often maintained until after hand contact, providing online visual feedback that can be used to correct the trajectory of the hand. In contrast, when walking, children and adults always break fixations of the obstacle before the foot lands—obstacle fixations provide feed-forward (as opposed to feedback) information.

Although children and adults seem to rely on fixations in laboratory tasks, other sources of evidence demonstrate that object and obstacle fixations are not mandatory. When making a peanut butter and jelly sandwich, adults occasionally reach for objects they had not previously fixated (Hayhoe et al., 2003). When navigating a complex environment, children and adults often guide locomotion from peripheral vision or memory, choosing not to fixate obstacles in their paths (Franchak & Adolph, 2010). These examples highlight the disconnect between what people do in constrained tasks and what they actually choose to do in tasks that more closely resemble everyday life. However, we know very little about how vision functions in everyday life, and the evidence is solely from studies with adults (Hayhoe et al., 2003; Land et al., 1999; Pelz & Canosa, 2001) and older children (Franchak & Adolph, 2010).

Scoring eye movements relative to reaching, walking, and crawling. We coded three actions during which infants moved a limb to a target object or obstacle. *Reaching encounters* were scored every time infants touched objects. The onset of the reaching event was the moment that the hand first contacted the object. Objects that could not be moved (such as large pieces of equipment and furniture) were not counted. Repeated touches of a single object (e.g., rapid tapping of a toy) that occurred within 2 s of the initial touch were excluded from analysis. *Walking encounters* were coded each time that infants walked up, down, or over an obstacle of a different height. *Crawling encounters* were coded each time that infants crawled hands-first up, down, or over an obstacle of a different height. The onsets of walking and crawling encounters were defined by

the moment the leading limb (foot or hand) touched the new surface.

Infants encountered many objects and obstacles as they explored the room—they touched objects at a rate of $M = 3.0$ times per minute ($SD = 0.59$) and walked up, down, or over obstacles at a rate of $M = 1.6$ times per minute ($SD = 1.2$). Crawling was less frequent; infants averaged only $M = 0.64$ encounters per minute ($SD = 0.67$) in a hands-first crawling posture; one infant never crawled on an obstacle. Although infants preferred to walk, balance while walking was precarious: Infants tripped or fell on 33.3% of walking encounters compared to only 3.1% of crawling encounters, $\chi^2(1, N = 104) = 11.1, p = .001$.

For each type of event, the coder determined if and when infants fixated the target object (reaching) or obstacle (walking or crawling) in the 5 s prior to limb contact. If there were multiple fixations of the target in the 5-s window, the coder scored the fixation that occurred closest to the moment of the event. In cases where the object was moving before the encounter, such as when the mother handed a toy to the infant, we also counted instances of smooth pursuit of the object lasting at least three consecutive frames. Because some obstacle surfaces were extensive (e.g., the lab floor), fixations of the obstacle were counted if infants held their gaze held on any part of the surface within one step's length from the actual point of hand or foot contact.

If infants fixated the target, coders scored the initiation of the fixation relative to the moment of the encounter—how far in advance infants began looking at the target obstacle or object. By definition, fixations had to be initiated before the moment of limb contact (Figure 3A). Coders also determined when infants terminated fixations—the moment infants broke fixation to the obstacle or object. Fixation termination could occur either before the event, indicating that infants looked away from their goal before the limb touched (Figure 3A), or after the event, indicating that infants maintained gaze on the target even after limb contact (Figure 3B).

Fixations of objects and obstacles. Infants' prospective fixations of objects and obstacles varied according to the action performed, $\chi^2(2, N = 309) = 13.7, p = .001$. Target fixations were most frequent when infants reached for objects (88.5%) and crawled on obstacles (90.3%). Infants fixated less frequently when walking up, down, and over obstacles—only 71.9%. What might account for a higher fixation rate for reaching and crawling compared to walking?

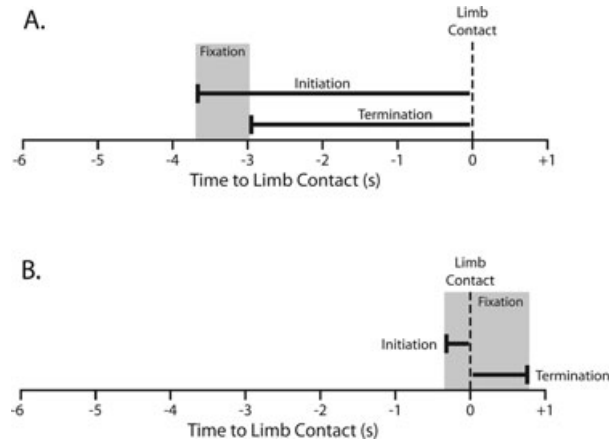


Figure 3. Hypothetical timeline showing fixation initiation and termination relative to the moment of limb contact (vertical dashed line). (A) Fixation beginning and ending before limb contact. (B) Fixation continuing through the moment of limb contact, providing online visual information of the target.

Possibly, differences in fixation rates reflect infants' expertise in the three motor skills. If greater experience with a skill implies more frequent fixations of target locations, fixation rates would be higher for crawling and reaching compared to walking. Infants in the current study had only $M = 1.1$ months of walking experience compared to $M = 7.1$ months of crawling experience. Accordingly, infants often tripped and fell while walking but seldom erred while crawling. We did not collect reaching onset ages, however, typically developing infants begin reaching in free sitting and prone positions at 5 months (Bly, 1994)—14-month-old infants in the current study should have accrued about 9 months of experience reaching for objects.

By the same logic, however, we should expect walkers to fixate obstacles more often as they become more experienced. But infants fixated obstacles much more often (72% of encounters) than 4- to 8-year-old children and adults who fixated obstacles before only 59% and 32% of encounters, respectively (Franchak & Adolph, 2010). The lower fixation rates of children and adults compared to infants indicates that, with experience, walkers do more with less: Adults guide locomotion efficiently using peripheral vision or memory, keeping foveal vision in reserve for serving other tasks. Furthermore, we found no evidence that prospective obstacle fixations improved walking performance for infants. They erred 33.4% of the time when fixating and 33.4% when guiding without fixations. Most likely, infants' errors resulted from poor execution of the movement rather than a lack of information about obstacles.

Physical constraints of the body provide a better explanation for the discrepancies in fixation rates between walking, crawling, and reaching. The hands are situated such that they are likely to be in front of the eyes, and previous research has demonstrated that infants' hands are often present in their visual fields while manipulating objects (Smith et al., 2011; Yoshida & Smith, 2008). Likewise, when in a crawling posture, surfaces on the ground fill infants' field of view. Frequent fixations of objects while reaching and obstacles while crawling might simply reflect how commonly they are in view. In contrast, when walking upright, fixating an obstacle at the feet requires a downward head movement. Just like infants' lower rate of fixations to mothers when they were outside the field of view, infants might fixate obstacles less frequently while upright because it requires an additional head movement to look at the ground surface.

Timing of fixations to objects and obstacles. We looked for converging evidence in the timing of infants' fixations, taking advantage of the temporal resolution afforded by the eye-tracker's recordings. For each fixation, we calculated the initiation time of the fixation relative to the moment of limb contact. Initiation times indicate how far in advance infants collected foveal information about the target object or obstacle. We binned fixations in six 1-s intervals before limb contact, and calculated the proportion of each infant's fixations that fell in each interval. Figure 4A shows the distributions of initiation times for reaching, walking, and crawling averaged across the six infants.

A 3 (action) \times 6 (interval) repeated measures ANOVA revealed an action by interval interaction across the three distributions, $F(10, 40) = 4.82$, $p < .001$. Reaching and crawling distributions show a sharp peak in the final second before contact, but in contrast, the walking distribution increases gradually up until the moment of contact. We followed up this observation by testing quadratic trend components in each distribution against a Bonferroni-corrected alpha level of .017. Quadratic models fit the fixation initiation distributions for reaching, $F(1, 5) = 151.0$, $p < .001$, and crawling, $F(1, 4) = 22.25$, $p = .01$. However, walking fixation initiations did not show a quadratic component, $F(1, 5) = 0.28$, $p = .62$.

Differences in the shapes of these distributions demonstrate that the consistency of visual-motor guidance varied when guiding the hands compared to guiding the feet. Quadratic trends for reaching and crawling show that infants most frequently initiated fixations in the final second before limb

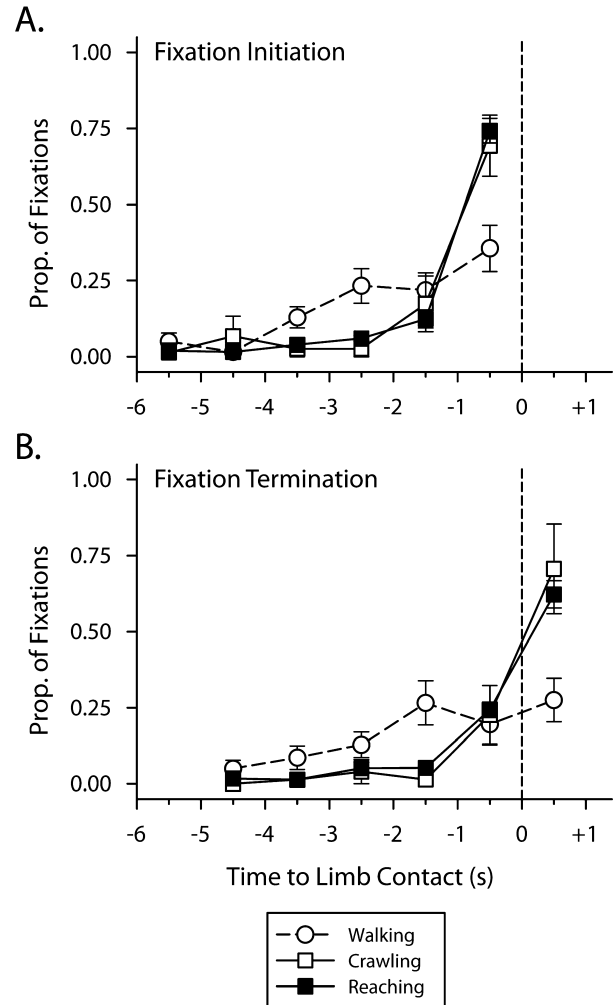


Figure 4. Histograms showing the timing of fixations to objects and obstacles for reaching, walking, and crawling encounters. Note. Vertical dashed lines indicate the moment of limb contact. Fixation initiation is shown in (A). Fixation termination is shown in (B).

contact, and rarely began earlier than 1 s in advance. Despite reaching for different objects and crawling on various obstacles, visual-motor coordination was consistent from encounter to encounter: Infants fixated the target in the final second 76.4% of the time when reaching and 60.7% of the time when crawling. The lack of a quadratic trend for walking suggests no single, dominant timing pattern for walking; infants exhibited greater variability in the initiation times for walking encounters compared to reaching and crawling.

We also considered fixation termination to determine how much time separates the collection of visual information from the action itself. Unlike fixation initiation, termination could occur after the encounter. Fixations that continued past the

moment of limb contact illustrate online visual guidance—infants watched as their hands and feet contacted the target. Like initiation, we binned termination times in six 1-s intervals. These intervals, however, ranged from 5 s before limb contact to 1-s after limb contact to include fixations that ended after the moment of contact.

Like fixation initiation, fixation termination distributions show characteristic patterns of visual-motor guidance for actions executed with the hands compared to the feet. We conducted a 3 (action) \times 6 (interval) repeated measures ANOVA on the proportion of termination times that ended in each interval. A significant Action \times Interval interaction demonstrated different distributions of termination times by action, $F(10, 40) = 3.41, p = .003$. Visual inspection of Figure 4B shows that, just like initiation, termination times for reaching and crawling have sharp peaks. Trend analyses confirmed a quadratic trend for reaching, $F(1, 5) = 77.8, p < .001$, and a marginal trend for crawling, $F(1, 4) = 9.07, p = .04$ (tested against a Bonferroni-corrected alpha of .017). Walking did not show a significant quadratic trend, $F(1, 5) = 0.13, p = .73$.

In addition, termination times indicate when infants generated online visual information that could be used to guide limbs to the target. Children and adults sometimes use visual feedback to guide the hand (Jeannerod, 1984; Kutz-Bushbeck et al., 1998). Infants usually looked at objects when the hand made contact. The peak of the reaching distribution occurred after the moment of contact—62.2% of reaching encounters were accompanied by online visual information. In contrast, whereas children and adults never use feedback to guide walking (Franchak & Adolph, 2010), infants watched obstacles 27.5% of the time at the moment of foot contact. Infants guided crawling online during 70.7% of encounters, lending further support for parity in the patterns of visual guidance for hand-actions.

Summary. Head-mounted eye tracking made it possible to study infants' eye movements during free, unconstrained reaching and locomotion. Eye movements functioned in support of ongoing motor action: Infants fixated objects and obstacles, generating visual information that could be used to guide actions prospectively. However, fixations were not required to perform actions adaptively. Like children and adults, infants were able to use peripheral vision or memory to guide reaching and locomotion (Franchak & Adolph, 2010; Hayhoe et al., 2003). But unlike older participants, infants predominantly relied on foveal vision, and, more-

over, used online visual guidance to control their actions, even when walking on obstacles.

This study also revealed differences in visual guidance for actions involving the hands compared to those with the feet. When reaching and crawling, infants frequently fixated key locations and displayed remarkably similar, consistent timings for initiating and terminating fixations. In contrast, infants did not fixate obstacles as frequently while walking, and the timing of fixations varied greatly from one encounter to the next. These results are surprising because walking and crawling share a similar function. Reaching serves a very different function, yet crawling more closely resembled reaching, presumably due to physical constraints of the body—the targets of actions involving the hands are more likely to be in the infants' field of view compared to actions involving the feet.

Conclusion

This study demonstrates the viability of head-mounted eye tracking as a method for studying infants' visual exploration during natural interactions, allowing perception, action, and social behavior to be observed in complex environments. Studying natural behavior is crucial—physical constraints of the environment shape social interactions, and physical constraints of the body affect how visual information of action is generated. Infants' spontaneous visual exploration provides the basis for characterizing opportunities for learning in development. Infants accumulate vast amounts of experience to support learning in every domain. Encounters with obstacles, objects, and mothers occur frequently and overlap in time—infants navigate obstacles on the way to new objects, all the while hearing their mothers' speech. Infants direct their eyes to relevant areas in the environment to meet changing task demands, seamlessly switching between different patterns of visual exploration.

Future research might ask questions about the intersections of different behaviors and the role of vision. How do infants use vision when switching between multiple, ongoing tasks? Maybe infants in the current study chose not to look to mothers at times when they were using vision to guide reaching or locomotor movements, or perhaps mothers' speech provided infants with information about obstacles in their paths, helping to focus their attention. Moreover, sequential analyses of eye movements can help reveal the real-time processes involved in social interaction and motor control. This study provides a first step in describing the

nature of the visual input—what is actually filtered by the visual system—as well as how visual information functions in everyday interactions.

References

- Adolph, K. E. (1997). Learning in the development of infant locomotion. *Monographs of the Society for Research in Child Development, 62*.
- Adolph, K. E. (2000). Specificity of learning: Why infants fall over a veritable cliff. *Psychological Science, 11*, 290–295.
- Adolph, K. E., Badaly, D., Garciaguirre, J. S., & Sotsky, R. (2010). *15,000 steps: Infants' locomotor experience*. Manuscript submitted for publication.
- Aslin, R. N. (2009). How infants view natural scenes gathered from a head-mounted camera. *Optometry and Vision Science, 86*, 561–565.
- Aslin, R. N., & McMurray, B. (2004). Automated corneal-reflection eye tracking in infancy: Methodological developments and applications to cognition. *Infancy, 6*, 155–163.
- Babcock, J., & Pelz, J. B. (2004). Building a lightweight eyetracking headgear. *ETRA '04: Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*.
- Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother–infant and peer–infant interaction. *Child Development, 55*, 1278–1289.
- Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word–object relations. *Child Development, 67*, 3135–3153.
- Berger, S. E., & Adolph, K. E. (2003). Infants use handrails as tools in a locomotor task. *Developmental Psychology, 39*, 594–605.
- Bly, L. (1994). *Motor skills acquisition in the first year*. San Antonio, TX: Therapy Skill Builders.
- Bronson, G. W. (1994). Infants' transitions toward adult-like scanning. *Child Development, 65*, 1243–1261.
- Clifton, R. K., Rochat, P., Robin, D. J., & Berthier, N. E. (1994). Multimodal perception in the control of infant reaching. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 876–886.
- Colombo, J. (2001). The development of visual attention in infancy. *Annual Reviews of Psychology, 52*, 337–367.
- Corbetta, D., Williams, J., & Snapp-Childs, W. (2007, March). *Object scanning and its impact on reaching in 6- to 10-month old infants*. Paper presented at the 2007 meeting of the Society for Research in Child Development, Boston, MA.
- Cowie, D., Atkinson, J., & Braddick, O. (2010). Development of visual control in stepping down. *Experimental Brain Research, 202*, 81–188.
- Di Fabio, R. P., Zampieri, C., & Greany, J. F. (2003). Aging and saccade-stepping interactions in humans. *Neuroscience Letters, 339*, 179–182.
- Falck-Ytter, T., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience, 9*, 878–879.
- Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146*, 668–670.
- Feinman, S., Roberts, D., Hsieh, K. F., Sawyer, D., & Swanson, D. (1992). A critical review of social referencing in infancy. In S. Feinman (Ed.), *Social referencing and the construction of reality in infancy* (pp. 15–54). New York: Plenum.
- Franchak, J. M., & Adolph, K. E. (2010). Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research, 50*, 2766–2774.
- Gibson, E. J., & Olum, V. (1960). Experimental methods of studying perception in children. In P. H. Mussen (Ed.), *Handbook of research methods in child development* (pp. 311–373). New York: Wiley.
- Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development*. New York: Oxford University Press.
- Goldfield, E., & Michel, G. (1986). Spatiotemporal linkage in infant interlimb coordination. *Developmental Psychobiology, 19*, 259–264.
- Hart, B. H., & Risley, T. R. (1999). *The social world of children: Learning to talk*. Baltimore: Paul H. Brookes.
- Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*, 49–63.
- James, W. (1890). *The principles of psychology*. London: Macmillan.
- Jeannerod, M. (1984). The timing of natural prehension movements. *Journal of Motor Behavior, 16*, 235–254.
- Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye tracking paradigm. *Proceedings of the National Academy of Sciences, 100*, 10568–10573.
- Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition from crawling to walking and infants' actions with objects and people. *Child Development, 82*, 1199–1209.
- Kellman, P. J., & Arterberry, M. E. (2000). *The cradle of knowledge: Development of perception in infancy*. Cambridge, MA: MIT Press.
- Kuhtz-Bushbeck, J. P., Stolze, H., Johnk, K., Boczek-Funcke, A., & Illert, M. (1998). Development of prehension movements in children: A kinematic study. *Experimental Brain Research, 122*, 424–432.
- 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.
- Mayer, D. L., & Fulton, A. B. (1993). Development of the human visual field. In K. Simons (Ed.), *Early visual development, normal and abnormal* (pp. 117–129). New York: Oxford University Press.
- Moore, C., & Corkum, V. (1994). Social understanding at the end of the first year of life. *Developmental Review, 14*, 349–372.

- 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., Canosa, R., & Babcock, J. (2000). *Extended tasks elicit complex eye movement patterns*. Paper presented at ETRA '00: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications.
- Sanderson, P. M., Scott, J. J. P., Johnston, T., Mainzer, J., Wantanbe, L. M., & James, J. M. (1994). MacSHAPA and the enterprise of Exploratory Sequential Data Analysis (ESDA). *International Journal of Human-Computer Studies*, *41*, 633–681.
- Sivak, B., & MacKenzie, C. L. (1990). Integration of visual information and motor output in reaching and grasping: The contributions of peripheral and central vision. *Neuropsychologia*, *28*, 1095–1116.
- Smith, L. B., Yu, C., & Pereira, A. F. (2010). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*, 9–17.
- Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates three-dimensional object completion. *Developmental Psychology*, *46*, 129–138.
- Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Erlbaum.
- Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, *4*, 197–212.
- von Hofsten, C. (1982). Eye-hand coordination in the newborn. *Developmental Psychology*, *18*, 450–461.
- von Hofsten, C., & Rönqvist, L. (1988). Preparation for grasping an object: A developmental study. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 610–621.
- Witherington, D. C., Campos, J. J., Anderson, D. I., Lejune, L., & Seah, E. (2005). Avoidance of heights on the visual cliff in newly walking infants. *Infancy*, *7*, 3.
- Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*, 229–248.
- Yu, C. (2007). Embodied active vision in language learning and grounding. In L. Paletta & E. Rome (Eds.), *Attention in cognitive systems* (pp. 75–90). Berlin: Springer.