Gaze behavior and the perception of egocentric distance

Daniel A. Gajewski

Department of Psychology,
George Washington University, Washington, DC, USA   ✉

Courtney P. Wallin

Department of Psychology,
George Washington University, Washington, DC, USA   ✉

John W. Philbeck

Department of Psychology,
George Washington University, Washington, DC, USA   ✉

The ground plane is thought to be an important reference for localizing objects, particularly when angular declination is informative, as it is for objects seen resting at floor level. A potential role for eye movements has been implicated by the idea that information about the nearby ground is required to localize objects more distant, and by the fact that the time course for the extraction of distance extends beyond the duration of a typical eye fixation. To test this potential role, eye movements were monitored when participants previewed targets. Distance estimates were provided by walking without vision to the remembered target location (blind walking) or by verbal report. We found that a strategy of holding the gaze steady on the object was as frequent as one where the region between the observer and object was fixated. There was no performance advantage associated with making eye movements in an observational study (Experiment 1) or when an eye-movement strategy was manipulated experimentally (Experiment 2). Observers were extracting useful information covertly, however. In Experiments 3 through 5, obscuring the nearby ground plane had a modest impact on performance; obscuring the walls and ceiling was more detrimental. The results suggest that these alternate surfaces provide useful information when judging the distance to objects within indoor environments. Critically, they constrain the role for the nearby ground plane in theories of egocentric distance perception.

Introduction

The perception of a scene can be characterized as the product of a sequential sampling process. Because the region of high visual acuity is limited, the eyes are directed from one location in a scene to another at a rate of around three times per second to resolve and encode the details (Henderson & Hollingworth, 1998; Rayner, 1998). In addition, information is extracted from the environment primarily during fixation, when the point of regard is stable (Matin, 1974). Human vision is also active. Eye movements are integral to the ongoing visual processing of a scene, and gaze behavior is an overt manifestation of attentional-selection strategies (for reviews, see Findlay & Gilchrist, 2003; Henderson, 2003). Since the classic studies of Buswell (1935) and Yarbus (1967), it has been known that informative regions of a scene are preferentially fixated and that where one looks largely depends on goals imposed by the observer's task. In this article we examine the role for eye movements in the perception of the distance between an object and an observer. There has been considerable interest in the role for eye movements in other spatial domains, such as in the control of locomotion (e.g., Hollands, Patla, & Vickers, 2002) and obstacle avoidance (e.g., Franchak & Adolph, 2010; Hayhoe, Gillam, Chajka, & Vecellio, 2008; Patla & Vickers, 1997; Rothkopf, Ballard, & Hayhoe, 2007). However, because these tasks are in the service of specific actions (e.g., steering and control of footholds and step elevation) and generally under continuous visual control, the role for eye movements in the development of a more general-purpose representation of distance remains unclear. There is some indication in the literature that eye movements can enhance the perception of depth intervals (Foley & Richards, 1972; Wist & Summons, 1976), particularly in reduced-cue contexts. However, work in the domain of distance perception is largely disconnected from what is known about how the visual system interrogates an image with shifts of attention and by the execution of saccadic eye movements. Where do observers look in a scene to build up a maximally

accurate representation of object distance? Are eye movements even needed? What regions in a scene are informative for making judgments of distance?

At least two theoretical frameworks bear on these fundamental questions. To begin, He and colleagues have put forth a body of work and an account of distance perception that does posit a very specific role for selection processes (He, Wu, Ooi, Yarbrough, & Wu, 2004; Ooi & He, 2007; J. Wu, He, & Ooi, 2008; B. Wu, Ooi, & He, 2004). Their account builds from the idea that the ground surface provides an important frame of reference for localizing objects (Gibson, 1950, 1979; Sedgwick, 1986; see also Bian, Braunstein, & Andersen, 2005, 2006). Specifically, they suggest that an accurate perception of distance depends on the acquisition of an accurate representation of the ground surface, which in turn depends on a *sequential surface integration process* (SSIP). Because the nearby distance and depth cues are more reliable, an accurate representation of the immediate ground surface provides an important constraint on the development of a ground representation at further distances where the distance and depth cues are less reliable. Thus, by this account, information about the nearby ground surface provides a kind of anchor for the integration of information about farther patches of the ground surface. The results of several experiments have supported this idea. For example, targets were localized less accurately when viewed through an aperture that occluded the nearby ground plane and performance was better when observers were asked to scan from the near ground to the object than vice versa (B. Wu et al., 2004). In addition, J. Wu et al. (2008) have shown that if a task biases participants' initial attention toward distances lying beyond the immediate ground surface, the accuracy of distance judgments suffers, presumably because the crucial information from the immediate ground plane has not been selected or picked up. Eye movements were not monitored in any of these studies, and the account is agnostic about whether selection can be accomplished covertly or if overt shifts of attention are required. Nevertheless, while attention can be directed covertly, attention and the direction of gaze are normally dynamically coupled (e.g., Henderson, 1992; Henderson, Pollatsek, & Rayner, 1989; Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995; Shepherd, Findlay, & Hockey, 1986). Thus, the benefit of scanning from near to far would predict an important role for eye movements.

A potentially important role for eye movements is also suggested by a dynamic framework we have developed on the basis of performance in a limited-viewing-time paradigm (Gajewski, Philbeck, Pothier, & Chichka, 2010; Gajewski, Philbeck, Wirtz, & Chichka, 2013). Participants judged the distance to a previewed object by walking to its remembered location with their eyes closed (blind walking). The duration of viewing was controlled with a liquid-crystal shutter window, which can deliver glimpses of various durations and follow them up with a masking stimulus (Pothier, Philbeck, Gajewski, & Chichka, 2009). When the duration of viewing was brief (9–220 ms), response sensitivity was quite high (slopes relating response distance to target distance were near 1) but coupled with a bias towards underestimation (as large as −25%). The high degree of response sensitivity, particularly when observed with a viewing duration that approached the limits of one's ability to even detect the object, suggests an important role for angular declination as a source of information about distance. Angular declination is the direction of the target below eye level (Ooi, Wu, & He, 2001), and the brief viewing durations afford little or no more than the extraction of the target's directional location in the visual field. In support of this idea, we have recently shown that floor-level targets viewed briefly in a well-lit environment are localized quite similarly to glowing targets in an otherwise dark room, a context where angular declination is arguably the only functional cue available (Gajewski et al., 2013). In contrast, angular declination is not informative about distance when targets are presented at eye level, and accordingly, response sensitivity in this case is low (Gajewski et al., 2010). However, even with floor-level targets, responses became more sensitive and less biased when the duration of viewing was extended to several seconds. The longer viewing duration presumably affords the extraction of additional useful sources of information. Because the limited viewing times employed (even when well above the detection threshold) were safely below the typical latency for a saccade, we presume that the benefit of the longer viewing duration is tied to the fact that more extended viewing allows for the execution of saccadic eye movements.

If eye movements are needed to develop a maximally accurate representation of distance, then gaze behavior is also expected to highlight the regions of a scene that are informative about distance. Consider the scenario that is illustrated in Figure 1. If the geographic slant of the ground surface is accurately represented, the distance D to an object resting on the ground is specified by a trigonometric function of eye height h and the angular declination of the object with respect to eye level α. Ooi, Wu, and He (2006) suggest that when cues to the ground surface are unavailable (as when illuminated targets are viewed in an otherwise dark room) or poorly specified (as when the nearby ground surface is occluded), targets are perceived as though resting on a surface that is increasingly elevated with distance (as though the ground has a nonzero slant η). This perceptual tendency Ooi et al. have termed the *intrinsic ground plane*. As can be seen, the target
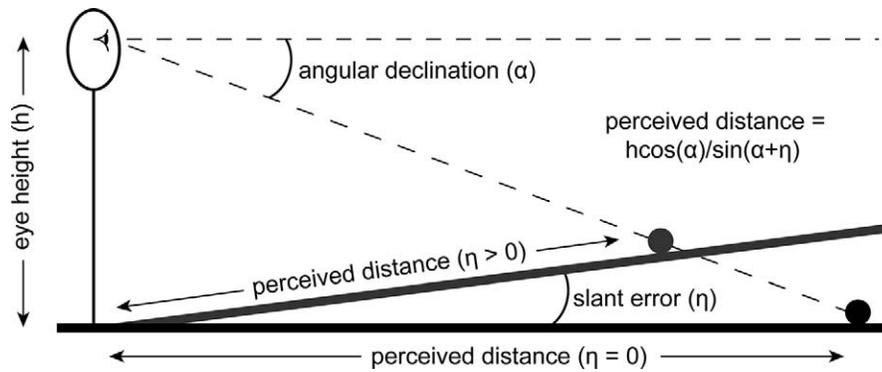
Figure 1. The perceived distance to a floor-level target can be represented as a trigonometric function of angular declination relative to eye level (α) and eye height (h). In addition, according to the SSIP framework, the ground surface is represented with some slant error (η) when cues to the nearby ground surface are not available. Judgments of distance are then distorted because the line of sight (or the angular direction of the target) no longer intersects with an accurate representation of the ground surface.

distance d specified by angular declination is underestimated in this case, because the target vector no longer intersects with the true ground surface. Critically, this underestimation increases with target distance, and the SSIP account suggests that the nearby ground surface must be fixated (or covertly attended) to generate a more accurate overall surface representation.

Performance changes across viewing durations in the limited-viewing-time paradigm were indeed consistent with the idea that increased time might afford the execution of eye movements to the near ground surface. In particular, a reduction in slant error predicts an increase in response sensitivity, which indeed occurred when viewing time was extended (Gajewski et al., 2010). However, performance was also marked by underestimation overall, which suggests that errors in the representation of ground slant may not be the sole factor hindering performance when viewing time is limited. Critically, the underestimation bias is diminished when a more extended viewing condition is administered in the first block of trials (Gajewski et al., 2010), and a single 15-s visual preview of the room environment has been shown to benefit performance on subsequent brief-glimpse trials (Gajewski et al., 2013). During debriefings, observers often spontaneously express frustration at their inadequate sense for the size of the room. Given the reliability of angular declination as an egocentric distance cue, it is unclear why room size should matter. Gajewski et al. (2013) have posited that a representation of the greater space could support performance by providing a structured spatial reference into which cues are integrated as they become available. Although the mechanism underlying such a benefit is admittedly speculative at this point, a representation of this kind appears to contribute to the scale of perceived distances within the environment. This framework predicts greater visual exploration of the room than would be predicted by SSIP. By this view, an enhanced representation of surface slant is important but is only

one piece of the puzzle. Fixations on the nearby ground surface would be expected, though other regions may also be informative, such as the edges formed by the meeting of walls with the floor and ceiling.

Given the theoretical bases of interest just discussed, the present study had three primary objectives. First, we wished to determine where observers look when required to make judgments of distance. Do they preferentially fixate the near ground surface, as might be expected by the SSIP theory? If so, could the gaze behavior be characterized as near-to-far scanning? Do they visually explore the room, as would be expected if an enhanced representation of the overall room space is important or useful? Second, we wished to determine whether preferred gaze strategies would depend on response mode. Our primary response mode (blind walking) was contrasted with a non-action-based response mode (verbal report). If the selection of information from the scene is task specific, we might expect different regions to be preferentially fixated. For example, it has been suggested that verbal report estimates of distance are more aided by landmarks (Andre & Rogers, 2006). In contrast, blind walking entails planning an action to be carried out over the surface that extends from the observer to the object. As a result, we might expect a stronger preference for fixating the near ground surface when blind walking is the response mode. Finally, we wished to determine the efficacy of the observed gaze strategies. Whether observers prefer to fixate the near ground surface or not, is performance better when this region is fixated, and does the efficacy itself depend on response mode? We begin with an observational study in which response mode (verbal report vs. blind walking) was manipulated. This approach was advantageous for allowing observers to choose their own gaze strategy while also supporting an analysis of performance as a function of gaze strategy, and it served as the basis for

the more pointed experimental manipulations that follow.

Experiment 1

Method

Participants

Twenty-eight students from the George Washington University community participated in exchange for course credit. All participants had normal or corrected-to-normal vision and were unaware of the aims of the study.

Stimuli and design

Targets were sheets of yellow foam placed on the floor at one of four distances (2.75, 3.50, 4.25, and 5.00 m). Targets were salient against a gray carpet with a subtle texture and were cut to project with the same approximate visual angle (about $0.67° \times 4.94°$) across distances. The room was a mostly empty laboratory space with a bookshelf to the left, a shelf and beam to the right, and an angled door at the back of the room (see Figure 2). The distance from the point of observation to the back wall was 7.4 m. Participants were randomly assigned to one of two response modes (blind walking or verbal report). Each target distance was employed five times, with order completely randomized.

Apparatus

The experiment employed a head-mounted eye-tracking system (ISCAN ETL-500). The eye tracker comprises two cameras fastened to a headband: a scene camera that captures video of the room from the participant's point of view and an eye camera that captures video of the participant's left eye reflected from a small two-way mirror. The mirror in front of the eye is supported from the side and is completely clear—it does not obstruct the participants' view of the ground plane or any other part of the scene. The video streams are fed to a computer that extracts gaze direction and generates a video of the scene with a cross depicting the moment-to-moment direction of gaze. The eye-tracking system provided image data at a rate of 60 frames/s, which was then recorded on a digital video disc.

Procedure

An experimenter greeted participants in a hall outside the laboratory and explained the procedure. Participants then inserted foam earplugs to control for
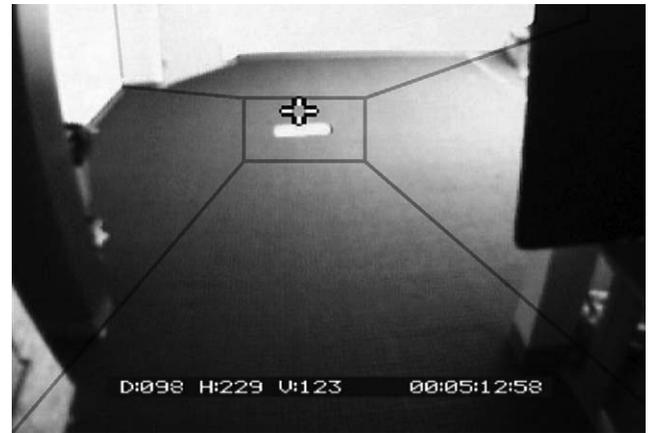

Figure 2. Still image extracted from the scene camera of the head-mounted eye tracker. The cross represents the momentary point of regard on the scene. Target, front, back, and side regions are illustrated with faded lines. These are for the purpose of illustration only; they did not appear in the scene or in the scored video clips.

potential auditory cues and were led into the darkened room with their eyes closed. Once positioned with their back to the stimulus environment, they were equipped with the eye tracker and the calibration procedure was initiated. Because we wished to control the viewing of the stimulus environment, we calibrated the eye tracker by having participants direct their gaze to a series of markers on a poster board approximately 2 m distant rather than using locations in the room as calibration points. After calibration, the participants closed their eyes and were turned and positioned to the observation point for the experiment. At the beginning of each trial, the experimenter engaged the DVD recorder and said, "Eyes open." A poster board obscuring their view of the room was held 12–18 in. from the participant's face. The experimenter then lowered the board to initiate viewing. During the viewing period, a second experimenter monitored the image from the eye camera and took notes on the quality of the eye tracking and the occurrence of blinks and track losses to aid later scoring. The duration of viewing was approximately 5 s and was controlled manually with the aid of the counter on the DVD recorder. The experimenter raised the poster board to terminate viewing, and then said, "Eyes closed." When blind walking was the response, the lights in the room were then turned off and the cable from the eye tracker to the computer was disconnected before the participant began walking. The eye tracker prohibited use of a blindfold, but it was suggested to the participants that the eye image would be used to determine whether they indeed kept their eyes closed. Even if they disregarded instructions and opened their eyes, vision would have been fairly uninformative, as the target was no longer present and the room was darkened (lit only by a dim flashlight

pointed at the participants' back, used by the experimenter to monitor their position as they walked). The participant walked unassisted, and typically veering was minimal. In rare instances in which veering brought participants close to a wall, they were nudged back into the straight pathway without providing feedback about the accuracy of their walked distance. A research assistant manually recorded the distance with respect to a measuring tape that extended along the side of the room and mostly out of the participant's field of view. When verbal report was the response, a verbal estimate was reported out loud and manually recorded. The verbal response was given in feet or meters, whichever metric was most familiar to the participant. At the end of the session, a calibration-check procedure was initiated. All four targets were placed on the floor and a video clip was recorded as participants were instructed to look at each of the targets in turn as well as at various points of interest in the room (such as the doorknob and the edge where the floor and back wall met). This video clip supported manual scoring of the trial clips.

Results and discussion

Eye-movement data were generated by scoring the videos using OpenSHAPA (www.openshapa.org), a data-analysis tool that allows the scorer to code the onset and offset of fixations and assign each fixation to a region of interest while stepping through frames of video. Each video clip depicted the stimulus environment from the scene camera with a cross representing the point of regard (POR) during scene viewing. The cross subtended approximately 3.29° in the video image. A screen shot from one of the trials is presented in Figure 2. Frames were considered part of a fixation if the POR cross was stable at a given location for at least six consecutive frames (100 ms). Each video was scored independently by two scorers with discrepancies resolved by a third scoring and/or through discussion. Scoring of videos was supported by notes recorded by an experimenter made on the basis of the eye video and by the calibration check recorded at the end of the session. One participant was excluded from analysis because of poor calibration and tracking. Seventeen trials from the retained participants were eliminated (3% of data) because track loss was sustained.

Analysis of gaze behavior was based on a number of metrics used to quantify the preference for looking at various regions within the scene. Within a given viewing episode, how frequently is the point of regard directed to a surface between oneself and the object as opposed to behind the object or to the sides? Given the SSIP framework, it is certainly of interest to know

whether or not observers demonstrate a preference for scanning behavior. In the present approach, scanning was operationalized in terms of more basic measures (described later)—eye-tracking videos were not directly scored as evidence for scanning. It should further be noted that the SSIP theory places greater weight on the need to scan than the need to simply fixate the nearby ground surface. For example, B. Wu et al. (2004) found a benefit for near-to-far scanning but not for far-to-near scanning when the field of view was constrained. Given that the same regions were presumably fixated in these conditions, order of selection appears to be of primary importance. However, several experiments from our laboratory suggest that visual information extracted from one viewing episode can readily be maintained in service of distance judgments on subsequent episodes when the viewing condition is more degraded (Gajewski et al., 2010; Gajewski et al., 2013). It is not clear why a role for memory should be specific to indoor environments. If a representation of the greater space indeed contributes to the scaling of perceived distance, one possibility is that the information maintained in memory for that purpose is not available or salient outdoors and/or because the scale of room-sized spaces varies more meaningfully than the scale of outdoor spaces. That is, the relatively large scale of outdoor spaces may not constrain the perceived distance of targets in the intermediate distance range, and in that case, information about scale is less useful and perhaps not maintained in memory. In any case, given a robust role for memory in our indoor task environment, the order of fixation should be less critical. For this reason, our analyses examined the regional locations and durations of fixations as well as the occurrence of scanning.

Regions of interest for scoring were defined with respect to the target and thus varied with target distance (see Figure 2). A fixation was judged to be within the target region if any portion of the cross was in contact with the object. The placement of boundaries separating the side regions from the front and back regions was somewhat arbitrary, as these are admittedly not distinct. Theoretical interest in fixations within the front region was driven by the idea that information from the nearby ground surface ought to support a representation of the surface extending forward and between the observer and the object (which varied in distance across trials). It would thus be inappropriate to simply divide the scene into static front and back regions. Fixations on or very near the boundaries were infrequent, particularly in the front. When these occurred, they were scored as fixations within the front region.

Because the viewing was controlled manually, the duration of viewing varied somewhat across participants and trials. The actual viewing durations were

	Region			
	Target	Front	Sides	Back
Proportion entered	.99 (01)	.51 (.07)	.23 (.06)	.32 (.05)
Total fixation time (ms)	3364 (245)	862 (149)	155 (48)	135 (28)
Entry count	1.85 (0.13)	0.74 (0.12)	0.37 (0.11)	0.38 (0.07)
Fixations per entry	1.61 (0.14)	1.81 (0.16)	1.07 (0.03)	1.06 (0.02)

Table 1. Summary of gaze-behavior measures (means and standard errors) for Experiment 1.

extracted from the video and defined as the time elapsed from the moment the target was revealed until the moment it was obscured by the card. The mean viewing duration was somewhat greater in the verbal condition than it was when blind walking was the response (5856 and 5479 ms, respectively), though the difference did not reach the level of significance ($p = 0.08$). The initial fixation was also uncontrolled, and gaze behavior at trial onset varied across participants. Observers were instructed to open their eyes prior to onset of viewing, which was controlled by the lowering of a large poster board. Participants often tracked their eyes with the card as it was lowered (i.e., they engaged in smooth pursuit of the falling edge of the card), and as a result their first fixation on the scene itself was the target of a saccade and a complete fixation. About as often, participants held their gaze steady as the card was lowered. If gaze was steady in that location for at least six frames (100 ms) from the onset of the trial, it was counted as the initial fixation and with an onset that coincided with the onset of viewing. In these cases, the fixation duration reflected only the portion of time that the scene was visible.

Initial fixation and target priority

In the limited-viewing-time paradigm (Gajewski et al., 2010), performance was stable over two consecutive blocks of 100-ms trials, presumably because the target was always prioritized and the viewing duration never afforded the selection of additional useful information. Thus, of interest here was the degree to which the target would be prioritized for overt attentional selection. Do observers look to the object first? The initial fixation was on the target 74% of the time, either because the participant was looking in the direction of the target object as the card was lowered or because the target object was the target of the first saccade from the card when it was being tracked. The initial fixation was within the front, side, and back regions 4%, 2%, and 21% of the time, respectively. In these cases, that target object was selected as the target of the first saccade 83% of the time. Overall, then, the target object could be considered prioritized 96% of the time.

Gaze-behavior analyses

Proportion entered

There are several ways to quantify where people look and for how long. We begin with the proportion of trials in which each of the regions was entered. That is, on average, how often do the eyes land within each of the regions at least once during a given trial? As already indicated, the target object was clearly prioritized. It is thus not surprising that the target region was entered at least once 99% of the time on average. In contrast, the front region was entered at least once only 51% of the time. The front region was entered more frequently than the side and the back regions (ps < 0.05, see Table 1), but 49% of the time participants never even looked in the front region. This outcome does not suggest a strong preference for a strategy that includes overt attentional selection of the nearby ground surface. There were no differences between blind walking and verbal report in the proportions of entries for any of the regions (all ps > 0.22).

Total fixation time

The second measure is the total amount of time spent fixating in each of the regions. There were no differences between blind walking and verbal report in total fixation time for any of the regions (all ps > 0.13). The average amount of time spent fixating the back and sides was exceedingly small (see Table 1), and time spent fixating the object was nearly four times that of the front region, $t(26) = 6.82$, $p < 0.001$. However, it should be noted that this measure includes zeros (i.e., durations were averaged over trials where the region was never fixated). The relatively small amount of time spent fixating the front region could be expected, given that observers looked there less frequently (and more so for the back and side regions). For the cases where both regions were fixated (i.e., fixation time conditionalized), the total time spent on the target ($M = 2636$ ms) was still greater than it was in the front region ($M = 1508$ ms), $t(24) = 4.28$, $p < 0.001$. These first two measures are relatively coarse but provide a compelling window to the overall priority assigned to each of the regions by the observers.

Entry count

The third measure is the average number of times observers looked within each of the regions during the course of a given trial (Table 1). This measure quantifies the visual exploration of the scene. How frequently do observers look from one region to another? Observers looked to the target region more frequently than any of the other three regions (ps < 0.001) and to the front more often than either the back or the sides (ps < 0.05). The average total number of entries across all regions was 3.34. Overall, the entry-count measure suggests a tendency to look from the target region to one or two other regions, mostly the front, and then back to the target region. However, as with fixation time, this measure averages over trials where the region was never entered. Again, of interest was the number of entries to the target and front regions when both were entered at least once. In this case, the number of entries for the target and front regions did not differ ($p = 0.41$). Observers looked to each of these regions 2.25 times on average, suggesting that when observers did look to the front region, they tended to look back and forth between regions. There were no differences between blind walking and verbal report in the number of entries for any of the regions (all ps > 0.15).

Fixations per entry

The fourth measure is the average number of fixations made within a region each time it is entered. This measure also quantifies visual exploration, but within a given region. On this measure, which by definition excludes trials where the region was never entered, the target and front regions did not differ ($p = 0.22$), and both were greater than the back and sides (ps < 0.01). There were no differences between blind walking and verbal report in the number of fixations per entry for any of the regions (all ps > 0.11). The number of fixations on the target may seem surprising, given that the size of the region is quite small, particularly relative to all other regions. However, the average time per entry on the target was nearly 2 s, which is an extraordinarily long time to remain in fixation. Frequently, observers' gaze would drift across the object and then a saccade would be made to correct for the drift. Often, observers shifted their gaze left and right from one edge of the object to the other, as though sweeping out the visual angle of the target. In contrast, a high number of fixations per entry in the front region would be expected if participants were scanning the foreground. The number of fixations per entry in this region ranged from one to eight but was greater than one only 56% of the time on average. Given that the front region was entered only 51% of the

time, the most liberal estimate of overall scanning frequency would be about 29%.

Eye-movement strategies

Based on the basic measures just discussed, we categorized each trial into one of three mutually exclusive eye-movement strategies based on regional preference. These are admittedly coarse-grained, but they provide a way of characterizing the selection of eye-movement strategies on a trial-by-trial basis. Does the selection of strategy change over the course of the session or does it depend more on distance? Alternatively, the preferred strategy may be idiosyncratic to the observer but very stable. We begin with the *steady-fixation* strategy (Steady), which was selected surprisingly often—about 36% of the time on average. A Steady trial occurred if participants directed their gaze immediately to the target and spent the remaining viewing time in that region. Next we consider the *target-and-front-only* strategy (Front), which was selected about 31% of the time on average. A Front trial occurred if the participant devoted viewing time to the target and front regions only. Because scanning was never strictly in the direction of near-to-far, we did not subcategorize the Front strategy. Finally, the *back-and-sides* strategy (Back-and-Sides), which was selected about 33% of the time on average, occurred if participants directed their gaze to either the back or one of the side regions after the initial fixation. This strategy can include fixations to the front region but represents the more global viewing strategy that we expected if observers were inclined to build up a representation of the greater space. The frequencies for none of the strategies depended on response mode (all ps > 0.61).

The selection of strategy was fairly stable across participants. The favored strategy was selected 73% of the time on average and ranged from 44% to 100%. Figure 3 shows the proportions for each of the strategies by distance and by trial number. Trends with distance and trial number were examined by generating multilevel logistic growth models. Because the strategy frequencies are not independent, these were generated for each strategy separately. Overall, the selection of strategy did not depend on distance (all ps > 0.44). It is especially interesting to note that the frequency of the Front strategy was not greater for the far distance. The selection of strategy did depend on trial number, however. The frequency of the Front strategy was initially relatively high. While the Steady and Back-and-Sides strategies increased modestly, these trends did not reach the level of statistical significance (ps > 0.14). The proportions of Front-strategy use declined over the course of the session, $t(26) = -2.56$, $p < 0.05$. Competing interpretations of this outcome can be considered. One possibility is that participants changed
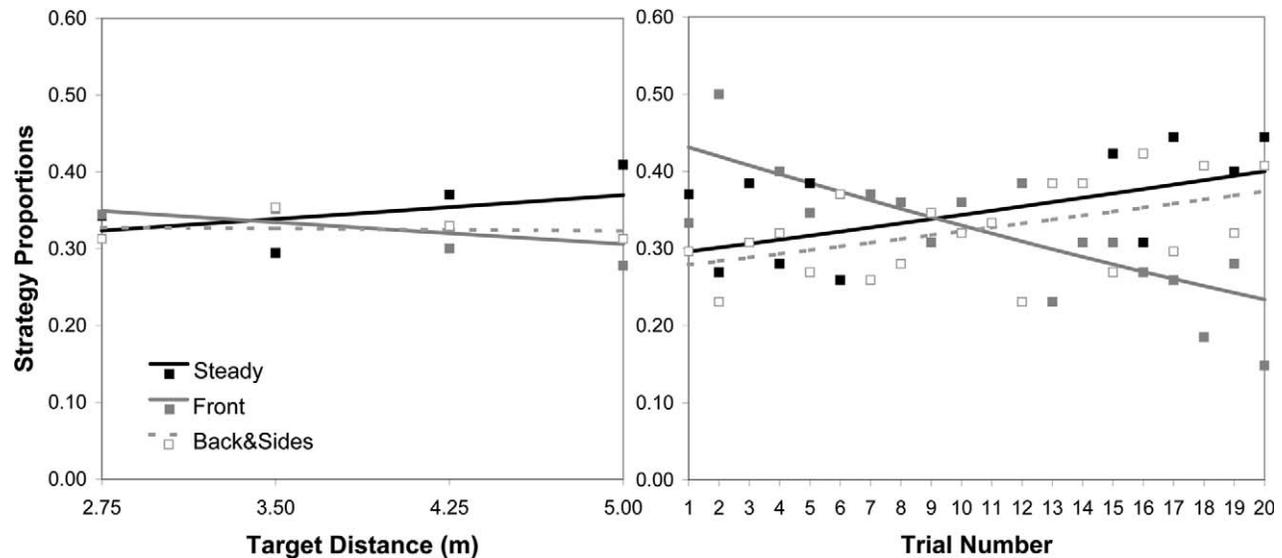
Figure 3. Gaze-strategy frequencies in Experiment 1 as a function of distance (left) and trial number (right). Depicted are the proportions of trials in which the strategy was selected, along with lines representing the fixed effects derived from multilevel logistic growth models.

their gaze strategy because the information needed was already available in memory from the earlier trials, a possibility consistent with the idea that the near ground surface is the locus for the important sources of information as well as with the block-order effects we have observed in previous studies. Another possibility, though, is that participants adopt a more effective strategy over time. That is, they learn that it is more useful to hold their gaze steady or direct their gaze at the back and sides. To discern between these possibilities, we examine performance as a function of gaze strategy (see below).

Performance analyses

Overall accuracy

Accuracy was examined in terms of response sensitivity (or simply sensitivity, hereafter) and bias. *Sensitivity* is the degree to which response distance differs systematically with differences in the distance of the target. *Bias* represents the overall tendency toward under- or overestimation. We employed a mixed (multilevel) modeling approach because it is well suited for designs that include a continuous independent variable (distance) and repeated observations from participants. Models reported included distance and intercepts as random factors. The parameter estimates for the fixed effects of distance (slopes relating response distance to target distance) correspond to sensitivity. Differences in sensitivity are apparent when a variable has an interactive effect with distance. Differences in bias can be examined by comparing mean responses

across distances in a separate model with intercept as the only random factor (see Experiment 2). However, of interest here was whether performance on a trial-by-trial basis would be predicted by the gaze strategy employed. For this reason, it was preferred to examine bias and sensitivity within the same model. To satisfy this goal, target distance was centered on the mean target distance, which aligned the intercept with the overall mean response. Differences in bias are thus apparent when a variable has an effect on the intercept. Bias is reported as a percentage of the mean target distance. In the text, F tests are reported for main effects and interactions; t tests are reported for contrasts. We begin with an overall performance analysis and follow with an analysis of performance as a function of gaze strategy.

Sensitivity was generally high (slope $= 1.17$) and did not depend on response mode ($F < 1$). The underestimation bias was greater with blind walking (-12%) than with verbal report (-1%), though this difference was not significant ($p = 0.22$) and was driven by one participant whose verbal estimates were given in meters and were, on average, nearly twice as great as the target distance. Because the bias and sensitivity for this participant were each greater than two standard deviations above the mean for that group, we excluded these data from subsequent analysis. An analysis of signed errors as a function of response mode and trial number did not suggest improvement over the course of the session. There was neither an effect of trial number nor a response-mode-by-trial-number interaction ($ps > 0.28$).

Accuracy as a function of gaze strategy

While gaze behavior did not depend on response mode, it was mostly idiosyncratic to the observer and could nevertheless play a role in performance. A primary objective was to determine whether accuracy depended in any way on the gaze strategy deployed on any given trial. Gaze strategy (Steady, Front, and Back-and-Sides) was included as a level 1 predictor for response distance in a model along with response mode and distance. That is, strategy was introduced as a variable with freedom to change on a trial-by-trial basis. Compelling main effects of strategy on sensitivity, $F(2, 444) = 5.72$, $p < 0.01$, and bias, $F(2, 444) = 6.85$, $p < 0.01$, were observed and are depicted in Figure 4. Because the effects of strategy did not depend on response mode (all $ps > 0.08$), contrasts between strategies were examined in a model that excluded the response-mode variable and its interaction terms. These show greater sensitivity for Back-and-Sides relative to Front, $t(448) = -2.96$, $p < 0.01$, but no other effects on sensitivity ($p > 0.14$). However, underestimation bias was greater for Front (bias $= -13\%$) than it was for Steady (bias $= -8\%$), $t(448) = 2.62$, $p < 0.01$, or Back-and-Sides (bias $= -8\%$), $t(448) = -3.00$, $p < 0.01$. This analysis does not suggest a benefit associated with preferential fixating of the front region, and it certainly does not suggest a cost associated with the Steady gaze strategy.

In sum, Experiment 1 provided several interesting and perhaps surprising outcomes. Most notable was the prevalence of the strategy of holding gaze on the target object. While the front region was *fixated* more frequently than the back and sides, *scanning* the front region was not even as frequent as the Steady gaze strategy. If an accurate representation of distance depends on minimizing slant error, extracting information about the near surface would be more important at further distances, since the effect of slant error would become compounded. Instead, there was a trend toward decreasing front fixations as distance increased. Finally, there was no indication that observers adopted different strategies depending on response mode. Side entries were numerically more frequent and front entries were numerically less frequent with verbal response than with blind walking, but neither approached the level of statistical significance. More striking was the fact that observers adopted varied strategies and that these were reasonably stable within participants.

The prevalence of the Steady gaze strategy, coupled with the fact that performance was unimpaired by a lack of eye movements, strongly argues against the idea that overt shifts of attention to various regions are critical for an accurate judgment of distance. While it is impossible to determine given the current exploratory design, this outcome at least suggests that observers adopt a covert attentional strategy of some kind. Previous work strongly suggests that angular declination is extracted quickly and that there is a benefit associated with
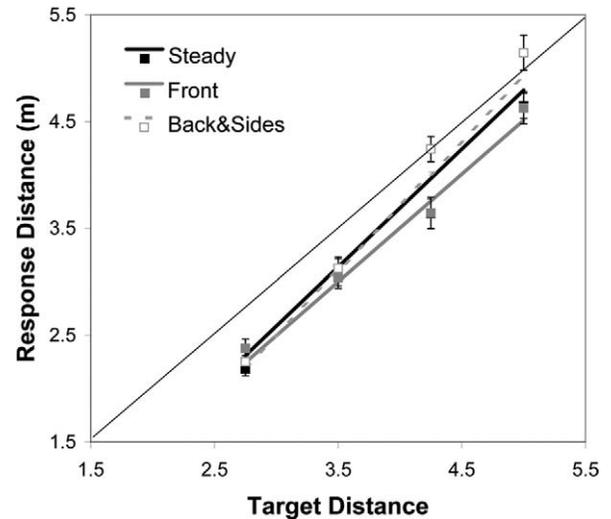

Figure 4. Response distance in Experiment 1 is shown as a function of target distance by gaze strategy. Depicted are the means with standard error bars and lines representing the fixed effects derived from a multilevel (mixed) model.

providing time to extract additional sources of useful information (Gajewski et al., 2010; Gajewski et al., 2013). If observers truly focus attention on the target region even when viewing affords exploration, it is unclear what could be gained by extended viewing time. There is at least some suggestion in the literature that binocular cues have a relatively slow time course (e.g., McKee, Levi, & Browne, 1990). However, binocular parallax (the stimulus to convergence) is not highly reliable in the intermediate distance range (Cutting & Vishton, 1995), and performance has been shown not to depend on binocular viewing for targets resting on the ground (Bian & Andersen, 2013; Philbeck & Loomis, 1997). One possibility is that extended viewing affords better processing of the ground surface immediately in front of and behind the target. This could aid in the determination of optical slant (the angle of gaze relative to the ground plane when the target is fixated), which has been posited to play a role in distance judgments (Durgin & Li, 2011; Li & Durgin, 2012) and would obviate the need to posit a role for covert attention. Experiment 2 examined the role for eye movements directly by imposing a gaze strategy rather than leaving it up to the observer.

Experiment 2

Method

Participants

Eighteen students from the George Washington University community participated in exchange for

course credit. All participants had normal or corrected-to-normal vision and were unaware of the aims of the study.

Stimuli and design

The stimuli (targets and room environment) were the same as in Experiment 1. Participants were randomly assigned to one of two eye-movement strategies (Free View or Steady Gaze). Each target distance was employed five times, with all distances sampled without replacement prior to repetition. That is, the experiment effectively was run in five 4-trial blocks, though they were not punctuated by breaks. This design was adopted so that sensitivity and bias could be readily examined over the course of the session.

Apparatus

The apparatus was the same as Experiment 1. Participants were set up with the head-mounted eye tracker and the experimenter monitored the output video during the session. The aim here was to examine the general utility of eye movements by setting up a contrast with the Steady Gaze strategy rather than to explore the selection of strategies when observers are free to move gaze. The eye tracker was thus used to assess compliance with the assigned eye-movement strategy. These image data were not recorded or further analyzed.

Procedure

The procedure was the same as in Experiment 1, except that an eye-movement strategy was imposed on the participant. One group of participants was told that the best strategy was to explore the scene by moving their eyes to various locations in the room. The second group was told that the best strategy was to maintain a steady gaze on the object. Participants in each group were instructed to adopt these strategies, and this was verified by the experimenter. Because there were no dependencies on response mode in Experiment 1, we opted to use blind walking as the response mode for all distance judgments in Experiment 2.

Results and discussion

Accuracy was assessed with the same mixed-model approach employed in Experiment 1. However, because there were no trial-by-trial predictors, differences in bias were assessed on the overall means and the means by block rather than by comparing intercepts on centered data. In addition, we include a metric for precision, the standard error of the estimate (SEE) for
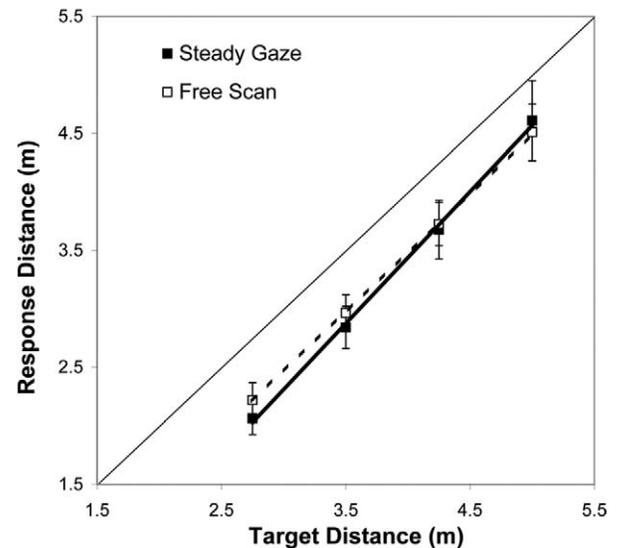

Figure 5. Response distance in Experiment 2 is shown as a function of target distance by gaze strategy. Depicted are the means with standard error bars and lines representing the fixed effects derived from a multilevel (mixed) model.

the best-fitting lines relating distance estimates to target distance. We recognize that this metric is not equivalent to the more commonly used measure of variable errors—the normalized standard deviations for each participant and condition. However, here we were interested in changes that might occur across blocks where the distances were not repeated. Our data fit well to a linear regression model, and, given those assumptions, we argue that the SEE is a reasonable proxy for variable error.

While marked by a bias towards underestimation, sensitivity was high in Experiment 2 and did not depend on viewing strategy (Figure 5). Viewing strategy had no overall effect on sensitivity (slope = 1.07), bias (−14%), or precision (SEE = 0.46 m), all $Fs < 1$. We performed one additional analysis to address the possibility that observers might accumulate the critical sources of information early in viewing—that is, reaching peak performance very early, thereby masking any linkage between eye movements and performance when analyzed across multiple trials. We have previously shown that prior viewing experience can facilitate subsequent performance when viewing time is limited (Gajewski et al., 2010). In the present case, the largest potential difference between viewing conditions might be expected early in the session. To determine whether there were performance differences early in the session that might have diminished over time, we ran models with block order as a continuous predictor. Block number was coded 0–4 so that the intercept reflected performance on the first block. In these models, the intercepts and the effects of block number were included as random effects. These analyses revealed no

differences between viewing conditions initially or over the course of the session. Viewing strategy had no effect on sensitivity, bias, or precision in the first block (all $ps > 0.16$), and the effect of viewing condition on these did not depend on block number (all $ps > 0.29$).

The results of Experiments 1 and 2 strongly converge on the idea that overt selection of the ground plane or any other region of the room is not required for an accurate judgment of distance. However, the experiments do leave open the question of whether there is important information within these regions that might be extracted covertly. This could be accomplished by holding the gaze steady and shifting attention covertly or by dispersing attention widely across the visual field. Given the importance of the ground plane in theories of distance perception generally (e.g., Gibson, 1950, 1979; Sedgwick, 1986) and in the SSIP theory specifically (e.g., B. Wu et al., 2004), it would be surprising if judgments were truly based only on the processing of information local to the object. If observers are adopting a covert attentional strategy, gaze behavior loses its ability to index the regions of the scene that are informative in the task environment. Controlling the visibility of select regions provides an alternate means of addressing this issue. Experiment 3 thus compared performance with the near ground surface obstructed by a blocker to performance when a full view of the scene was afforded. It was similar in spirit to the B. Wu et al. (2004) study where effects of occlusion were observed. Here we tested participants in an indoor environment and with block order included as a variable of interest to examine for possible effects of visual familiarity with the space.

Experiment 3

Method

Participants

Twenty-eight students from the George Washington University community participated in exchange for course credit. All participants had normal or corrected-to-normal vision and were unaware of the aims of the study.

Stimuli

Targets were sheets of yellow foam placed on the floor at one of four distances (3, 4, 5, and 6 m) with angular size held constant, as in Experiments 1 and 2. Experiment 3 was conducted in a different lab environment, which extended to 9.54 m from the observation point. The stimulus portion of the room was entirely empty. There was a discontinuity in the wall about 4 m out on the left side, but there were no doors or shelves in the room. The participants' view to the near right (out to 1.5 m) was obstructed by a barrier that was used to control lighting and to provide the research assistants an object to hide behind during scene viewing. Targets were salient against a bluish-gray carpet with a fine-grained mottled texture.

Apparatus, design, and procedure

There were two viewing conditions administered in blocks, with block order counterbalanced and included as a variable of interest. Participants viewed targets with a large poster board obstructing their view of the foreground (Blocker) or with no obstruction (Full View). The blocker was attached to a height-adjustable stand and was set so that approximately 3° of the ground in front of the target remained visible. That is, the height of the blocker varied with distance. To keep eye level and viewing position constant across trials, participants stood with their chin in a chin rest. Prior to the beginning of the session, the experimenters measured the participants' chin and eye heights and calibrated the height of the blocker accordingly. Once the experiment was complete, a calibration check was executed to determine the actual proximity of the blocker to the front edge of the target. The observed gap between the target and the blocker was approximately 1.7° on average.

In both viewing conditions, participants were instructed to maintain a steady gaze on the target during the viewing period. Eye movements were not monitored with the eye tracker, but large gaze shifts away from the object, such as to the foreground or to the side wall, could be readily detected by the experimenter. While we wished to hold gaze strategy constant across conditions, the focus was on controlling the sources of information available to the observer. When eye movements were detected, the experimenter reminded the participant to hold gaze steady. These occurrences were rare, and the data from these infrequent trials were retained. Each target distance was employed twice, with all distances sampled without replacement prior to repetition. That is, the experiment was run in two blocks, one for each viewing condition, and each block was subdivided into two sets of trials. Blind walking was the response mode for all distance judgments. The viewing duration was approximately 5 s and was controlled manually as in Experiments 1 and 2.

Results and discussion

As can be seen in Figure 6, there was a modest difference between viewing conditions that depended on block order. Response sensitivity was generally high, though there was a marginal interaction of viewing
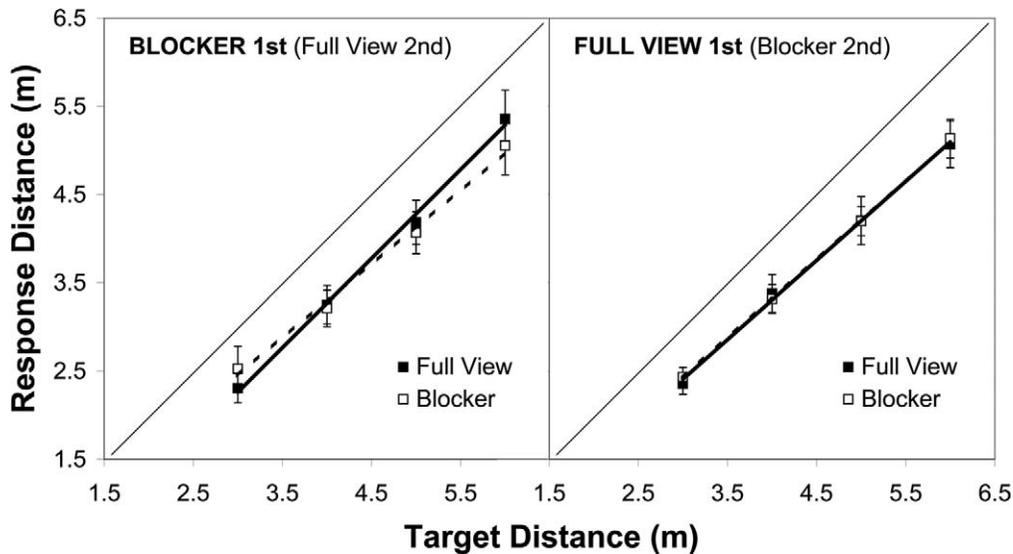
Figure 6. Response distance in Experiment 3 is shown as a function of target distance by viewing condition (Full View vs. Blocker). Performance when the Blocker trials were administered first is shown on the left; performance when the Full View trials were administered first is shown on the right. Depicted are the means with standard error bars and lines representing the fixed effects derived from a multilevel (mixed) model.

condition and block order, $F(1, 380) = 3.38$, $p = 0.07$. When the blocker condition was administered first, sensitivity was greater in the Full View condition (slope $= 1.01$) than in the Blocker condition (slope $= 0.85$), $t(380) = 2.56$, $p < 0.05$. When the Full View condition was administered first, there was no effect of viewing condition (slope $= 0.90$), $p = 0.97$. Repetition number had no effect on sensitivity (all $ps > 0.20$). Bias was primarily not affected by viewing condition or block. However, there was marginal interaction between viewing condition, block order, and repetition number, $F(1, 78) = 3.78$, $p = 0.06$. When the Full View condition was administered in the first block, there was a trend toward a reduction in bias across repetitions (-19% and -14% for the first and second repetitions, respectively), $t(78) = 1.90$, $p = 0.06$. Bias was otherwise not affected by viewing condition, block order, or repetition number ($M = -17\%$, all $ps > 0.36$). This pattern, while admittedly modest, is consistent with the idea that information about the near ground surface may be available but not immediately extracted when a full view is afforded, perhaps because gaze is held steady. In contrast, when the view is occluded, observers never gain access to this information and therefore never improve. Precision was greater in the second block (SEE $= 0.33$ m) than the first block (SEE $= 0.45$ m), regardless of viewing condition, $F(1, 78) = 6.13$, $p < 0.05$.

While there does appear to be some cost associated with obscuring the near ground surface, the real-world magnitude of this cost is quite small. Indeed, the differences we have previously observed between brief- and extended-viewing conditions were more substantial (Gajewski et al., 2010). This outcome raises the possibility

that the near ground surface may be less important indoors when other surfaces are available, such as the walls and/or ceiling. We have previously found that a visual preview of the room is sufficient to support performance when viewing time is limited (Gajewski et al., 2013). One possibility following from this work is that, in addition to supporting a representation of the ground surface, information extracted about the walls and ceiling could support a better representation of the size and shape of the room and/or a better sense of the scale of the space, which in turn could influence judgments of target distance. Experiment 4 examined the role for the other surfaces by comparing performance when the target was viewed through an aperture that obstructs the walls, floor, and nearby ground surface to performance when a full view of the scene was afforded. Experiment 4 closely parallels the configuration employed by B. Wu et al. (2004) in the study where effects of occlusion were observed. Again, however, here we tested participants in an indoor environment (see also Creem-Regehr, Willemsen, Gooch, & Thompson, 2005) and with block order included as a variable of interest to examine for familiarity effects.

Experiment 4

Method

Participants

Twenty-eight students from the George Washington University community participated in exchange for course credit. All participants had normal or corrected-
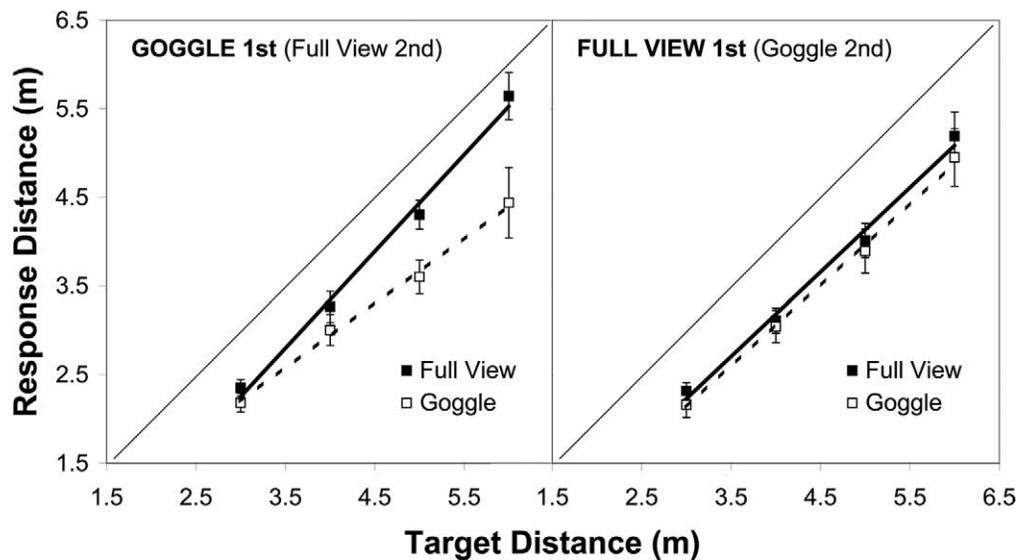
Figure 7. Response distance in Experiment 4 is shown as a function of target distance by viewing condition (full view vs. with goggle). Performance when the goggle trials were administered first is shown on the left; performance when the full-view trials were administered first is shown on the right. Depicted are the means with standard error bars and lines representing the fixed effects derived from a multilevel (mixed) model.

to-normal vision and were unaware of the aims of the study.

Stimuli

The stimuli and task environment were the same as in Experiment 3.

Apparatus, design, and procedure

The design was the same as in Experiment 3 and the procedure was similar. In the obscured-viewing condition here, participants viewed the stimuli through an aperture that occluded the side walls and ceiling in addition to the near ground surface. The aperture was created by masking up a set of clear laboratory goggles with electrical tape. A small opening allowed a view of the scene that subtended a visual angle of approximately $14.5° \times 17.2°$ (high $\times$ wide). Because the required alignment of the two apertures would vary between participants, we opted to run the experiment monocularly in both conditions. A head-angle calibration procedure was employed to ensure that participants began each trial with the target in view. This procedure also ensured that the view of the scene was stable across trials. Outside the laboratory before the experiment began, a marker was placed on a stick (an 8-ft. 1×2 positioned 0.5 m in front of the participant). Based on the participant's eye height, the marker was positioned so that it would correspond to the gaze angle centered for the range of object distances employed. The participant then adjusted the goggles and their head angle to center the marker. At the

beginning of each trial, the participant was instructed to open their eyes. With the poster board obstructing their view, they were told to find the marker on the stick and then close their eyes but hold their head steady. The experimenter then put the stick to the side and the trial proceeded as it did in all previous experiments. The design and procedure were otherwise the same as in Experiment 3.

Results and discussion

As can be seen in Figure 7, the effect of the occluder here was compelling. There was an effect of viewing condition on sensitivity that depended on block order, $F(1, 380) = 11.25$, $p < 0.001$. When the goggle condition was administered first, sensitivity was greater in the full-view condition (slope = 1.09) than it was in the goggle condition (slope = 0.74), $t(380) = 5.19$, $p < 0.001$. When the full-view condition was administered first, there was no effect of viewing condition on sensitivity (slope = 0.94), $p = 0.66$. Sensitivity increased across repetitions, $F(1, 380) = 4.41$, $p < 0.05$, but there were no interactive effects of repetition number on sensitivity (all $ps > 0.22$). The effect of viewing condition on bias also depended on block order, $F(1, 78) = 7.76$, $p < 0.01$. When the goggle condition was administered first, bias was greater in the goggle viewing condition (−27%) than it was in the full-view condition (−14%), $t(78) = 5.24$, $p < 0.001$. When the full-view condition was administered first, there was no effect on bias (−20%), $p = 0.20$. There were no effects of repetition number on bias and no interactions (all $ps >$

0.33). There were no effects on precision (SEE = 0.45 m), all ps > 0.15.

The pattern of results here argues strongly against the idea that localization in the steady-gaze condition is based on information local to the object. While viewing the scene through the aperture does occlude more of the scene than the blocker did in Experiment 3, the object and the local ground surface were fully available, even when seen through the aperture. This outcome suggests that participants indeed extract information about the greater space covertly while holding their gaze steady. Interestingly, the block-order effect suggests that the information extracted from the full-view condition, presumably information about the size of the room or scale of the space, can be maintained in memory to support judgments of distance when these cues are occluded. This aspect of the data is consistent with the patterns we have observed when manipulating viewing duration (Gajewski et al., 2010; Gajewski et al., 2013) and point to a critical need to control for the visual experience with a block design.

The pattern of results observed here contrasts with that observed in Experiment 3, suggesting that nearby ground surface is less important indoors when other surfaces are available, such as the walls and/or ceiling. To be fair, viewing was monocular in Experiment 4 but not in Experiment 3. If binocular cues were strong in this context, they certainly could have mitigated the effect of the blocker, though previous studies do little to suggest a contribution of binocular cues in the manipulated distance range and when targets are at floor level (e.g., Bian & Andersen, 2013; Philbeck & Loomis, 1997). In addition, B. Wu et al. (2004) found a benefit of near-to-far scanning with a constrained field of view. The role for scanning may be more potent when the viewing conditions are more degraded. In Experiment 5, we directly compared near-to-far scanning and steady-gaze conditions with the field of view in both conditions limited by the aperture. If the benefit for scanning is more pronounced in this configuration, there should be a compelling performance advantage for the near-to-far condition. Indeed, in the B. Wu et al. study, performance with occluded near-to-far scanning was as good as it was with an unobstructed view of the scene.

Experiment 5

Method

Participants

Twenty-eight students from the George Washington University community participated in exchange for course credit. All participants had normal or corrected-to-normal vision and were unaware of the aims of the study.

Stimuli and apparatus

The stimuli, apparatus, and task environment were the same as in Experiment 4, except that large black poster boards were placed as blockers to crop participants' view from the top and sides, with the top blocker adjusted according to each participant's eye height. This ensured that these regions were equally obstructed in both obstructed-viewing conditions.

Design and procedure

The design was the same as in Experiment 4. Two viewing conditions (goggle-steady and goggle-scan) were run in blocks, with block order counterbalanced. Each block comprised two sets of four trials using four distances. As a reference for performance, we included a full-view condition. Because information extracted from the full-view condition was expected to eliminate the deleterious effects of subsequent obstructed-viewing conditions, this block of trials was run last for all participants. The goggle-steady condition was the same as in Experiment 4. In the goggle-scan condition, participants began the trial with their head tilted down and gaze directed at the ground directly in front of their feet. Because their view was constrained by the aperture in the goggle, scanning was accomplished by tilting the head upward. They scanned up until their head was in the horizontal position and then closed their eyes and repeated. B. Wu et al. (2004) found similar effects for one and two scans; we wished to maximize the probability of finding an effect in the present study and so allowed our participants to make two scans. Scanning speed was approximately 3 s per sweep and was demonstrated for the participant in advance. Participants were instructed to scan at a steady pace through the object and up to the horizontal position. In the goggle-steady condition, participants were afforded a view of the target and only the immediately surrounding ground surface. In the goggle-scan condition, participants were afforded a view of the ground from themselves to the object.

Results and discussion

As can be seen in Figure 8, Experiment 5 provided no indication that near-to-far scanning is beneficial in the present task environment. There was an effect of viewing condition on sensitivity, $F(1, 596) = 41.67$, $p < 0.001$, but no other effects or interactions (all ps > 0.14). Sensitivity was greater in the full-view condition (slope = 0.94) than in either of the two obstructed-
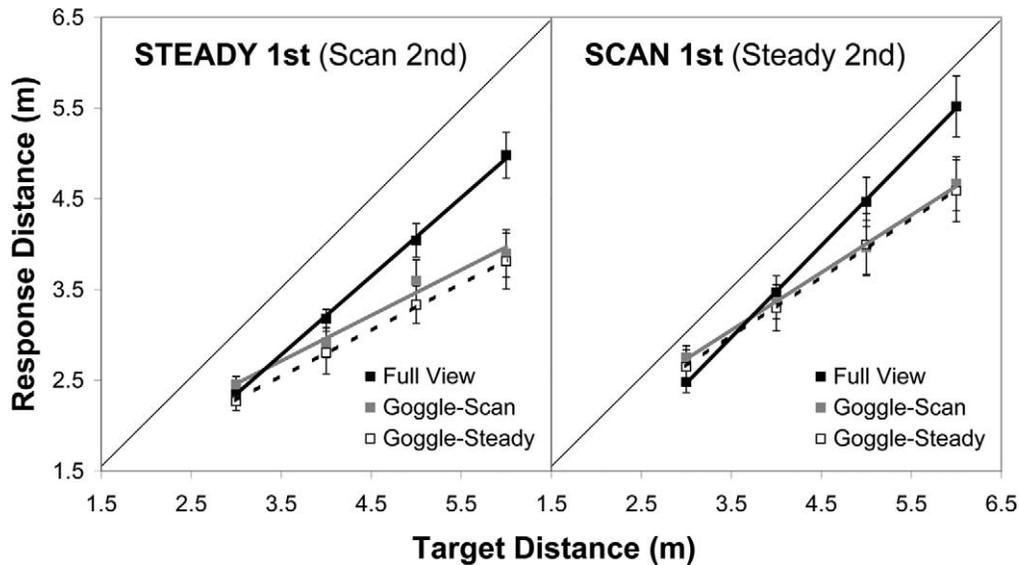
Figure 8. Response distance in Experiment 5 is shown as a function of target distance by viewing condition (full view vs. steady with goggle and scan with goggle). Performance when the goggle-steady trials were administered first is shown on the left; performance when the goggle-scan trials were administered first is shown on the right. Depicted are the means with standard error bars and lines representing the fixed effects derived from a multilevel (mixed) model.

viewing conditions ($ps < 0.001$; slopes for the goggle-steady and goggle-scan conditions were 0.58 and 0.57, respectively). Sensitivity did not differ between the two goggle conditions, $p = 0.72$. Similarly, there was an effect of viewing condition on bias, $F(1, 130) = 23.72$, $p < 0.001$, but no other effects or interactions (all $ps > 0.08$). Underestimation bias was less pronounced in the full-view condition (bias $= -15\%$) than in either of the two obstructed-viewing conditions ($ps < 0.001$; biases for the goggle-steady and the goggle-scan conditions were -26% and -23%, respectively). Bias did not differ between the two goggle conditions, $p = 0.13$. There were no effects on precision (all $ps > 0.11$; mean SEE $= 0.38$ m).

The compelling full-view advantage observed here and in Experiment 4 suggests that the information about the greater visual space is important for localizing floor-level objects, at least indoors and when the floor can be assumed to be flat. The SSIP account suggests that an accurate representation of distance depends on the ground-plane representation, because the intersection of the eye-to-target vector with the surface is needed to compute distance. Experiment 3 showed little impact of occluding the nearby ground plane, and Experiment 5 showed no benefit of making it visible by scanning near to far. Instead, obscuring the walls and ceiling had the greatest impact on performance. This outcome is consistent with the idea that an enhanced representation of the greater space might have a scaling influence on perceived distance. However, it should be noted that an important role for the ground plane finds continued support in the present data if it is assumed that the edges formed by the

meeting of the floor and walls enhance one's representation of ground-surface slant. There is little or no indication that features of the nearby ground plane are critically extracted in the present indoor task environment.

General discussion

The present study was framed around a fundamental question: Where do observers look to judge the distance to an object? While there was an exploratory element to our initial approach, there were several theoretical bases of interest. First, much theoretical weight has been placed on the idea that information about the ground surface plays a crucial role in perceiving the distance of objects (e.g., Bian et al., 2005; Gibson, 1950; B. Wu et al., 2004), yet the role for eye movements in the extraction of this information has until now been entirely unknown. J. Wu et al. (2008) provided data suggesting that selective attention to the nearby ground surface plays a role in performance, but it was unclear whether overt or covert attention would be needed. In Experiment 1, observers did show a preference for fixating the space between themselves and the object compared to other locations in the room scene, but participants only looked to this region on about half of the trials on average. Scanning in Experiment 1 was infrequent and never strictly near to far, and near-to-far scanning was not beneficial in Experiment 5. If fixating the nearby ground surface were the optimal strategy, a failure to reliably adopt

this strategy would have been a logical possibility. However, performance was not better when participants exhibited a preference for fixating the front region than when they held their gaze steady on the object. Further, there was no cost of holding gaze steady in Experiment 2, where viewing strategy was manipulated experimentally. If information about the nearby ground surface were as important as suggested by the SSIP theory (He et al., 2004; B. Wu et al., 2004) and more broadly, one would have to suppose that the information was extracted covertly rather than by the execution of eye movements. However, while there was at least some cost associated with obscuring the near ground surface in Experiment 3, the more compelling effect of obstruction was observed in Experiments 4 and 5, which obscured the walls and ceiling as well as the nearby ground surface. The overall pattern of results suggests that information about the nearby ground surface may not be as important as previously supposed, at least with indoor environments where alternate surfaces are visible and can be used to support the perception of distance to an object. What, then, is the additional source of information that observers are extracting, and what are the implications for theories of distance perception?

Durgin and Li (2011; Li & Durgin, 2012) suggest that gaze declination and optical slant are each critical variables in the perception of distance. Optical slant is the orientation of the ground surface relative to the direction of gaze when the target is fixated. Given that the visible ground surface local to the object was never obscured in our studies, differences in perceived optical slant across viewing conditions would not be expected. However, the goggle condition did obscure the visible horizon (i.e., the edge formed where the floor and the back wall meet), and in that condition the observer's head was also likely pitched a bit more downward to ensure the targets were always in view. Rand, Tarampi, Creem-Regehr, and Thompson (2011) have recently suggested that observers might encode angular declination with reference to the visible horizon in smaller scale environments when the actual horizon cannot be seen. Li and Durgin (2009) have provided data suggesting that head orientation (a component of gaze declination) can be proprioceptively exaggerated with a gain factor of about 2. Both of these possibilities are consistent with the idea that observers in our studies benefitted from the extraction of visual information about the surrounding space, because all costs associated with wearing the goggles were eliminated when the full-view condition was administered first. The visible horizon is an element of the surrounding space that was only seen in the full-view condition, and any potential errors associated with one's sense of head orientation would have to have been improved by memory for the visual space from the first block.

An important role for the extraction of information about the surrounding space was also suggested as part of a dynamic framework for distance perception proposed by Gajewski et al. (2013) based on block-order effects in the limited-viewing-time paradigm (see also Gajewski et al., 2010). Viewing durations that afford detection of the floor-level targets (9–24 ms) have proven sufficient to support a sensitive response to distance, presumably because angular declination is a reliable cue that can be quickly extracted. While there was very little benefit observed when the viewing duration was extended up to the time frame of a typical eye fixation (220 ms), performance has been shown to improve more markedly when viewing time is more extended (5000 ms). Critically, this same high level of performance has been observed with limited viewing durations when visual experience is provided in advance, such as a preceding block of extended-viewing trials or even a single 15-s visual preview of the room without a specified target object. The role for visual experience in performance suggests that it is important to have a representational structure in place for integrating new sources of information as they become available. We have argued that an abstract representation of the space, a mental model or situation model (Zwaan & Radvansky, 1998), could serve this purpose. A similar idea has been expressed by Loomis and colleagues (Avraamides, Loomis, Klatzky, & Golledge, 2004; Loomis, Klatzky, Avraamides, Lippa, & Golledge, 2007; Loomis, Lippa, Klatzky, & Golledge, 2002; Loomis & Philbeck, 2008). We advance this view by suggesting that an enhanced representation of the space exerts a scaling influence on the computation of distance based on the otherwise most dominant cue, angular declination.

Given the very strong role for angular declination when viewing time is limited, we assumed that information about the near ground surface would play an equally strong role in the benefit of longer viewing durations. The current study constrains that account, at least for indoor environments. The SSIP places great weight on the idea that an accurate representation of the ground surface depends on the extraction of nearby ground cues; in this view, errors in perceived distance arise because the line of sight to the target does not intersect with an accurate representation of the ground surface. Our results suggest that the ground surface need not play such a crucial role, because obscuring the near ground surface had a very minor effect. Our data are also consistent with the results of Creem-Regehr et al. (2005), who failed to find an effect of obscuring the nearby ground plane but found an effect of limiting the overall field of view, at least when head position was held stationary. Their stimulus environment was a hallway with nearby wall surfaces. Nevertheless, we are not suggesting that the

nearby ground surface plays no role whatsoever. Rather, we are suggesting that the role it plays by minimizing surface slant error is reduced when viewing is indoors and/or when alternate surfaces are available. In an indoor environment, the visible walls and ceiling provide edges that may be more salient cues to surface slant. These alternate surfaces also appear to support a perception of environmental scale and may even be built up by the very same sequential surface integration process. However, representations of these alternate surfaces must mediate the perception of target distance very differently than is suggested for the ground surface in the SSIP account. Their influence must be distributed more globally (i.e., by specifying environmental scale), since these surfaces are irrelevant for deriving geographic slant. Again, the pattern of results here does stand in contrast to those observed outdoors (B. Wu et al., 2004). One possible basis for the discrepancy is that the ground plane is a more powerful cue to environmental scale outdoors because alternate surfaces are farther away. Further research is needed to determine the nature of the difference between indoor and outdoor environments.

Finally, we considered the possibility that gaze behavior might exhibit a degree of task specificity in the current context. If observers overtly select from different regions of the scene depending on response mode, differential performance could be explained in terms of the cues extracted rather than by the behavior goals of the observer per se. There were no compelling effects of task on eye-movement strategy and only a very modest trend towards a performance difference that depended on eye-movement strategy. In particular, we found somewhat more back and side fixations for verbal report and a modest benefit for the Front strategy with blind walking but not with verbal report. It has been suggested that landmarks are more important when verbal report is the response mode (Andre & Rogers, 2006); perhaps if our task environment were more cluttered we would have seen more distinct differences. The more compelling conclusion that arises from the present study on the whole is that the selection of eye-movement strategy makes surprisingly little difference for judgments of target distance.

Keywords: distance perception, angular declination, eye movements

Acknowledgments

Corresponding author: Daniel A. Gajewski.
Email: gajewsk1@gwu.edu.
Address: Department of Psychology, George Washington University, Washington, DC, USA.

References

Andre, J., & Rogers, S. (2006). Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Perception & Psychophysics, 68,* 353–361.

Avraamides, M. N., Loomis, J. M., Klatzky, R. L., & Golledge, R. G. (2004). Functional equivalence of spatial representations derived from vision and language: Evidence from allocentric judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 801–814.

Bian, Z., & Andersen, G. J. (2013). Aging and the perception of egocentric distance. *Psychology and Aging, 28,* 813–825.

Bian, Z., Braunstein, M. L., & Andersen, G. L. (2005). The ground dominance effect in the perception of 3-D layout. *Perception & Psychophysics, 67,* 802–815.

Bian, Z., Braunstein, M. L., & Andersen, G. J. (2006). The ground dominance effect in the perception of relative distance in 3-D scenes is mainly due to characteristics of the ground surface. *Perception & Psychophysics, 68,* 1297–1309.

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

Creem-Regehr, S. H., Willemsen, P., Gooch, A. A., & Thompson, W. B. (2005). The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments. *Perception, 34,* 191–204.

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. J. Rogers (Eds.), *Perception of space and motion: Handbook of perception and cognition* (2nd ed., pp. 69–117). San Diego, CA: Academic Press.

Durgin, F. H., & Li, Z. (2011). Perceptual scale expansion: An efficient angular coding strategy for locomotor space. *Attention, Perception, & Psychophysics, 73,* 1856–1870.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision:*

The psychology of looking and seeing. Oxford, UK: Oxford University Press.

Foley, J. M., & Richards, W. (1972). Effects of voluntary eye movement and convergence on the binocular appreciation of depth. *Perception & Psychophysics, 11,* 423–427.

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.

Gajewski, D. A., Philbeck, J. W., Pothier, S., & Chichka, D. (2010). From the most fleeting of glimpses: On the time course for the extraction of distance information. *Psychological Science, 21,* 1446–1453.

Gajewski, D. A., Philbeck, J. W., Wirtz, P. W., and Chichka, D. (2013). Angular declination and the dynamic perception of egocentric distance. *Journal of Experimental Psychology: Human Perception and Performance*, advance online publication, doi:10.1037/a0034394.

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

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

Hayhoe, M., Gillam, B., Chajka, K., & Vecellio, E. (2008). The role of binocular vision in walking. *Visual Neuroscience, 25,* 1–8.

He, Z. J., Wu, B., Ooi, T. L., Yarbrough, G., & Wu, J. (2004). Judging egocentric distance on the ground: Occlusion and surface integration. *Perception, 33,* 789–806.

Henderson, J. M. (1992). Visual attention and eye movement control during reading and picture viewing. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 260–283). New York: Springer-Verlag.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7,* 498–504.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–293). Oxford, UK: Elsevier Science.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception & Psychophysics, 45,* 196–208.

Hoffman, J. E., & Subramaniam, B. (1995). The role of attention in saccadic eye movements. *Perception & Psychophysics, 57,* 787–795.

Hollands, M. A., Patla, A. E., & Vickers, J. N. (2002). "Look where you're going!": Gaze behaviour associated with maintaining and changing the direction of locomotion. *Experimental Brain Research, 143,* 221–230.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35,* 1897–1916.

Li, Z., & Durgin, F. H. (2009). Downhill slopes look shallower from the edge. *Journal of Vision, 9*(11):16, 1–15, http://www.journalofvision.org/content/9/11/16, doi:10.1167/9.11.16. [PubMed] [Article]

Li, Z., & Durgin, F. H. (2012). A comparison of two theories of perceived distance on the ground plane: The angular expansion hypothesis and the intrinsic bias hypothesis. *I-Perception, 3,* 368–383.

Loomis, J. M., Klatzky, R. L., Avraamides, M., Lippa, Y., & Golledge, R.G. (2007). Functional equivalence of spatial images produced by perception and spatial language. In F. Mast & L. Jancke (Eds.), *Spatial processing in navigation, imagery, and perception* (pp. 29–48). New York: Springer.

Loomis, J. M., Lippa, Y., Klatzky, R. L., & Golledge, R. G. (2002). Spatial updating of locations specified by 3-D sound and spatial language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(2), 335–345.

Loomis, J. M., & Philbeck, J. W. (2008). Measuring perception with spatial updating and action. In R. L. Klatzky, B. MacWhinney, & M. Behrmann (Eds.), *Embodiment, ego-space, and action* (pp. 1–43). New York: Psychology Press.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81,* 899–917.

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research, 30,* 1763–1779.

Ooi, T. L., & He, Z. J. (2007). A distance judgment function based on space perception mechanisms: Revisiting Gilinsky's (1951) equation. *Psychological Review, 114,* 441–454.

Ooi, T. L., Wu, B., & He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature, 414,* 197–200.

Ooi, T. L., Wu, B., & He, Z. J. (2006). Perceptual space in the dark is affected by the intrinsic bias of the visual system. *Perception, 35,* 605–624.

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.

Philbeck, J. W., & Loomis, J. M. (1997). Comparison

of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 72–85.

Pothier, S., Philbeck, J., Chichka, D., & Gajewski, D. A. (2009). Tachistoscopic exposure and masking of real three-dimensional scenes. *Behavior Research Methods, 41,* 107–112.

Rand, K. M., Tarampi, M. R., Creem-Regehr, S. H., & Thompson, W. B. (2011). The importance of a visual horizon for distance judgments under severely degraded vision. *Perception, 40,* 143–154.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 85,* 618–660.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, 7(14):16, 1–20, http://www.journalofvision.org/content/7/14/16, doi:10.1167/7.14.16. [PubMed] [Article]

Sedgwick, H. A. (1986). Space perception. In L. K. K. R. Boff & J. P. Thomas (Eds.), *Handbook of perception and human performance: Volume 1. Sensory processes and perception* (pp. 21.21–21.57). New York: Wiley.

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology, 38A,* 475–491.

Wirtz, P. W., & Chichka, D. (2013). Angular declination and the dynamic perception of egocentric distance. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication, doi:10.1037/a0034394.

Wist, E. R., & Summons, E. (1976). Spatial and fixation conditions affecting the temporal course of changes in perceived distance. *Psychological Research, 39,* 99–112.

Wu, B., Ooi, T. L., & He, Z. J. (2004). Perceiving distance accurately by a directional process of integrating ground information. *Nature, 428,* 73–77.

Wu, J., He, Z. J., & Ooi, T. L. (2008). Perceived relative distance on the ground affected by the selection of depth information. *Perception & Psychophysics, 70,* 707–713.

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

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162–185.