Control of attention and gaze in complex environments

Jelena Jovancevic  
Center for Visual Science, University of Rochester, Rochester, NY, USA

Brian Sullivan  
Center for Perceptual Systems, University of Texas at Austin, Austin, TX, USA

Mary Hayhoe  
Center for Perceptual Systems, University of Texas at Austin, Austin, TX, USA

In natural behavior, fixation patterns are tightly linked to the ongoing task. However, a critical problem for task-driven systems is dealing with unexpected stimuli. We studied the effect of unexpected potential collisions with pedestrians on the distribution of gaze of subjects walking in a virtual environment. Pedestrians on a collision course with the subject were surprisingly ineffective in evoking fixations, especially when subjects were additionally occupied with another task, suggesting that potential collisions do not automatically attract attention. However, prior fixations on pedestrians did increase collision detection performance. Additionally, the detection of potential collisions led to a short-term change in the strategy of looking at subsequent pedestrians. The overall pattern of results is consistent with the hypothesis that subjects typically rely on mechanisms that are initiated top–down to detect unexpected events such as potential collisions. For this to be effective, subjects must learn an appropriate schedule for initiating search through experience with the probabilities of environmental events.

Keywords: attention, saccades, active monitoring, task demand

Introduction

One of the crucial problems in vision research is understanding the principles that guide the selection of information from visual scenes. Vision is an active process that enables us to have access to relevant information when it is needed. It is clear that it is not possible for the visual system to process all the information in the visual array (Ullman, 1984); thus, there must be mechanisms that guide this selection process (Findlay & Gilchrist, 2003). Deployment of gaze is an overt manifestation of this allocation of attention (Henderson, 2003). Where do people look in natural scenes? Early studies demonstrated that uninformative parts of the scene rarely get fixated (Buswell, 1935). How are “interesting and informative” parts of the image determined or chosen by the visual system? One potential answer is that stimulus-based information generated from the image attracts attention and, thus, gaze, as a part of a bottom–up or exogenously driven system. Another possible answer is that gaze reflects task-directed acquisition of information as a part of a top–down or endogenous system. Presumably, some combination of these factors is at work in natural visual environments, but the relative importance or effectiveness of the two factors is unclear.

Investigation of the problem of deployment of gaze in visual scenes has taken several directions. In one approach involving scene statistics, investigators have found that high spatial frequency content and edge density play a role in attracting fixations (Mannan, Ruddock, & Wooding, 1997). Furthermore, local contrast was found to be higher and two-point correlation lower for fixated parts of the scene (Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Parkhurst & Niebur, 2003; Reinagel & Zador, 1999). In another approach, visual saliency of the image is computed using a model of the known properties of primary visual cortex (Itti & Koch, 2000, 2001; Koch & Ullman, 1985; Torralba, 2003). This “saliency-map” approach uses image features such as color, intensity, contrast, and edge orientation to generate a map of feature salience for each dimension. These maps may be combined to create a single saliency map that indicates regions of interest in an image that should attract fixations. Saliency models can make predictions on the distribution of gaze in a scene, and these predictions can be correlated with human data (Oliva, Torralba, Castelhano, & Henderson, 2003; Parkhurst, Law, & Niebur, 2002). Note that the approaches above are essentially correlational techniques and do not establish a causal link between fixation locations and image properties. Saliency models can be inflexible and typically have no way of dealing with task-relevant objects that are not salient, without some modeling of top–down processes. What is most important, however, is that they account for only a small portion of the variance in the deployment of gaze. For example, Parkhurst et al. (2002) found the correlation between the location of highest salience and the observed fixation
locations to be, on average, 0.45 for images of complex natural scenes and 0.55 for computer-generated fractals.

Other research has concentrated on task-related knowledge, seeking to explain control of gaze in scene perception. For example, Yarbus’ (1967) classic experiments revealed the importance of instructions on the location of fixations, suggesting that cognitive factors are important in defining the locus of gaze, in addition to stimulus factors. However, picture-viewing studies such as this are not easily controlled, as the experimenter often has no access to what the observer is doing from moment to moment during the viewing period. In addition, viewing a picture of a scene is very different from acting within that scene because a different set of information is needed to guide behavior. Wallis and Bülthoff (2000), for example, showed that drivers and passengers in a virtual environment have different sensitivity to changes in the scene.

Several advances in technology such as new mobile eye trackers that can be used in natural environments and the development of complex virtual environments now allow investigation of active gaze control in natural tasks in controlled conditions (Droll, Hayhoe, Triesch, & Sullivan, 2005; Shinoda, Hayhoe, & Shrivastava, 2001; Triesch, Ballard, Hayhoe, & Sullivan, 2003; Turano, Geruschat, Baker, Stahl, & Shapiro, 2001). Recent eye-movement research has focused on extended visuomotor tasks such as driving, walking, sports, and making tea or sandwiches (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land, 1998; Land & Furneaux, 1997; Land & Lee, 1994; Land, Mennie, & Rusted, 1999; Shinoda et al., 2001; Turano, Geruschat, & Baker, 2003). These studies have found that the eyes are positioned at a point that is not the most salient but is relevant for the immediate task demands. Fixations are tightly linked in time to the evolution of the task, and very few fixations are made to regions of low interest regardless of their saliency (Hayhoe et al., 2003; Land et al., 1999; Sprague & Ballard, 2003, in press). Models of the underlying mechanisms of gaze changes that rely on top-down definition of the target have been proposed (Rao, Zelinsky, Hayhoe, & Ballard, 2002; Wolfe, 1994). However, the process by which something becomes a search target is the issue that this article is concerned with.

Naturalistic eye-movement studies have revealed that fixations appear to have the purpose of obtaining quite specific information. For example, cricket players fixate the bounce point of the ball just ahead of its impact because this provides them with critical information in estimating the desired contact point with the bat (Land & McLeod, 2000). These task-specific computations have been referred to as “visual routines” (Ballard, Hayhoe, Pook, & Rao, 1997; Hayhoe, 2000; Roelfsma, Lamme, & Spekreijse, 2000; Ullman, 1984). Visual routines can make use of higher level information to limit the amount of information that needs to be analyzed to that relevant to the current task, thus reducing the computational load. For example, in a block-copying experiment by Ballard, Hayhoe, and Pelz (1995), observers had to copy simple, colored block patterns from a model area to a work area on a computer screen using a mouse to pick up and move blocks from a resource area. In copying one block, observers often made two fixations on a model block: The first fixation is presumably used to identify the color of the block to be copied, the second to acquire the information about the location of the block in the model. Thus, it appears that the two fixations falling on the same object serve the purpose of obtaining two different pieces of information, depending on the momentary task demands. This has been referred to as the “just-in-time” strategy by Ballard et al. because the observer gets only the specific information needed for the particular part of the task just in time for its execution. The fact that the visual information selected is specific to a particular task goal led Ballard et al. (1997) to suggest that visual routines promote computational efficiency.

However, if eye movements are controlled by the task, how does one access perceptual information that is not on the current agenda? In normal ongoing behavior, it is not always possible to anticipate what information is required. How does the visual system divide attention between current task goals and unexpected stimuli that may be important and may change the task demands? This issue has been referred to as the “scheduling” problem (Hayhoe, 2000; Shinoda et al., 2001). These authors suggest two possible answers to this problem. One is that attention is attracted exogenously by the stimulus. The other possibility is that attention is attracted endogenously according to the observer’s internal agenda. Traditionally, basic visual responses have been thought to be driven from the “bottom-up” by the properties of the stimulus. Two lines of research, attentional capture and inattentional blindness, provide insight into how unexpected stimuli are noticed. Folk and Gibson (2001) referred to attentional capture as instances in which attention is drawn to stimuli without the subject’s volition. Implicit measures (e.g., response times and eye movements) are generally relied on to infer shifts of attention (Jonides & Yantis, 1988; Theeuwes, Kramer, Hahn, & Irwin, 1998). Experiments by Theeuwes (1992, 1994), Jonides and Yantis (1988), Yantis and Hillstrom (1994), and others found that an abrupt onset of a salient feature singleton may capture attention in a stimulus-driven, bottom-up fashion. This conclusion has been challenged by others, who claim that the ability of even stimuli such as a unique color or abrupt onset to attract attention is modulated by the current attentional set (Folk, Remington, & Johnston, 1992; Folk, Remington, & Wright, 1994; Gibson, 1996a, 1996b; Gibson & Kelsey, 1998). The inattentional blindness studies, on the other hand, directly probe subjects’ awareness of unexpected stimuli. For example, stimuli that capture attention implicitly may not capture awareness (McCormick, 1997). Recent studies of inattentional blindness have tried to elucidate the factors that lead to noticing of unexpected objects (Mack & Rock, 1998;
Newby & Rock, 1998). In these studies, the authors used brief presentations of simple shapes and found that about 25% of subjects reported no awareness of the unexpected item (Mack & Rock, 1998). More recently, Most, Simons, Scholl, and Chabris (2000), Most et al. (2001), and Scholl, Noles, Pasheva, and Sussman (2003) used sustained and dynamic computerized task to assess the factors that might lead to noticing unexpected events. The authors found that almost 30% of the subjects failed to notice a unique item that traveled across the display for 5 s. Thus, the extent to which unexpected stimuli are processed is unclear. The extent to which these findings generalize to natural vision is also unclear. In ordinary life, the visual system deals with the entire visual field and not with displays that usually subtend only a small portion of the visual field by the use of simple, easily segmented geometric forms as stimuli. Unrestrained subjects generate head and body movements that can create a continuous stream of image motion. Also, the temporal evolution of behavior occurs over several minutes, which is not easily addressed in standard experimental paradigms. In a recent video-based study on selective looking and inattentional blindness, Simons and Chabris (1999) explored whether the unusualness of an unexpected object influences the likelihood of detection. They found that subjects frequently (73%) failed to report a salient visual object in the context of a competing task set, thus suggesting that top-down signals may underlie the phenomenon of “inattentional blindness.” However, the relative role of top-down control versus bottom-up salience in natural vision is not clear.

This problem was addressed in previous experiments with virtual driving (Shinoda et al., 2001). In these experiments, subjects’ ability to detect Stop signs, which were visible for restricted time periods, was examined. Subjects’ performance in detection of Stop signs was found to be heavily modulated by both the instructions and the local visual context (location of the Stop sign in the intersection or midblock). The authors concluded that fixations on Stop signs were primarily controlled based on active search, and thus, the problem of scheduling appeared to be solved by learning an appropriate strategy for a particular context. It is unclear how broadly these results hold, however. The Stop sign in the Shinoda experiment was relatively small and stationary with respect to the scene. Also, its behavioral significance is mostly symbolic (in the sense that ignoring it had no direct consequences in the absence of traffic).

The goal of the present investigation was to probe the question of what controls the distribution of attention in natural vision, with the use of a more salient stimulus that might be more effective in attracting attention exogenously. Because salience is undefined in the absence of a particular model, we chose a situation where the exogenous capture of attention might reasonably be expected and would be particularly advantageous, namely, obstacle avoidance. We devised a virtual environment where observers walked along a footpath with virtual pedestrians. The logic was to examine subjects’ sensitivity to an unexpected event that was chosen based on behavioral relevance and probable saliency. The unexpected event in this experiment occurred when a virtual pedestrian changed its trajectory to a collision path with the observer for a limited time period. When an object in a subject’s field of view is on a collision course, this situation creates a looming stimulus on the retina. Such looming stimuli are commonly thought to be powerful stimuli for attracting attention, and it has also been shown that the bottom–up information needed to attract gaze is present in looming stimuli and that subjects are sensitive to this information (Franconeri & Simons, 2003). Previous research has shown that the rate of expansion of the looming stimulus is one of the sources of information used to compute time to collision (Regan & Gray, 2000; Tresilian, 1999); hence, it is reasonable to suppose that such a stimulus might attract attention in the context of natural walking. Further, neurons in motion-sensitive areas of visual cortex (MST) appear to be sensitive to radial expansions as generated by looming stimuli (Duffy & Wurtz, 1995). Compared with the Stop signs used in Shinoda et al. (2001), a pedestrian on a collision course should be more salient due to the larger visual angle subtended and the smaller retinal eccentricity and should provide a stronger test of the effectiveness of bottom–up signals in attracting attention in natural vision.

Detection of the stimulus (i.e., the colliding pedestrian) should be revealed by a fixation. Detection of the potential collision is somewhat ambiguous, as it might result either from an exogenous mechanism responding to the new configuration of the flow field or from top–down monitoring of the pedestrians in peripheral vision. However, we expect that a top–down mechanism would be less reliable in attracting attention and more susceptible to other attentional demands. Exogenous factors, however, might be expected to be more robust and “capture” attention as shown in a variety of experiments (Jonides & Yantis, 1988; Theeuwes, 1992, 1994; Theeuwes & Godijn, 2001; Theeuwes et al., 1998; Yantis & Hillstrom, 1994).

Because walking is a relatively easy task, in a separate condition, subjects were asked to follow a pedestrian leader. This condition was introduced to modify the attentional demands on subjects. If attentional resources are required to detect the potential collision, the concurrent perceptual task of following a leader should modify the probability of detection. If, on the other hand, the change in the flow field has the power to attract attention bottom–up, then the detection of a colliding pedestrian should be relatively unaffected by the additional task. This condition has the added advantage of stabilizing subjects’ gaze on the leader directly ahead. This introduces less variability in the retinal location of the collision events than with unconstrained gaze. In a separate set of trials, an additional manipulation altered the saliency of colliding pedestrians by increasing their speed during their “collision period.” This variation has the effect of increasing the speed of the looming stimulus and, thus, of possibly increasing the
saliency of the bottom–up signal. If observers rely on bottom–up scene analysis to initiate a particular visual computation, then they should be sensitive to these saliency manipulations.

**Methods**

**Apparatus**

Subjects wore a Virtual Research V8 Head-Mounted Display, as shown in Figure 1. The helmet was equipped with a 3rd Tech HiBall-3000 motion tracker. This is a high-precision analog/optical tracking system that tracks linear and angular motion (6 df) at 2,000 Hz over a 4.8 × 6 m region. To remove measurement jitter, we filtered the HiBall’s position information using the equation:

\[
\text{Position}(t) = 0.9 \times \text{Position}(t-1) + 0.1 \times \text{NewPositionData}(t).
\]

The system latency for updating the scene conditioned on a movement of the HiBall was estimated as ~37–49 ms depending on the time the data were received by the rendering computer. These values were obtained using the head-mounted display and graphics rendering system (described below) described in Triesch et al. (2003). Our previous measures of the rendering system latency (~34–46 ms) were added with an estimate of 3 ms for the HiBall latency (the time from a sensor signal to the reception of a position and orientation data packet by the rendering system) as provided by the manufacturer. The update rate was sufficient such that subjects did not experience a noticeable lag between head motion and the visual update or any consequent motion sickness.

A segment of a 3D model of a town (Performer Town) created by SGI was used so that subjects could walk down a virtual footpath in the town. The footpath includes four corners that correspond to walking along four sides of the 4.8 × 6 m experimental room—a distance of about 21.6 m.

The dimensions of the virtual world are geometrically matched to the real world so that there is no visuo-vestibular or visuo-motor conflict generated by movement through the scene (except for the stereo-conflict between accommodation and vergence inherent in head-mounted displays). The visual display was generated by a Silicon Graphics Onyx 2 computer at a rate of 60 Hz and was rendered in stereo on two LCD screens in the headset, each having a resolution of 640 × 480 pixels and a visual angle of 48° × 36°.

Two devices monitored the movements of the eyes. An Applied Science Laboratory (ASL) 501 Video-Based Eye Tracker monitored the position of the left eye with a 60-Hz temporal resolution and with an accuracy of approximately 1°. Eye position was calibrated by having the subject look at each of the nine points on a 3 × 3 grid. The calibration was repeated six times (between each trial) during the session to make sure that the noise of the eye tracker and the movement of the helmet on the head did not reduce the quality of the tracking. Eye, head, and gaze direction were recorded throughout the experiment and saved in the data file. In addition to the data stream, a video record of the scene, with eye position super-imposed, was captured using a Hi-8 video recorder. An image of the left eye was superimposed on the video record at the top-left corner to allow monitoring of potential track losses. The deviation in the pedestrian path was triggered during a saccade (see below). Because of real-time delays in the ASL signal, it was necessary to use a different eye tracker to trigger these changes during saccades. To do this, a limbus eye monitor (ASL 210) with a 1,000-Hz resolution that was mounted on the right eyepiece monitored the velocity of the eye. The overall latency from the start of a saccade to the scene updating is ~50 ms when triggered by a saccade. The rendering computer monitored the limbus tracker velocity signal in real time and detected a saccade each time the signal was increased to more than 100 deg/s for more than 5 ms. These

![Figure 1](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933537/ on 11/21/2018)
parameters were used to screen the data for saccades with an amplitude of about $\geq 15^\circ$ to ensure that the saccade did not end before the scene update had occurred. The saccade contingent updating system in virtual reality and the evaluation of its performance are more thoroughly described in Droll et al. (2005) and Triesch et al. (2003).

**Walking task**

The basic task for subjects in the experiment was to walk around a virtual block and to avoid virtual pedestrians represented by simple, colored “robot-like” figures (see Figure 2a). We could not maintain a 60-Hz frame rate with more realistic characters; thus, the robot-like figures were used. Although unrealistic, these were highly salient stimuli. Subjects were instructed to walk on a footpath at their usual speed and to avoid the pedestrians in the environment.

Figure 2a shows the path and pedestrians as seen by the subject within the helmet, whereas Figure 2b shows the same path from a bird’s-eye view. There were six robots: four walking in the direction opposite to that of the subject and two walking in the same direction as the subject. Pedestrian paths were set by predetermining eight individual waypoints per pedestrian around the monument (Figure 3). This ensured that the pedestrians were evenly distributed on each side of the white line in the center of the sidewalk path, that there were no collisions between pedestrians, and that there was enough space for the subject to walk between them. Pedestrians were also evenly distributed around the monument so that, in one lap, a subject would see all the pedestrians. However, because the speeds of pedestrians were not equal, the layout of pedestrians on the path was constantly changing, and a subject would thus never see the same layout twice. Pedestrians walking in the direction opposite to the subject were colored green, pink, red, and purple, whereas the two pedestrians walking in the same direction as the subject were colored gray and blue.

Subjects were paid undergraduate students at the University of Rochester. Observations were made on 16 subjects. Subjects were given several practice trials without pedestrians to get used to the virtual environment. When they were comfortable enough to walk at their customary speed, the experiment began. Before the experimental trials began, all subjects were reminded to walk only on the sidewalk and to avoid all oncoming pedestrians. All subjects performed this task with ease. All subjects participated in two main conditions. In one condition, subjects were asked to simply walk on the sidewalk at a normal pace while avoiding pedestrians. In another condition, subjects were given the same instructions as in the previous condition but were also instructed to follow a yellow pedestrian at a constant distance. Waypoints that determine the paths for pedestrians were chosen in such a way that pedestrians stayed off to either side of the white line to ensure that the pedestrians were minimally occluded by the pedestrian leader. Nine subjects had three consecutive No-Leader condition trials, followed by three consecutive Leader condition trials. For the seven remaining subjects, the order of the two conditions was randomized across the six trials. We analyzed the probability of fixation across trials. There were no statistical differences between trials (depending on the order) in either of the two groups of subjects (a group with three consecutive No-Leader trials, followed by the three consecutive Leader trials, and a group with random presentation of the trials) or across these groups; thus, in the further analysis, the trials were treated regardless of order.

**Potential collisions**

The unexpected event in the Leader and No-Leader conditions was the onset of a collision path of a pedestrian.
Given the behavioral relevance of a potential collision (occurring with 10% frequency), who, after 1 s, went back to its original path (see Figure 4). These pedestrians will be referred to as the “colliders.” This situation provides two sources of information that might attract gaze: the collider’s change of angle and the looming caused by the collision course itself. The collider’s instantaneous orientation change for a collision trajectory was triggered during a saccade to isolate the looming cue as the only source of information for bottom–up attention. An additional constraint limited triggered changes to occur only when the pedestrian was between 3 and 5 m from the subject to ensure that the collider never actually “bumps” into the observer. While this decreases the behavioral impact of a potential collision, we did not want the colliders arbitrarily brought to the observers’ attention. If colliders were to continue their collision path and sublend the majority of the scene, the results would be difficult to interpret because detection would be obligatory. In addition, such overt behavior of the pedestrians could also cause dramatic changes in fixation strategy. To enable constant visibility of a collider on its collision course, we made sure that collisions were not triggered when pedestrians were turning the corner. When the above conditions were met, a collision path for a given pedestrian was determined by taking the current position and heading of a subject, adding 1 m to this vector, and setting this as the next waypoint for the collider’s path. After 1 s on this path, a collider would go back to its original path, by walking toward the next nearest of its eight predetermined waypoints. Subjects were not given any feedback on the occurrence of the collider.

Given the behavioral relevance of a potential collision and the large visual angle a collider subtends (average range, 2.9° × 11° to 4.3° × 22°), as well as the relatively small retinal eccentricity (colliders appeared in the periphery over a range of 3° to 22° off center), one would expect that such a stimulus would be relatively easily detected. However, to increase the saliency of a collider even further, in one condition, colliders not only changed their path but also increased their speed by about 25% during the collision period only, before returning to their normal speed. In the constant-speed condition, collider speeds ranged from 0.8 to 0.85 m/s within and outside of the collision period, whereas in the increased-speed condition, their speeds ranged from 1 to 1.06 m/s during the 1-s collision period and 0.8 to 0.85 m/s outside of it. This speed increase of 25% during collision course was relative to the precollision speed of the pedestrian. However, because, on average, subjects travel at 1 m/s and the pedestrians approach them at 1 m/s, relative to the observer, the constant- and increased-speed colliders appear to be approaching at 2 and 2.25 m/s, respectively. This means that the actual increase in speed of a collider during the collision period was 12.5%. Given that the threshold for detection of the increase in speed is 8% (McKee & Nakayama, 1984), our increase of 12.5% should be detectable. The sensitivity to the stimulus was estimated before the beginning of the experiments. Both the change in path and the change in speed were found to be easily detectable when the subject was explicitly looking for the event. This was performed by having a stationary observer wearing an HMD fixate a predetermined spot straight ahead. An array of virtual pedestrians was then presented, with some of them occasionally going on a collision course. Colliders with increased speed were randomly interspersed with the colliders with constant speed. The observer was asked to report whether there was a collider and whether it had increased its speed or not. The pedestrians were evenly distributed on the path. Both the change in path and the change in speed were reliably detected. Different groups of subjects were involved in the actual experiment in the previously described two subconditions (direction change and direction change + increased speed). Nine subjects participated in a subcondition with constant-speed colliders, and seven subjects participated in a subcondition with increased-speed colliders. The data for two subjects

Figure 3. An illustration of a pedestrian’s path around the monument. The red square represents the pedestrian and the stars mark the predetermined eight waypoints.

Figure 4. An illustration of a collision path. The red pedestrian goes on a collision path (marked by an arrow) with a subject (represented by a white ellipse); after 1 s, he goes back to his original path, walking toward the next nearest waypoint (marked by a star).
in the increased-speed condition had to be discarded due to poor eye track, leaving five subjects whose data were analyzed. All subjects participated in three blocks with six circuits in each of the conditions.

**Analysis**

To evaluate whether these potential collision events attracted attention, we measured fixations on colliders during the collision period.

Fixations were determined using in-house Fixation Finder software that implements an adaptive velocity-based algorithm that adapts the velocity threshold depending on an estimate of the noise level present in the signal for each subject. The initial threshold for saccade detection from the eye-movement velocity was set to 65 deg/s. This high threshold was used because of the noise level present in the track signal and the low temporal resolution of the tracker. All recorded fixations needed to meet the criteria of having an angular velocity of less than 65 deg/s for at least 60 ms. Successive eye position samples occurring less than 30 ms apart, and with a displacement of less than 1°, were collapsed into one. This automated scoring of fixations was found to be comparable to that of manual frame-by-frame analysis of video records. In-house Matlab (MathWorks) functions were used to analyze eye movements, including identifying the object where each fixation fell on and the duration of the fixation. An automated labeling system for determining the visibility of pedestrians at any given moment was used. The subjects’ gaze vector was projected in 3D and tested for which surface it first intersects to determine the object of fixation. Every frame for each fixation for which surface gaze intersected was evaluated. Due to noise in the tracker and subject motion, a given fixation may intersect with multiple surfaces. The surface with most intersections in a given fixation was designated as the object label for that fixation. Pedestrians were surrounded by a bounding box with a height of 1.7 m and a width and depth of 0.35 m to simplify calculations. The virtual dimensions of pedestrians are 1.7 m in height and 0.75 m in width and depth. The extra 5 cm was added due to the noise in the track and served to achieve a 93% agreement rate between automated analysis and manual coding. At 1 m, this means that a pedestrian has a boundary that extends ~0.75° on each side to allow for noise in the eye-tracking signal. The automated labeling only labeled each pedestrian type. Any fixation on an object that was not a pedestrian was labeled as “other.” Eye position was monitored for all 16 subjects. Automated fixation analysis is sensitive to noise in the tracking signal; hence, it was necessary to screen out data resulting from track losses. The automated analysis scans the raw horizontal and vertical eye position data for track losses by excluding abnormally large or negative values. Track losses were evaluated across all trials for all conditions. On average, about 1% of frames per trial (about 1 s of data for a 100-s trial) was marked as track losses. There were no significant differences between conditions. The automated analysis uses a simple rule when it encounters fixations that are possibly broken up by a track loss. If the track position before and after the loss remains within 2.5° of one another, the fixation is not ended and further data are analyzed for the end of fixation. If the position criterion is exceeded, the fixation is ended and a new fixation is marked when valid data are received.

Because of the difficulty in maintaining an accurate track within the virtual reality helmet, only a subset of subjects had adequate eye position data throughout the experiment to merit analysis with the automated fixation finder. Analysis of the video records resulted in the selection of 14 subjects (9 in the constant-speed condition and 5 in the increased-speed condition) for whom eye position was judged adequate for automated analysis. As a check of the validity of the automated analysis, the eye movements of four subjects (two in the constant-speed condition and two in the increased-speed condition) were coded manually and separately by two of the authors using a frame-by-frame analysis of the video record. A frame-by-frame comparison of the two coders’ fixation records showed 93% (SE = 2) agreement. The automated analysis output was compared to manual labeling using a total of 10 trials selected from four subjects. Fixations on pedestrians were compared frame by frame between manual and automated analyses. The automated analysis matched 93% (SE = 0.9) of the fixations found in the manual coding, on average, for each trial. On average, 9% (SE = 1.1) of the fixations labeled by the analysis were deemed false positives. Examination of the data showed that most of the false positives identified were due to the point of gaze being located in an ambiguous position between two closely spaced pedestrians. Given that the automated analysis nearly matched the agreement between human observers, all subsequent analyses of subject fixation data were conducted using only the automated analysis.

Statistical analysis of the data was performed using analysis of variance (ANOVA) with one factor. Post hoc analysis using Tukey contrast was performed where appropriate to determine the pairwise significance of three or more samples, for which the analysis with ANOVA showed a statistical effect.

**Results**

Several aspects of subjects’ performance were analyzed and are presented here. Because the fixation locus and focus of attention are tightly linked, for these analyses, eye movements were taken as a measure of the subjects’
attentional state (Corbetta et al., 1998; Findlay & Gilchrist, 2003; Shinoda et al., 2001). Note that while eye tracking provides a record of the current object being fixated, the visual system must certainly deal with objects in the periphery (such as planning the next saccade). Although we use foveation as an attentional measure, it does not provide a complete description of the subject’s possible focus of attention. In the following, we will present data that suggest that peripheral monitoring may be a common strategy in the walking environment. We will first examine the general strategy that the subjects employ when fixating normal pedestrians. We will then describe the fixation patterns on colliders. Next, we will consider the effect of the increased speed of colliders on fixation patterns. Then, the effect of the detecting of a collider on fixations of normal pedestrians will be examined. We will then present the effect of color, distance, degree of rotation, and the number of pedestrians on-screen on fixation patterns. Finally, we will present the effect of collider fixations on walking performance, such as the change in subject velocity and distance to the leader between precollision and collision periods.

Fixations on normal pedestrians

We extracted fixation patterns on normal pedestrians during all trials. To examine the general strategy subjects use to distribute gaze while walking in the presence of pedestrians, we analyzed the data in such a way that all normal pedestrians were grouped according to the time they appeared in the field of view, in independent 1-s bins. Then, of all the pedestrians present in each time bin, the number of those that were fixated was determined. This is shown in Figure 5 (note that the points do not sum up to 1 because probabilities in each time bin are computed separately; error bars in this and all subsequent figures are ±1 SEM between subjects). In the No-Leader condition, out of all the pedestrians present in the 1-s period after they appeared, 25% were fixated. Of all the pedestrians present in the field of view between 1 and 2 s after their appearance, 65% were fixated, and so on. Our data show that subjects are most likely to fixate normal pedestrians 1–2 s after they appear in the field of view. This is similar but not identical to the point furthest from the observer. At 2 s, pedestrians are typically, but not always, at a 5-m distance. Subjects are clearly sensitive to new pedestrians, and the general strategy appears to be that they inspect them at a distance, possibly to predict their paths.

Figure 5 also shows the fixation patterns for the Leader trials. The overall trend of fixating pedestrians when they first appear is the same, but the added task of following a leader reduces the number of fixations by about a factor of 2.

Figure 6 (blue bars) shows the spatial distribution of fixations on pedestrians and other objects in the environment for the No-Leader condition. In this analysis, all pedestrians are placed in one bin and compared to fixations on other objects in the environment (the walkway, grass, etc.). Note that most fixations (~70%) are not on pedestrians. Most fixations are directed toward other objects, most likely for navigational purposes (such as the walkway, the white line on the walkway, or the central stone structure that the pathway winds around). Figure 6 (purple bars) shows the same analysis for the Leader condition. Here, we see that fixations on the leader now account for most fixations. Fixations on pedestrians and other objects are cut nearly in half compared to the No-Leader condition.

Fixations on colliders

Following our observations on normal pedestrians, we examined fixation patterns on colliders. Detection of the collider could occur either if the new flow field configuration (rather than the change in configuration, which is
masked by performing the change during a saccade) attracts attention bottom–up or if the observer is using the peripheral retina to actively monitor (top–down) other moving objects in the field. Perfect detection could result from either of the two, but failure to detect is more frequent and more consistent with top–down monitoring. If active monitoring is required to detect the collider, then there should be instances when the observer is not monitoring and consequently misses the collider. For example, an additional perceptual task of following a leader reduces the attentional resources necessary for monitoring, thus making potential collisions more difficult to detect. If, on the other hand, the flow field configuration itself has the power to attract attention bottom–up (Franconeri & Simons, 2003), then we might expect that a pedestrian on a collision path would almost always be fixated. Movie 1 shows an example of a missed collider in slow motion. A subject following the leader misses a purple pedestrian that goes on a collision path and then goes back to its original path.

An example of a successful detection of a collider is shown in Movie 2. A purple pedestrian goes on a collision path with the subject, followed by a fixation on that colliding pedestrian.

Most of the pedestrians that turned into colliders did so in the first 3.3 s after their appearance on-screen (98%).

Additionally, the probability of fixating a pedestrian varies with the amount of time the pedestrian has been present on-screen (see Figure 5). Because collider onset times vary and the probability of fixating a pedestrian varies with duration on-screen, we need to examine the effect of colliders holding time on-screen approximately constant. Initially, to control for time on-screen, we placed all pedestrians that turned into colliders into three bins according to the collision period (0.2–1, 1–2, and 2–3 s). There were 131, 107, and 59 colliding pedestrians in each bin, respectively. Figure 7 schematically illustrates the paths of normal pedestrians and a colliding pedestrian. To compare the probability of fixations on colliders and normal pedestrians during collision time, we matched pedestrians that turned into colliders with normal pedestrians in the following way: using the distribution of collision onset times, we created three simulated collision time bins for normal pedestrians. Because there is no actual collision onset time for normal pedestrians, the center of each bin was selected as the collision onset time for them (0.6, 1.5, and 2.5 s). Pedestrians appear on-screen for a variety of durations, spanning from 1 to 5 s (see Figure 8). Pedestrians were included in each bin depending on their on-screen duration. For instance, a pedestrian present on-screen for 5 s would be evaluated for fixations in each bin, whereas a pedestrian on-screen for only 1.6 s would only be evaluated in the first bin. This process yielded 2,235, 1,444, and 577 pedestrians for the bins whose collision onset times were 0.6, 1.5, and 2.5 s respectively.

These normal pedestrians were then matched to the colliding pedestrians within the three previously described time bins. After this analysis, it was found that although fixation probabilities do vary over time, the difference between the colliders and pedestrians in each collision period onset bin was constant over time. Thus, further
analyses of noncolliding pedestrians and colliders are presented where bins across time have been collapsed into one.

Figure 9 displays the probability of fixation for the collision period. During this period, in the No-Leader condition, the probability of fixating a collider was 60% (with an average duration of 360 ms), whereas the probability of fixating a noncolliding pedestrian was 41%, \(F(1,23) = 6.924, p = .015\) (via one-way ANOVA). Fixations on colliders lasted 320 ms on average in the No-Leader condition and lasted 240 ms in the Leader condition. Although there was a significant increase in fixations in the No-Leader condition, this effect did not hold in the Leader condition. In the Leader condition, the probabilities of fixating colliders and noncolliding pedestrians are roughly equivalent to 25%. Colliders were fixated for 270 ms on average. To show that there were no systematic differences in fixation patterns on normal pedestrians and those destined to become colliders at a later time, we also analyzed these two groups of pedestrians in the precollision period. We found no difference in the probability of fixations between these two groups (probabilities of fixations for the colliders and normal pedestrians were .48 and .47 for the No-Leader condition, respectively, and .29 and .28 for the Leader condition, respectively).

**Fixations: Conditional probabilities**

Subjects often fixate a given pedestrian more than once. This is particularly interesting in the case of pedestrians that turn into colliders. We investigated what percentage of fixated pedestrians were fixated a second time in the period when they turn into a collider. We will refer to these as refixations, or \(p\) (fixation during the collision period | fixation before the collision period). After first fixating a pedestrian, it is possible that the subject continues to monitor it in his peripheral vision. If the pedestrians are being monitored, it is plausible that when that pedestrian starts walking toward the subject, the potential collision is going to be easier to detect. This should be revealed by a greater probability of “refixating” this pedestrian. If, however, these pedestrians are not monitored in peripheral vision, this initial fixation should not matter and the probability of fixating a collider should be the same whether the pedestrian was fixated earlier or not. We have taken all the pedestrians that turn into colliders that were fixated prior to the collision period (40%) and computed the percentage of these that had a refixation during the collision period. In the No-Leader condition, there is a significant difference between colliding pedestrians and matched noncolliding control pedestrians; 60.4% of collider pedestrians were refixated, whereas only 39.7% of matched noncolliding control pedestrians were refixated, \(F(1,19) = 14.9, p = .001\) (one-way ANOVA; Figure 10a). The increased rate of refixation for collider pedestrians in the No-Leader condition indicates that previously fixated pedestrians may be peripherally monitored. In the Leader condition, there is no difference in the probability of refixating between these groups: 30.8% versus 30.1%, respectively, \(F(1,20) = 0.004, p = .949\) (one-way ANOVA; Figure 10a).

We also examined the reversed conditional probability: Given that the collider was fixated, what is the probability that it had been fixated prior to the collision period or \(p\) (fixation before the collision period | fixation during the collision period)? We will refer to these as prior fixations. We have taken all fixated colliders (60%) and computed the percentage of these that had a prior fixation before the collision period. In the No-Leader condition, there is a significant difference between colliders and controls. Sixty-five percent of colliding pedestrians had a prior fixation, whereas only 40% of matched noncolliding control pedestrians had a prior fixation, \(F(1,22) = 8.581, p = .008\) (one-way ANOVA; Figure 10b). This effect between colliders and controls disappears in the Leader condition, where the probabilities of prior fixations for colliders and

![Figure 8](https://jov.arvojournals.org/pdfaccess.ashx?url=data/journals/jov/933537/) The distribution of the on-screen durations of non-colliding pedestrians. The peak at the beginning is most likely due to head movements when pedestrians are close to the subject and can easily go in and out of view.

![Figure 9](https://jov.arvojournals.org/pdfaccess.ashx?url=data/journals/jov/933537/) The probability of fixations on colliders versus time-matched normals during the collision period. There were more fixations on colliders but only when the subject was not following a leader.
controls were 23% and 20%, respectively, \( F(1,19) = 0.0035, \ p = .952 \) (one-way ANOVA). By comparing the Leader and No-Leader conditions, we found that the probability of a prior fixation for colliders significantly dropped in the Leader condition: 65% versus 20%, \( F(1,22) = 20.871, \ p = .001 \) (one-way ANOVA). Presumably, the additional task of following a leader reduces the attentional resources necessary to monitor pedestrians. Subjects appear to be more sensitive to the onsets of collision paths in those pedestrians that they have recently fixated. This is consistent with the hypothesis that subjects are monitoring these pedestrians in their periphery.

**Effect of collider speed on potential collisions**

We were surprised at the low rate of fixations to colliders. One possible reason is that the change in trajectory was not salient enough to capture attention or gaze. To investigate how increased saliency influences fixations, we increased the speed of the colliders by 25%. If attention were captured bottom–up, one would expect that a more salient stimulus would increase the probability of fixations. We found no effect of speed of colliders on the fixations. In the constant-speed and increased-speed conditions, probabilities of fixations were 58% and 61%, respectively, in the No-Leader condition and 34% and 30%, respectively, in the Leader condition (Figure 11a and 11b). Additionally, the latencies from the collision onset time to the first fixation of that collider were examined. We found no significant differences between constant-speed and increased-speed colliders. The fact that a more salient stimulus had no effect on fixations suggests that peripheral monitoring and not bottom–up capture of attention was responsible for the increased probability of fixations on colliders in Figure 11.
Effect of collider detection on fixations on normal pedestrians

Shinoda et al. (2001) hypothesized that subjects handled environmental uncertainty by using learnt knowledge of the probabilistic structure of the environment to initiate task-specific computations at likely points. For example, when subjects were asked to follow normal traffic rules, they spent much more time looking in the neighborhood of the intersection than when asked only to follow the car in front. This suggests that subjects rely on the regularities in the environment to schedule their searches. We were interested to find out if the subjects’ performance in the walking task can be modified by their expectations about the environment. In other words, does the prior experience of detecting a potential collision produce more vigilant fixation patterns on normal pedestrians in case they too are potential colliders? We analyzed the number of fixations on pedestrians following a collision period. Although there was a trend for increasing the number of fixations on pedestrians following a fixation of a collider versus a nonfixated collider, we found no significant differences. However, there could be some short-term effects in fixation duration (similar to the analysis in Droll et al., 2005). We analyzed total fixation durations on pedestrians in a 3-s period (the average time a pedestrian is present on-screen) following a collision period. If monitoring the trajectory of pedestrians is under top–down control, successful detection of colliders should presumably change the pattern of fixations in such a way to enable easier detection and/or avoidance of potential future collisions. A good strategy would be to increase fixations of normal pedestrians as potential sources of collisions. This is consistent with the research that shows that eye movements are driven by prospects of reward. The sensitivity of eye-movement patterns to stimulus probability is consistent with a system that is shaped by learning (Hikosaka, Takikawa, & Kawagoe, 2000). This is also consistent with models that use second-order reward statistics (such as uncertainty and risk) to choose where the center of gaze will be allocated next (Sprague & Ballard, 2003). In our case, following a detection of a collider, subsequent fixations are distributed in such a way that could plausibly reduce the uncertainty of the environment and reduce the possibility of future collisions.

The total time spent in fixating pedestrians increased from 550 ms following a miss to 830 ms following a fixated collider in the No-Leader condition (where a “miss” means simply that the collider was not fixated). In the Leader condition, total fixation duration increased from 280 ms following a miss to 550 ms following a detected collider (Figure 12). These differences were found to be significant in both the No-Leader and the Leader conditions, $F(1,26) = 5.651, p = .025$ and $F(1,25) = 13.59, p = .0011$, respectively (one-way ANOVA). This suggests that subjects not only rely on immediate potential collisions to allocate gaze but also base short-term pedestrian fixation strategies on prior experience. A potential collision event apparently leads subjects to increase the priority of monitoring pedestrians.

Effect of stimulus attributes of colliders

A number of different factors could be affecting the pattern of fixations on both the normal pedestrians and colliders. The effect of the number of pedestrians on-screen, the effect of the distance from the observer, the effect of the degree of rotation of a collider, and, finally, the effect of the color of pedestrians on fixations were examined.

Number of pedestrians on-screen and fixations

It is a natural assumption that the degree of complexity of the scene might somehow affect the fixations within the scene. For example, more pedestrians in the scene would presumably divide the attentional resources available, thus lowering the probability of fixating a collider. We found no evidence of this. The probability of fixating a collider was not significantly affected by the number of pedestrians on-screen during the collision period (Figure 13). In the presence of the added task of following a leader, the overall probability of fixating colliders are lower compared to No-Leader trials (as discussed above), but there is still no significant effect produced by the number of pedestrians on-screen (Figure 13).

Distance from the subject and fixations

It is possible that some of the colliders were more noticeable due to the mere fact that they were closer to the observer. To test for the effect of the distance of the collider from the observer on the probability of fixating a collider,
we divided all colliders into three bins based on the distance from the observer at the time of the rotation onset. No significant effect of the distance from the observer was found in either condition (Figure 14). Again, however, the overall probability of fixating a collider in the Leader condition was lower than that in the No-Leader condition.

**Degree of rotation of a collider and fixations**

To explore the possibility that the degree of rotation of collider affects the probability of fixating it, we divided all colliders into bins relative to the degree of rotation while on a collision course, and the probability of fixating colliders within those bins was computed. Considerable variability within and between subjects was observed. In the No-Leader condition, very small (0–5° and 5–10°) and very large (20–25°; Figure 15, dark red trace) rotational angles seemed to have attracted more fixations than medium-sized deviations (10–15°). These differences, however, were not statistically significant. This distribution changes in the Leader condition where subjects seem to fixate colliders with medium-sized rotations (10–15°) more than others (Figure 15, light red trace). Significant differences were found between groups in the probability of fixations relative to different degrees of rotation of colliders (0–5°, 5–10°, 10–15°, 15–20°, and 20–25°), $F(4,48) = 2.692$, $p = .042$ (one-way ANOVA). Post hoc analysis using Tukey contrast showed significant differences between Group 3 (10–15°) and Groups 1, with smaller changes in trajectory (0–5°; $p = .038$), and 5, with large changes (20–25°; $p = .0054$). However, we found no relation between the degree of rotation and the distance of the subject to the pedestrian leader. Thus, the reasons for these peculiarities in the patterns of fixations on colliders with respect to degree of rotation are not yet clear.

**Color of pedestrians and fixations**

Colliders had fairly saturated colors and were highly visible. However, it is possible that some colors might be more salient than others. To investigate the effect of color of colliders on fixations, we divided all colliders into groups according to their color and we computed probabilities of fixations. No significant differences were found between different groups in either condition (No Leader and Leader; Figure 16).

**Effect of colliders on walking performance**

Thus far, the analyses presented have been restricted to fixation behavior. Although fixation patterns suggest that a collider has attracted attention in some way, it may not reveal covert detections. We therefore investigated other potential indicators that show that subjects had detected colliders: (1) a change in the distance to the leader or (2) a change in the subject’s velocity.
To investigate whether fixations of colliders affected the walking performance of the observer, we computed the change in the average distance to the leader between the range of ±500 ms around the collision period onset and the range of 500 to 1,500 ms after the collision period onset moment. These data were analyzed for effects due to fixation during the collision period and fixations prior to its onset. These ranges were chosen to allow for latencies in the subject’s visuomotor response. Although the changes in distance were small, there was a significant increase, $F(1,24) = 4.843, p = .0376$, in the change in the distance for fixated colliders versus that for nonfixated colliders. This is shown in Figure 17a. There was no significant change in the distance to the leader during the collision period when a collider was not fixated; that is, the value in Figure 17a is not significantly different from zero. A similar analysis was conducted to investigate whether fixations of noncolliding pedestrians affected subjects’ walking performance. We found no significant change in the distance to the leader between fixated and nonfixated noncolliding pedestrians.

We also investigated whether a fixation on a pedestrian prior to the fixation during its collision course had any effect on the subsequent change in the distance to the leader. If subjects have fixated a pedestrian prior to its collision course onset, they may be more prepared for the subsequent collision course. We found that in colliders fixated prior to collision course onset, there was a significantly smaller change in the distance to the leader than for those without prior fixations, $F(1,16) = 5.047, p = .039$ (one-way ANOVA; see Figure 17b). We performed a comparable analysis to investigate whether having a prior fixation on a noncolliding pedestrian affects a subsequent change in the distance to the leader. We found no significant change in the distance to the leader for the noncolliding pedestrians that had a prior fixation versus the pedestrians without a prior fixation.

An analysis was then performed on the change in average velocity of the observer for both the No-Leader and the Leader conditions. Although there was a trend to slow down more for fixated colliders than for nonfixated ones in both conditions, these differences were not significant. Similarly, no significant change in velocity was found for fixated versus nonfixated noncolliding pedestrians.

Prior fixations on a pedestrian before collision course onset influence subject velocity in a manner consistent with those previously described for the change in the distance to the leader. If a pedestrian was fixated both before and after it turned into a collider, the change in velocity of the observer during the collision course was significantly smaller than for those colliders that did not have a prior fixation (see Figure 18). These differences were significant for both the No-Leader and the Leader trials, $F(1,19) = 4.79, p = .041$ and $F(1,15) = 4.674, p = .047$, respectively. A comparable analysis was then performed for pedestrians that do not turn into colliders. No significant changes in

**Figure 16.** Effect of the color of colliders on the probability of fixations. No apparent preference for any particular color was observed, except for red in the Leader condition.

**Figure 17.** (a) The change in the average distance to the leader during the collision course was greater for the fixated colliders than for those that were not fixated. (b) Effect of a prior fixation on the change in the distance to the leader. Change in the distance to the leader was smaller for the colliders that had a prior fixation before they turned into colliders than for those that did not have a prior fixation.
velocity were found for pedestrians only fixated in the simulated collision period versus pedestrians with a prior fixation.

A possible explanation for the result of having a smaller change in velocity for the colliders with a prior fixation is that when subjects first fixate a pedestrian before it turns into a collider, they keep monitoring it in peripheral vision and attempt to keep track of its future trajectory. This may make them more prepared to notice any changes in the pedestrians’ trajectory, thus minimizing the need for any abrupt adjustments in their own behavior. These results are consistent with those obtained in the analysis of the probability of prior fixations on colliders, in that they provide evidence that subjects keep monitoring the pedestrians they already fixated in the periphery. We suggest that pedestrians that were first fixated before turning into colliders be peripherally monitored to reduce the uncertainty about the future trajectory of such pedestrians. Continued monitoring means that a change in pedestrian trajectory does not affect the behavior of the observer as much as it does when this uncertainty is greater, as shown by the smaller decrease in velocity for fixated colliders with a prior fixation versus those without.

**Discussion**

The goal of this experiment was to more precisely elucidate the mechanisms by which the visual system controls attention and gaze in complex scenes. Despite the evidence that gaze is tightly controlled by the ongoing task in the context of natural behavior, top–down control may have limited effectiveness in dynamic environments because of the difficulty of dealing with unexpected events. In a previous attempt to explore this problem, Shinoda et al. (2001) found that subjects were successfully able to detect unexpected, briefly presented Stop signs based on learnt, top–down strategies. In the present experiment, we sought to provide a more challenging test for top–down hypotheses by devising a situation where bottom–up exogenous attentional capture was more likely. We investigated performance by detecting potential collisions with pedestrians. This provided a much larger and presumably more salient stimulus event than the Stop signs. The probability of fixating colliding pedestrians was greater than that of fixating normal pedestrians. Surprisingly, this increase was relatively modest. An increase of only 18% in the No-Leader trials was observed so that about 40% of colliders were missed. With a rather long colliding pedestrian presentation time (1 s), the task of detecting a collider should be relatively easy if our visual system relied on bottom–up mechanisms to guide gaze. Although this poor detection of colliders is surprising, this result is ambiguous in itself. Eighteen percent of detections could be a consequence of either bottom–up attentional capture or active monitoring of events in peripheral vision. However, the overall pattern of results appears more consistent with a top–down interpretation.

First, even the limited ability of the colliders to attract gaze was abolished when another task was added. This sensitivity to an added task is consistent with a top–down mechanism, although other explanations are possible. Sixty percent of total fixation time was spent on the leader; thus, the time spent fixating pedestrians and other objects was cut in half compared to the No-Leader condition. It is possible that the failure to fixate colliders was because subjects assume that the leader provides a safe path and because they are relieved of some of the burden of obstacle and pedestrian avoidance. It is also possible that the leader is a more salient object and, as such, attracts most fixations. However, it seems odd that in the condition with no leader but with a number of colorful, salient pedestrians, 70% of fixations are not on pedestrians. It seems unlikely that adding one more pedestrian (the leader) would attract 60% of the gaze solely based on salience and must be at least in part due to task demands. Another alternative is based on Lavie’s theory that the added perceptual load of following a leader interferes with the detection of colliders. King, Dykeman, Redgrave, and Dean (1992), Lavie (2005), and Lavie, Fockert, and Viding (2004) demonstrated that a high perceptual load reduces interference from distractors. In the present situation, we can consider the following task as a source of higher perceptual load and the collider as a distractor stimulus. Several imaging studies also indicate reduced activity caused by distractors in a variety of cortical areas in conditions of high perceptual load (e.g., Rees, Frith, & Lavie, 1997; Rees, Russell, Frith, & Driver, 1999; Vuilleumier, Armony, Driver, & Dolan, 2001; Yi, Woodman, Widders, Marois, & Chun, 2004). Rees et al. (1997) found that irrelevant motion background evoked responses in motion-selective cortices (MT, V1/V2) in low-load conditions but not in high-load conditions. These findings suggest that even salient stimuli may not
evoke significant cortical activity or attract attention if perceptual/cognitive resources are not available. This is also consistent with the inattentional blindness findings of Mack and Rock (1998). Thus, if the 18% detection of the collider in the No-Leader condition is a result of attentional capture by bottom-up signals, it appears to be heavily modulated by the overall allocation of attentional resources.

If top–down control is indeed responsible for the detection of colliding pedestrians, what is responsible for the 18% increase in the probability of detecting colliders in the No-Leader condition? What is the mechanism by which top–down systems deal with such unexpected events? One plausible explanation is that the observers are using the peripheral retina to actively monitor (top–down) other moving objects in the field. The relatively high frequency with which subjects fixated normal pedestrians suggests that they are the focus of significant attentional resources. Other evidence comes from our results on the refixations of pedestrians after they turn into colliders. We reason that if one actively monitors pedestrians in periphery, one is more likely to detect a change in them. The finding that 60% of fixated pedestrians had another fixation during collision period (compared with 40% for the control value) speaks in favor of this interpretation. Perhaps the strongest evidence for active monitoring comes from our finding that there is a change in strategy of looking at pedestrians following a detection of a collider. Namely, total time spent on pedestrians is increased following a detected collider. This suggests that subjects are sensitive to environmental events and modify the distribution of attention accordingly. Additionally, the standard error between subjects in a variety of fixation measures is quite small, indicating regularity in the fixation strategies that subjects employ. It is also worth noting that subjects do not appear to learn that the colliders remain on a collision trajectory for 1 s and that they do not actually intersect with the subjects’ path. Why is it that subjects cannot learn to ignore colliders across trials? Our subjects never suffer the negative consequences of an actual collision; thus, would they perhaps fixate the colliders less as this fact was learned? We conjecture that obstacle/pedestrian avoidance is a highly learned activity that may be difficult to extinguish within an hour of testing. Collisions with obstacles and pedestrians in real life are quite rare. Presumably, there is some avoidance system engaged to prevent such collisions even when the frequency of actual collisions is rather low.

The strength of our conclusion that top–down mechanisms are an important factor in avoiding potential collisions in natural vision rests heavily on the veridicality of the virtual environment. It may be the case that the visual stimulus occasioned by colliding virtual pedestrians is simply less salient than in real life. Consequently, it would be desirable to perform a comparable experiment in a real environment. Such an experiment was recently performed in our laboratory, and the preliminary results seem to be consistent with the results presented in this article. While colliders were detected at a greater rate than noncolliders in the No-Leader condition, it is still surprising that the virtual colliders were still frequently undetected (about 40% missed in the No-Leader condition). As previously described, the collision events were easily visible when subjects were explicitly attempting to detect them. Manipulating the speed of the collider to increase the saliency even further failed to evoke more fixations. Because the initial stimulus tests were conducted in a stationary manner, it may be the case that with the added task load of walking, the 12.5% change in collider velocity is not easily detectable by the moving observer, which explains our lack of significant effects for this manipulation. Additionally, we examined the latencies from the collision onset time to the first fixation of that collider. No significant differences were found between constant-speed colliders and increased-speed colliders. Despite the issues described above, the fact that the increased stimulus speed failed to effect the probability of fixations of colliders may be interpreted as evidence of a top–down influence on gaze deployment. Added to this, different stimulus properties such as the number of pedestrians on-screen, the distance from the observer, the color of pedestrians, and the degree of rotation had no consistent effect on fixations.

Our finding that detection of colliding pedestrians seems to be largely under top–down control is consistent with a recent computational model of attentional selection and gaze allocation by Sprague and Ballard (2003, in press). This model is helpful in understanding how such a top–down system might be implemented. In their model, a simulated person walks in a similar environment to that used by the subjects in the present experiment. The model subject has three tasks: to follow the sidewalk, to avoid obstacles, and to pick up objects. The allocation of attention to these three tasks is entirely under top–down control, and the appropriate schedule must be learnt by experience with the frequency of obstacles and pickup objects and how often it is necessary to look at the path to stay on it. This model uses second-order reward statistics of uncertainty and risk to choose between ongoing competing tasks. More direct comparisons of human performance with this model will help explore the limits of a purely top–down system (Rothkopf, Ballard, Sullivan, & de Barbaro, 2005). The development of such models provides a formal system for modeling complex behavioral sequences, a critical component in being able to go beyond a simple description of fixation patterns as “task driven” to being able to predict, in detail, the observed fixation sequences in natural behavior (Rothkopf et al., 2005). Recent work on the neurophysiological basis of eye movements has revealed that the saccadic eye-movement circuitry is sensitive to the reward structure of the task (Hikosaka et al., 2000; Platt & Glimcher, 1999). Hayhoe and Ballard (2005) have argued that this sensitivity to reward serves as a substrate for mediating the linkage between fixations and task structure in natural behavior. That is, it provides a basis for the kind of learning of where and how frequently to fixate different
regions as in the Sprague and Ballard model or as demonstrated by the human subjects in the walking environment used in the present experiment. Indeed, Schultz (2000) showed that the behavior of the dopaminergic neurons in the substantia nigra pars compacta, a part of the basal ganglia system (a reward system integral to the generation of saccadic eye movements), can be predicted by mathematical models of reinforcement (Montague, Hyman, & Cohen, 2004; Schultz, 2000).

Arguably, subjects might be detecting pedestrians and colliders covertly, thus making it difficult to evaluate the magnitude of the bottom–up component. We, however, found no measurable change in behavior (no significant change in subjects’ velocity or distance to the leader) during collision courses when subjects did not fixate the collider. It seems likely that overt gaze location specifies the primary distribution of attention and that subjects were simply unaware of the collision trajectory.

Is there, then, involuntary attentional capture by certain salient stimuli? Studies by Theeuwes and Godisn (2001), Theeuwes (1992, 1994), Jonides and Yantis (1988), Yantis and Hillstrom (1994), and others found that abrupt onsets of salient stimuli singleton may capture attention in a bottom–up fashion. As previously discussed, these findings have been challenged by others, who claim that the ability of salient stimuli to capture attention is modulated by the current attentional set (Folk et al., 1992, 1994; Gibson, 1996a, 1996b; Gibson & Kelsey, 1998). Given the complexity of subjects’ motor behavior and of the visual scene in natural contexts, the ability of abrupt onset stimuli to capture attention in all of these studies may be exaggerated. The results of our study are more consistent with the body of work on inattentional blindness (Mack & Rock, 1998), which suggests that when subjects are engaged in one task, they are remarkably insensitive to other perceptual tasks.

Conclusions

Although walking is relatively simple, it entails a variety of subtasks: maintaining a heading, keeping track of one’s surroundings and footing, and avoiding potential obstacles. This study suggests that the way subjects distribute their attention across a scene is determined by a relatively small number of behavioral goals with varying priorities. This is consistent with a recent computational model of attentional selection and gaze allocation by Sprague and Ballard (2003, in press). When encountering pedestrians, subjects appear to inspect them at a distance, possibly to predict their paths. The overall pattern of results is consistent with the hypothesis that the visibility of colliders depends on active monitoring, according to a schedule depending on the observer’s current task and estimate of environmental probabilities.

Acknowledgments

This research was supported by National Institutes of Health Grants EY-05729 and RR-09283.

Commercial relationships: none.
Corresponding author: Jelena Jovancevic.
Email: jjovancevic@cvs.rochester.edu.
Address: Center for Visual Science, University of Rochester, Rochester, NY, USA.

References


ignoring to sustained inattentional blindness. Psychological Science, 12, 9–17. [PubMed]


