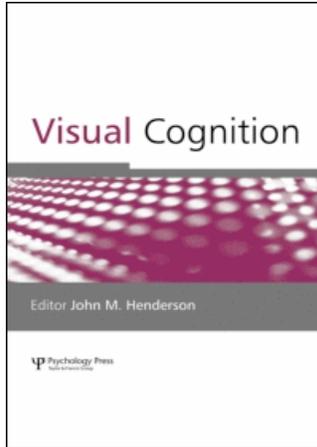


This article was downloaded by:[Hayhoe, Mary M.]  
On: 26 March 2008  
Access Details: [subscription number 790941544]  
Publisher: Psychology Press  
Informa Ltd Registered in England and Wales Registered Number: 1072954  
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Visual Cognition

Publication details, including instructions for authors and subscription information:  
<http://www.informaworld.com/smpp/title~content=t713683696>

### Is attention drawn to changes in familiar scenes?

Hacer (üke) Karacan <sup>a</sup>; Mary M. Hayhoe <sup>bc</sup>

<sup>a</sup> Department of Cognitive Science, Middle East Technical University, Ankara, Turkey

<sup>b</sup> Center for Perceptual Systems, University of Texas, Austin, TX, USA

<sup>c</sup> Center for Visual Sciences, University of Rochester, New York, USA

First Published on: 19 September 2007

To cite this Article: Karacan (üke) , Hacer and Hayhoe, Mary M. (2007) 'Is attention drawn to changes in familiar scenes?', *Visual Cognition*, 16:2, 356 - 374

To link to this article: DOI: 10.1080/13506280701453854

URL: <http://dx.doi.org/10.1080/13506280701453854>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## Is attention drawn to changes in familiar scenes?

Hacer (Üke) Karacan

*Department of Cognitive Science, Middle East Technical University,  
Ankara, Turkey*

Mary M. Hayhoe

*Center for Perceptual Systems, University of Texas at Austin, TX, and Center  
for Visual Sciences, University of Rochester, New York, USA*

In this study, we examined the mechanisms that control attention in natural scenes. We asked whether familiarity with the environment makes subjects more sensitive to changes or novel events in the scene. Previous investigation of this issue has been based on viewing 2-D images of simple objects or of natural scenes, a situation that does not accurately reflect the challenges of natural vision. We found that familiarity with the environment significantly increased the time spent fixating regions in the scene where a change had occurred. Together with previous work (Brockmole & Henderson, 2005a, 2005b), our results support the hypothesis that we learn the structure of natural scenes over time, and that attention is attracted by deviations from the stored scene representation. Such a mechanism would allow deployment of attention to objects or events that were not explicitly on the current cognitive agenda.

The capacity limits of attention and working memory set fundamental constraints on the way that the information in visual scenes is processed. Numerous experiments reveal that only a small fraction of the information in an image can be attended during a brief presentation, and retained in working memory. There is general consensus that only the “gist” of a scene is retained in working memory after a brief presentation, along with three or four items, or “object files”, and other higher level semantic information (Hollingworth & Henderson, 2002; Irwin & Andrews, 1996; Luck & Vogel, 1997). An important and relatively unexplored issue is how these limitations

---

Please address all correspondence to Mary M. Hayhoe, Center for Perceptual Systems, University of Texas at Austin, 1 University station, A8000, TX 78712, USA. E-mail: [mary@mail.cps.utexas.edu](mailto:mary@mail.cps.utexas.edu)

The authors wish to acknowledge the assistance of Jelena Jovancevic, Brian Sullivan, and Kelly Chajka in the design of the virtual environment and the conduct of the experiment. The work was supported by NIH grants EY05729 and RR09283.

play out in the context of vision in the natural world. How do we ensure that attention is directed to the right place at the right time? Given the limited bandwidth that attention sets on visual processing, this is a difficult problem that the visual system must solve.

It is often assumed that the solution to the problem is that attention is attracted by salient stimuli or events in the visual array. There is evidence that high spatial frequency content, edge density, and local contrast play a role in attracting fixations (Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Mannan, Ruddock, & Wooding, 1997; Parkhurst & Niebur, 2003; Reinagel, Godwin, Sherman, & Koch, 1999). It has also been demonstrated that visual saliency, based on image features such as colour, intensity, contrast, and edge orientation, can account for some of the variance in gaze distribution when viewing 2-D images (Itti & Koch, 2000, 2001; Koch & Ullman, 1985; Torralba, 2003). In addition, sudden onset stimuli have considerable ability to capture attention even if the observer's attention is directed elsewhere (Jonides & Yantis, 1988; Theeuwes, 1994; Yantis, 1998). Other stimuli, such as a unique colour or shape, or motion stimuli, may also capture attention (Chastain, Cheal, & Kuskova, 2002; Franceroni & Simons, 2003; Yantis & Hillstrom, 1994). Attentional capture may also be accompanied by oculomotor capture, where gaze is diverted to the novel event (Irwin, Colcombe, Kramer, & Hahn, 2000; Theeuwes, Kramer, Hahn, & Irwin, 1998). It is not clear whether such stimuli can invariably attract attention whatever the task (Folk, Remington & Johnston, 1992; Gibson & Jiang, 1998; Gibson & Kelsey, 1998; Turatto & Galfano, 2000; Yantis, 1998). However, Brockmole and Henderson (2005a, 2005b) found that abrupt onsets attracted a fixation when subjects viewed photographic images of natural scenes, regardless of whether they were explicitly told to search for a new object that might appear, or whether less specific memory instructions were given. These authors also found that new objects attracted gaze even when presented during a saccade, although less reliably and with longer latency than those presented during a fixation, as might be expected given the nature of the retinal transient signal produced when the eye is stationary.

All of the work showing stimulus-based effects of attentional or oculomotor capture has been done with 2-D experimental displays and either simple geometric stimuli or photographic renderings of natural scenes. There are reasons to suppose that attentional and oculomotor capture might not be entirely effective in the context of natural visually guided behaviour. Even when subjects are instructed to inspect images of natural scenes, the situation does not accurately reflect the challenges of visually guided behaviour in real, 3-D environments. Acting within a scene generates a very different retinal image sequence than inspecting 2-D images. Instead of a single stationary image, a complex image sequence is generated on the retina as the observer moves through the scene. Sudden onsets are rare in

real environments, and many kinds of information are important to the subject and need to be attended, not just sudden onsets. For example, observers need to be aware of irregularities in the pavement, or the location of obstacles. Consequently a mechanism that relied on sudden onsets for attracting attention is not likely to be very effective. Motion can effectively capture attention when presented in an otherwise stationary display (Franceroni & Simons, 2003), but might be masked in natural environments where retinal transients are continuously generated by the observer's motion through the environment (Jovancevic, Sullivan, & Hayhoe, 2006), and a variety of moving objects might be present, as in a city street. Thus, it remains problematical how observers distribute attention in an appropriate manner in natural vision.

The challenge of controlling attention and gaze in natural environments stems not only from the nature of the time-varying stimulus array, but also from the attentional demands of controlling action within the environment. Recent work in natural tasks has demonstrated that the observer's cognitive goals play a critical role in the distribution of gaze during ongoing natural behaviour (Hayhoe & Ballard, 2005; Land, 2004). In extended visuomotor tasks such as driving, walking, sports, and making tea or sandwiches, fixations are tightly linked, step-by-step, to the performance of the task (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land & Furneaux, 1997; Land & Lee, 1994; Land, Mennie, & Rusted, 1999; Patla & Vickers, 1997; Pelz & Canosa, 2001; Turano, Gerguschat, & Baker, 2003). Subjects exhibit regular, often quite stereotyped fixation sequences that are tightly linked, in time, to the actions, and very few irrelevant areas are fixated (Hayhoe et al., 2003; Land et al., 1999). This does not entirely solve the problem, however. In a stable environment such as a table top, a subject's behavioural goals can be achieved by fixating only the task-relevant objects. In other environments, however, such as driving or walking down the street, the goals are less well defined, and it is not always possible to anticipate what information is required. This has been described as the "scheduling problem" (Hayhoe, 2000; Jovancevic et al., 2006; Shinoda, Hayhoe, & Shrivastava, 2001). How does the visual system handle unexpected stimuli? There is some evidence that subjects can handle unexpected events by monitoring the environment in a strategic manner determined by the context (Rothkopf, Ballard, Sullivan, & de Barbaro, 2005; Sprague & Ballard, 2007). For example, drivers monitor the neighbourhood of intersections to locate stop signs, and walkers monitor other pedestrians' trajectories to avoid collisions (Shinoda et al., 2001; Jovancevic et al., 2006). This work suggests that observers can learn the dynamic properties of the environment in order to distribute gaze in an optimal manner (Hayhoe & Ballard, 2005; Hayhoe, Droll, & Mennie, 2007).

A different kind of solution to the problem is suggested by the finding that long-term memory for scene content is quite good (Hollingworth, 2004; Hollingworth & Henderson, 2002; Melcher, 2006; Melcher & Kowler, 2001). Few scenes are entirely novel, and observers make large numbers of fixations within a given scene. For example, within 5 min an individual will have made of the order of 1000 fixations. Over repeated exposures to real scenes, subjects may be able to build up quite elaborate long-term memory representations. It is well known that, in many circumstances, observers are insensitive to changes in scenes made during saccades, or when masked by some other kind of transient such as a blank screen or a blink (O'Regan, 1992; Rensink, 2000; Simons, 2000). This "change blindness" phenomenon has typically been interpreted to mean that very little information is retained in short-term visual memory from one fixation to the next as observers move their gaze around a scene (Henderson & Hollingworth, 1999). However, the finding that there may be robust long-term memory representations has important implications for how changes might be detected in more typical circumstances. A detailed long-term memory representation of a scene may be useful in guiding subjects' attention to changes. If the scene is efficiently coded in long-term memory, different mechanisms might be available for coding new information. Subjects may compare the current image with the stored representation in a manner similar to that suggested by Rao and Ballard (1999), and a mismatch, or "residual" signal may serve as a basis for detecting a change. Such mismatches might serve as a basis for attracting attention to changed regions of scenes. This may allow subjects to be particularly sensitive when there are deviations from familiar scenes, and thus attention may be drawn to regions that do not match the stored representation. Evidence for this was found by Brockmole and Henderson (2005b). When subjects were given 15 s pre-exposures to images of natural scenes, new objects were able to attract gaze in a subsequent brief exposure, even when the object was presented during a saccade, and there was no retinal transient associated with its appearance, and no opportunity to construct a short-term memory representation. The authors suggest that the preexposure allowed subjects to construct a long-term memory representation of the scene, as a basis for discriminating the new object. Thus, when the scene is familiar, changes may be more readily detectable.

In the present experiment, we investigated this issue further. In particular, we asked whether the effect of scene familiarity generalized to 3-D environments, when the observer is acting within the environment. As described above, it is necessary to observe active, visually guided behaviour in complex environments because the stimulus conditions and cognitive goals are so different from viewing 2-D images, even when those images represent natural scenes (Droll, Hayhoe, Triesch, & Sullivan, 2005; Hayhoe & Ballard, 2005; Jovancevic et al., 2006; Triesch, Ballard, Hayhoe, &

Sullivan, 2003; Wallis & Bühlhoff, 2000). Since real natural environments are hard to control, we devised an immersive virtual environment, where subjects walked along a footpath in the presence of a variety of stationary objects. We examined whether the opportunity to become familiar with the environment influenced the distribution of gaze, and, in particular, whether subjects preferentially fixate scene changes when they have become familiar with the environment. In order to understand the way that familiarity might improve detection of changes, we also manipulated the kind of changes that were made. Previous work has revealed attentional capture may occur with objects that vanish, as well as those that appear (e.g., Theeuwes, 1991), although offsets might be less effective than onsets (e.g., Boot, Kramer, & Peterson, 2005; Brockmole & Henderson, 2005a). In addition to new objects, and objects that were removed, we also examined changes in the position of objects, as well as replacing an existing object with a different one, since both these factors are relevant in change detection manipulations, and may reveal the nature of the detection process. Finally, we were interested in changes that did not engender a retinal transient. Since changes accompanied by a retinal transient are typically easier to detect than those that are not (see, e.g., Brockmole & Henderson, 2005b), this represents the most challenging situation for the observer, and, as described above, many of the situations where an observer needs to attend will not necessarily be accompanied by a transient signal.

## METHODS

### Apparatus

Subjects wore a Virtual Research V8 head mounted display. The helmet was equipped with a 3rdTech HiBall-3000 motion tracker. This is a high-precision analog/optical tracking system that tracks linear and angular motion (6 degrees of freedom) at 2000 Hz over a  $4.8 \times 6$  m region. The system latency for updating the scene conditioned on a movement of the Hiball was estimated as  $\sim 37$ – $49$  ms, depending on when the data was 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 ( $\sim 34$ – $46$  ms), were added with an estimate of 3 ms for the Hiball latency (timing 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 that subjects did not experience a noticeable lag between head motion and the visual update, or suffer any motion sickness that typically results from visual-vestibular conflict.

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 with  $640 \times 480$  pixel resolution each and a visual angle of  $48^\circ \times 36^\circ$ . An Applied Science Laboratory (ASL) 501 video-based eye tracker monitored the position of the left eye with 60 Hz temporal resolution and approximately one degree in accuracy. Eye position was calibrated by having the subject look at each of nine points on a  $3 \times 3$  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 superimposed, 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.

### Walking task

The virtual environment consisted of a portion of SGI Performer Town. Subjects walked along a rectangular footpath that encircled a monument in the virtual environment. The footpath corresponded to walking along four sides of the  $4.8 \times 6$  m experimental room, a distance of about 22 m. The dimensions of the virtual world were geometrically matched to the real world so that there was no visuovestibular or visuomotor conflict generated by movement through the scene. Objects in the environment were located along the outer side of the footpath (see Figure 1). The monument served the function of occluding the view until the subject turned the corner. Simple coloured robot-like figures acted as pedestrians in the environment. There were six pedestrians, two walking in the same direction and four in the opposite direction, at different speeds, so that their configuration varied continuously as the observer walked around the central monument.

In the environment, at each side of the monument there were two objects, making a total of eight (see Figure 2, which shows the layout of the objects in the scene, and the sequence of changes made during the trial). A trashcan, a mailbox, a billboard, and a house were stable objects that were located on the two sides of the monument where no change occurred. In addition, a water fountain, a street lamp, and another house were stable, but were located at the two sides of the monument which were subjected to change. For the first change, a gazebo was replaced with either a dog or a fire hydrant. Here, two different objects, of approximately equal size, were used in order to test whether the results were influenced by the specific properties of the replaced object (see Figure 2B). The second change was the

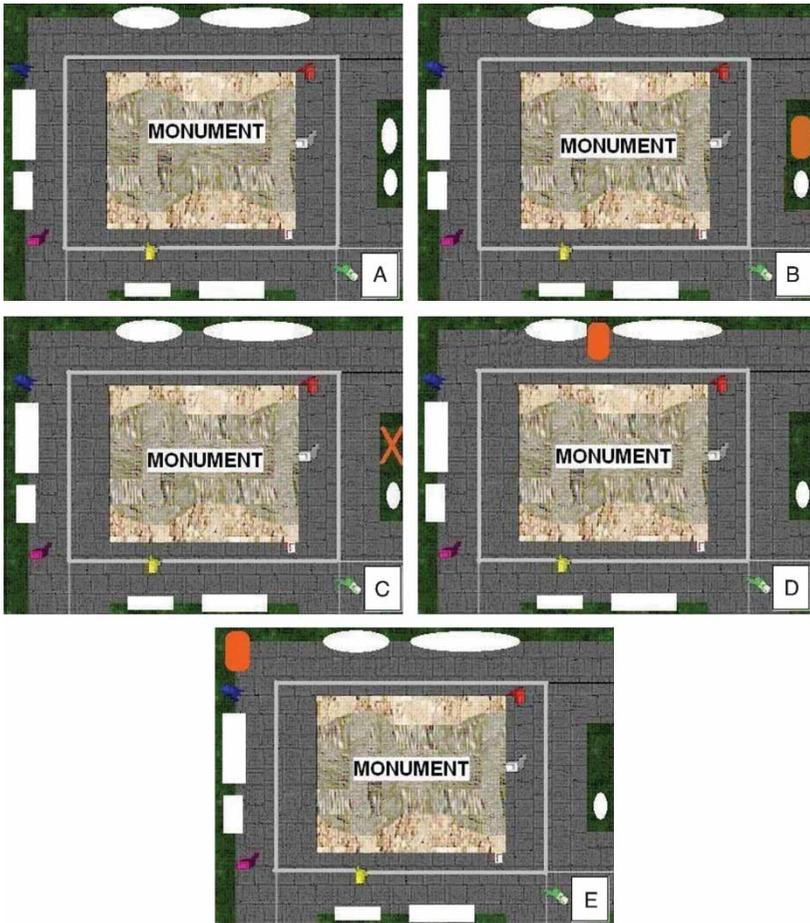


**Figure 1.** A snapshot of the environment. To view this figure in colour, please see the online issue of the Journal.

disappearance of this replaced object. This location remained unoccupied until the end of the trial (see Figure 2C–E). When a new object appeared, it was placed in between two existing objects as shown in Figure 2D. That object then moved to a new location on the corner on the subsequent circuit around the monument (see Figure 2E). The new object that appeared in the environment was again either a dog or a fire hydrant. Half of the subjects saw the dog as the new object (and the fire hydrant as the replaced object), while the other saw the fire hydrant instead (with the dog as the replaced object).

## Procedure

The experiment was conducted with 38 subjects who were undergraduate and graduate students at University of Rochester. Walking around the monument six times was counted as one trial. (Subjects always walked in the same direction.) The experiment was carried out in a between-subjects format; two groups were defined according to subjects' familiarity with the environment. Each group included 19 subjects. The first group was labelled the "inexperienced group" since they had no familiarization trials. The second group had three familiarization trials (i.e., 18 turns around the monument) in the experimental environment before the changes occurred,



**Figure 2.** (A) Bird's-eye view of the environment. White circles and white rectangles represent changing and stable objects respectively. (B) An object was replaced with another object (all changes are represented with an orange rectangle). (C) An object disappeared (the cross sign shows the previous location of the disappeared object). (D) A different object appeared at a new location. (E) The new object moved to a different location. To view this figure in colour, please see the online issue of the Journal.

and were labelled the “experienced group”. The subjects were sequentially allocated to one of these two groups.

The experienced group had a total of four trials: three familiarization trials, and one trial when the changes occurred. The inexperienced group had only one trial, when the same changes were made. The inexperienced group walked around the monument once before any changes occurred. (Thus, the inexperienced group had a small amount of experience, resulting

from the first circuit around the monument.) On the second circuit, the gazebo was replaced with either a dog or a fire hydrant. On the third circuit, the replaced object disappeared. On the fourth circuit, a new object (a fire hydrant if the replaced object was a dog, and vice versa) appeared in between the two existing objects. On the fifth circuit, this object moved to a new location that was previously unoccupied by an object. (No changes occurred on the final circuit.) The same sequence of changes was followed for the experienced group, but they occurred on the fourth trial, instead of the first.

Before the experiment, subjects signed a consent form and were shown the actual path they would walk in the laboratory. In order to ensure the safety of the subjects, a laboratory assistant walked with the subjects during the experiments while holding the connection cables of the system. The possible aftereffects of the usage of virtual environments were also explained, and they were asked to inform the experimenter if they had any discomfort during the experiment. (No subjects reported discomfort.) Subjects were instructed to become familiar with the environment while avoiding the pedestrians. The instructions were the same for both groups. The only difference between two groups was the number of turns around the monument before the changes occurred (a total of 19 for the experienced group, and 1 for the inexperienced group).

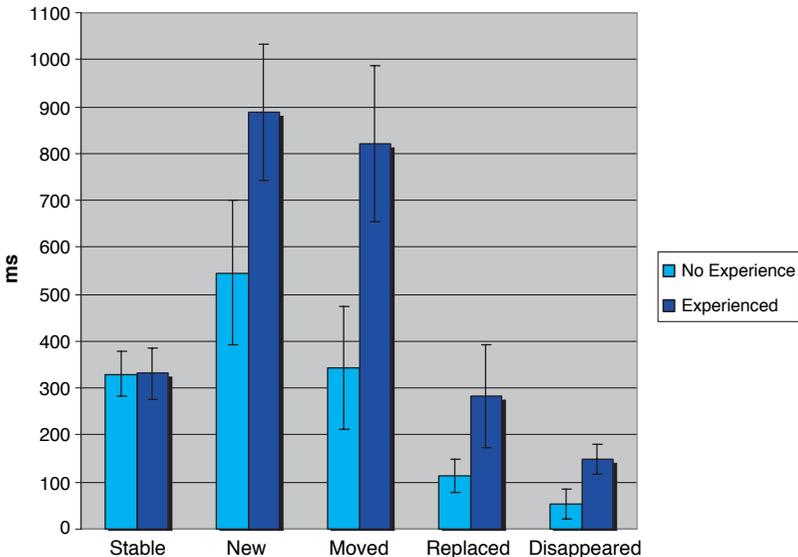
## Data analysis

All analyses were done on the video records which were analysed frame by frame to find the fixations. Gaze was divided up into fixations on the background, pedestrians, or the objects. (Fixations were defined as a constant location within a 1 deg radius for a period of 100 ms or more.) We were interested in the total amount of time fixating a particular object. This might be composed of a single long fixation or several shorter fixations on the object. We will refer to this as gaze duration to avoid confusion with the duration of a single fixation. The total duration of all fixations on stable objects was found, and also on all changing objects, separated by category of change. These values were divided by the number of circuits around the monument, and the number of objects to give the total time fixating a single object in one circuit around the monument, averaged over objects. For the object that disappeared, fixation on the object's prior location was measured. Only the trial where the changes occurred was analysed (i.e., Trial 4 for the experienced group and Trial 1 for the inexperienced group). All six laps in the trial were examined. The first lap was used only for calculating the average fixation durations for stable objects. Although there were some stable objects on side of the monuments where changes occurred,

only those stable objects on sides where no changes occurred were included in the analysis, in case there was some influence of the nearby changes.

## RESULTS

The average gaze duration on stable and changing objects is shown in Figure 3 for the two groups, for each category of change. Both groups had essentially identical fixation durations on the stable objects. However, subjects who had experience in the environment looked substantially longer at the changing objects. The average increase over the four different change conditions was approximately 170 ms,  $t(36) = -2.53$ ,  $p = .008$ , while no difference was observed between the two groups for stable objects,  $t(36) = -0.016$ ,  $p = .49$ . Comparing the groups for the different types of change, the effect of experience was significant for objects that were moved or that disappeared,  $t(36) = -2.27$ ,  $p < .05$ , and  $t(36) = -2.07$ ,  $p < .05$ , respectively, and close to significance for the new and replaced objects,  $t(36) = -1.61$ ,  $p = .057$ , and  $t(36) = -1.47$ ,  $p = .07$ , respectively. This suggests that prior experience in the environment indeed increased the likelihood that subjects would look at a change in the scene.

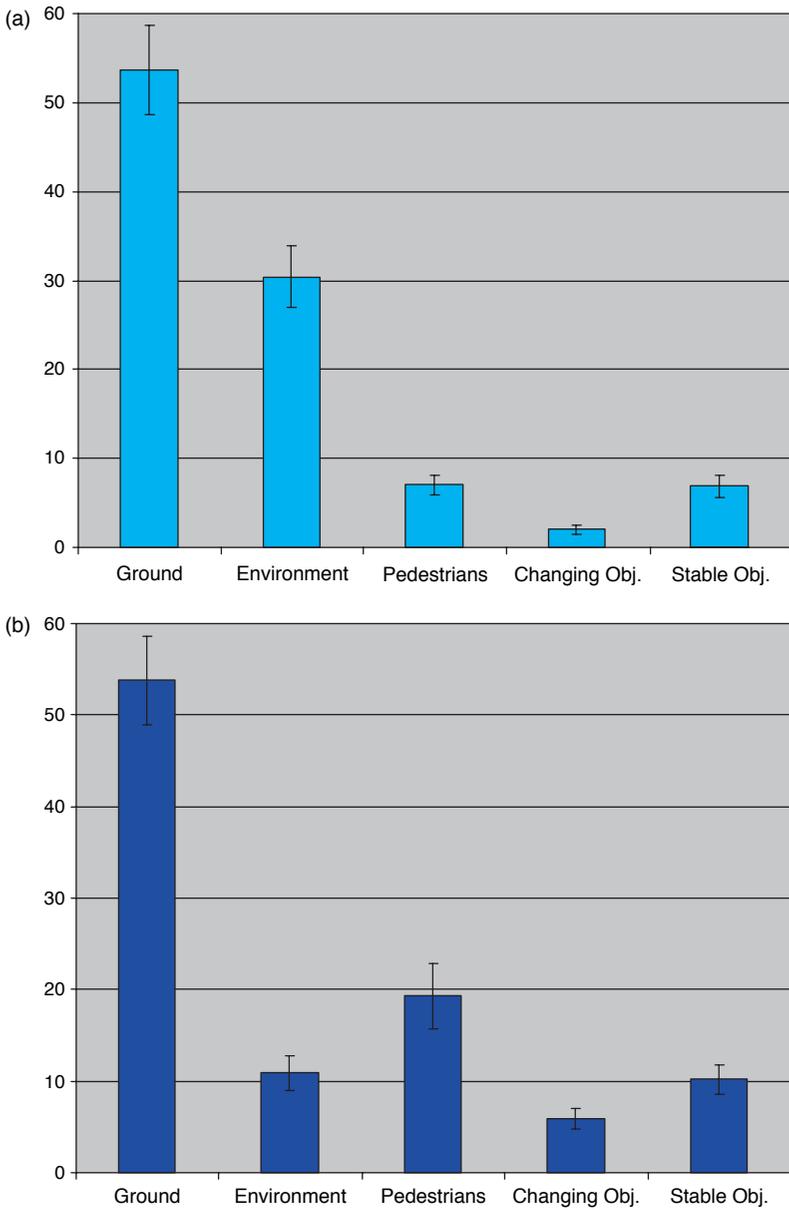


**Figure 3.** Gaze durations for different changes. Total time gaze fell on an object on a given circuit around the monument, for stable and changing objects, for experienced and inexperienced groups. Error bars are  $\pm 1$  standard error of the mean across subjects. To view this figure in colour, please see the online issue of the Journal.

The type of change also had a significant effect on fixation duration, for both experienced and inexperienced groups,  $F(4, 18) = 5.1$ ,  $p < .05$ , and  $F(4, 18) = 4.13$ ,  $p < .05$ , respectively. For the experienced group, the new and moved objects were fixated longer than stable objects,  $t(18) = 6.1$ ,  $p < .01$ , and also longer than the objects that were replaced or disappeared,  $t(36) = 2.57$ ,  $p < .05$ , and  $t(36) = 4.6$ ,  $p < .01$ , respectively. A similar ordering is observed for the inexperienced observers, although all fixation durations were less for that group. For both groups, objects that disappeared were fixated for shorter periods than their respective baselines for stable objects. For the inexperienced group, replaced objects were also fixated for shorter periods than stable objects,  $t(19) = 4.6$ ,  $p < .001$ . Note that no differences were observed depending on whether the dog or the fire hydrant was used as a novel object, so the data were collapsed over these conditions.

### Gaze distribution

We also examined how subjects distribute their gaze in the environment. The location of the fixations was classified into fixations on the path, surrounding environment (for example the grass, the monument, or in the distance), pedestrians, or objects, either changing or stable. The proportion of time spent fixating on each of these regions is plotted in Figure 4a and 4b for the inexperienced and experienced groups, respectively. Most of the fixations (54%) were located on the walking path for both experienced and inexperienced groups. This may reflect the ongoing demands of walking and staying on the path. The inexperienced group devoted 30% of the total gaze duration to the surrounding environment, but this dropped to 11% following experience in the environment. This difference is significant,  $t(36) = 2.024$ ,  $p = .050$ . Presumably these fixations contribute to the long-term memory representation of the environment. Interestingly, the experienced group devoted more gaze time to pedestrians than the inexperienced group,  $t(36) = -1.968$ ,  $p = .057$ . Thus, the distribution of fixations in the scene changes as a function of experience in the scene. The smallest amount of time was spent on the objects in both groups. Note that in this figure, the stable objects are accorded a greater percentage of the gaze distribution because there are more stable objects than changing ones. When gaze duration is appropriately normalized to give gaze duration per object, as described above in the Methods section, the data in Figure 3 shows the relative allocation of gaze to the changing versus the stable objects. As expected from the data in Figure 3, the difference between the two groups in the proportion of fixations on changing objects in Figure 4 is significant,  $t(36) = -2.53$ ,  $p = .008$ , and the difference for stable objects is not significant.



**Figure 4.** Fixation distribution of (a) inexperienced group, (b) experienced group. To view this figure in colour, please see the online issue of the Journal.

TABLE 1  
Number of the participants that detect the changes explicitly

<i>Changes</i>	<i>Object replaced</i>	<i>Object disappeared</i>	<i>Object appeared</i>	<i>Object moved</i>
Groups				
Inexperienced	1	1	8	4
Experienced	3	3	11	8

### Correlations between subject reports and fixations

After completing the experiment in the virtual environment, subjects were asked whether they had noticed any change that had occurred in the environment during the experiment. If their response was positive, they were asked to verbalize their explicit knowledge about those detected changes. Table 1 shows the number of participants in each group that detected the different kinds of changes explicitly. Since there are 19 subjects in each group, clearly many of the changes went unnoticed.

A score for each subject was calculated by summing up the number of correct detections. Since there were four changes, these scores varied from 0 (i.e., no change was detected) to 4 (i.e., all changes were detected). When these scores (i.e., number of changes noticed) were compared with the average fixation durations on the changing objects we found that there was a significant correlation between two, Spearman's  $\rho = 0.586$ ,  $N = 38$ ,  $p < .01$ . We also investigated the different change types. This time we compared the fixation durations of subjects who detected the changes with the ones who did not. The results showed that the fixation durations on the novel object,  $t(36) = 2.6$ ,  $p = .116$ , the moved object,  $t(36) = 2.828$ ,  $p = .101$ , and the disappearing object,  $t(36) = 0.007$ ,  $p = .932$ , did not differ significantly depending on whether the subject explicitly detected the change or not. However, there was a significant difference for the replaced object,  $t(36) = 31.411$ ,  $p < .01$ . This result bears further investigation. In summary, overall awareness of the changes were quite low, and fixations were not always accompanied by an awareness of the change. Thus, the fixation duration may be a more sensitive indicator of attentional prioritization than explicit awareness, as has been observed in previous studies (Jovancevic et al., 2006; Shinoda et al., 2001).

### DISCUSSION

The results of this experiment reveal that familiarity with the environment is an important factor in the distribution of gaze in the environment. In particular, subjects spend more time fixating the changed objects if they are

familiar with the environment. Novel and displaced objects were most effective in attracting gaze, and subjects spent approximately 500 ms longer fixating these objects than stable objects. The similarity of the data for displaced and novel objects presumably results from the appearance of the displaced object in a previously unoccupied location. When a new object appeared in a previously occupied location, fixation times were much less. These results suggest that the appearance of objects in novel locations is indeed an important factor in attracting both attention and gaze in the context on ongoing natural behaviour, consistent with previous studies with nonimmersive displays (Brockmole & Henderson, 2005a, 2005b; Theeuwes et al., 1998). It is also important to note that these changes occurred while the objects were out of the field of view of the subject, and so were not accompanied by a retinal transient. Thus, their attentional prioritization must be a consequence of the difference between the current image and the subject's stored memory representation of the scene. This is consistent with Brockmole and Henderson's (2005a, 2005b) result showing that changes occurring during a saccade had the power to attract fixations when subjects had the opportunity to construct a memory representation of the scene during a prior 15 s exposure to the image.

The present results, together with those of Brockmole and Henderson (2005a, 2005b), point to a powerful factor in controlling attention in the context of natural behaviour. Most environments are familiar. Even environments that are novel will contain many statistical similarities with environments that have been experienced in the past (for example, street scenes, classrooms, stairways, and halls). Thus, a mechanism that draws attention to regions that do not fit stored memory representations will be generally useful. What are the mechanisms by which this attraction occurs? Some insight may be given by models of early cortical processing where feedback signals from higher cortical areas carry a model-based prediction to lower areas (Rao & Ballard, 1999). When there is a mismatch between the input and top-down predictive signals, the residual signal is transmitted to higher cortical areas and generates a revised prediction. If the predictive signal is based on a stored memory representation of the scene, scene changes may generate a mismatch, or residual signal, that prompts a reevaluation of the scene, and may thereby attract attention. In the current volume, Brockmole and Henderson show that semantically inconsistent objects were fixated sooner than consistent objects. Since stored scene representations presumably contain high-level semantic information, an inconsistent object would generate a bigger mismatch signal. Thus, Brockmole and Henderson's recent finding provides additional support for a mismatch mechanism. This kind of a mechanism seems likely to be more robust than simple saliency-based models such as Itti and Koch (2000), where intrinsic stimulus properties are hypothesized to attract attention.

Such models are inflexible, and do not account for more than a small proportion of the fixations observed in natural behaviour (Hayhoe & Ballard, 2005; Jovanevic et al., 2006; Land, 2004; Rothkopf & Ballard, in press). A mismatch mechanism, such as that indicated here, can do the job of saliency-based models by allowing attentional deployment to regions that are not part of the ongoing task, and where something unexpected might occur. Such a mechanism is not easily classified as top-down or bottom-up, and can serve as a useful adjunct to attentional deployment that is governed by the immediate cognitive agenda. It seems close to the idea of Bayesian “surprise” that Itti and Baldi (2005) suggest might serve as a mechanism for attracting gaze. However, Itti and Baldi calculate surprise with respect to stimulus properties, whereas we are suggesting that surprise is measured with respect to a memory representation. This is of course speculative, but the robustness of the effects observed in the current experiment suggests that a model of this general kind is required.

There are some slightly puzzling aspects of the results. When objects are removed from the scene, the time spent fixating the region where the object had been was less than the time spent fixating stable objects. This implies that objects that disappear *are not prioritized by attention* in the absence of retinal transients, consistent with Brockmole and Henderson (2005a). On the other hand, more time was spent fixating these locations when the subject was familiar with the environment. This effect of familiarity suggests that the disappearance of the object *does indeed lead to attentional prioritization*. A similar result holds for objects that are replaced, although the difference between the groups is less reliable. It is not clear why the inexperienced subjects looked at the replaced object less than stable objects. It is possible that there is some uncontrolled factor, such as object size, that accounts for the reduced likelihood of a fixation on the replaced object. Note that the billboard and house, two of the stable objects, are much larger than the dog or the fire hydrant, consistent with this speculation. This, and other idiosyncrasies of the design, limit the conclusions about the specific difference we observed between the different types of change. For example, the object that moved to a new location was the same object that was novel in the preceding circuit. A fixation on that object when it moved might therefore not be independent of the events on the previous lap.

The results also have significance for understanding change blindness. Poor performance in change blindness experiments has typically been interpreted as evidence for limited scene memory (O’Regan, 1992; Rensink, 2000; Simons, 2000). Change blindness clearly reflects the limitations of short-term memory, although it is now clear that visual long-term memory for scenes is quite extensive (Hollingworth, 2004; Hollingworth & Henderson, 2002; Melcher, 2006; Melcher & Kowler, 2001). Why, then, have the demonstrations of change blindness been so effective? The present results,

and those of Brockmole and Henderson (2005a, 2005b), suggest that part of the explanation is that experiments on change blindness almost invariably present observers with unfamiliar scenes. Since observers have no stored model of what the scenes *should* look like, observers are forced to deploy attention serially through the scene and use limited capacity working memory. This explanation would account for missing certain, somewhat arbitrary, changes such as a man's trousers changing colour. Additional evidence for the importance of familiarity with the scene comes from work by Droll, Hayhoe, Triesch, and Sullivan (2004). These authors made changes to the colour of one of a set of red and blue virtual blocks that the subject was required to sort. When a block changed colour from red to blue, or vice versa, subjects were relatively insensitive to the colour change. Changing the block colour to one not present in the scene, such as green, was immediately apparent. This was true whether or not other blocks were present in the current view. This supports the attention-getting power of a stimulus that differed from the scene that observers were familiar with in the experimental context, which contained only red and blue blocks. These observations may help explain our intuitions that vision is generally effective in allowing us to see what we need to see in the natural world, despite the impressive demonstrations of change blindness in the context of experiments. It is not necessary to rely entirely on such a mechanism, however. As mentioned above, subjects also deploy vision actively to search for behaviourally important information. Subjects were very good at detecting briefly presented Stop signs in a virtual driving environment if the Stop signs were located at intersections. They were less effective at detecting signs in less likely locations, such as in the middle of a block (Shinoda et al., 2001). Similarly, subjects appear to detect potential collisions by actively monitoring other pedestrians in peripheral vision (Jovancevic et al., 2006). Note that a substantial proportion of the subjects' fixations fell on pedestrians, rather than on the stationary objects, in the present experiment (Figure 4). Thus, the distribution of gaze in a scene is strongly influenced by both the subject's learnt agenda, and deviations from previously learnt state.

In summary, we showed that subjects' familiarity with a 3-D immersive virtual environment significantly increases the probability that gaze will be attracted to changes in the scene. These results are consistent with the results of Brockmole and Henderson (2005a, 2005b, this issue), which were conducted with 2-D images. The results support the hypothesis that we learn the structure of natural scenes over time, and that attention and gaze are attracted by deviations from the normal state.

## REFERENCES

- Ballard, D. H., & Rothkopf, C. (2005). Learning visual representations with projection pursuit [Abstract]. *Journal of Vision*, 5(8), 373a.
- Boot, W. R., Kramer, A. F., & Peterson, M. S. (2005). Oculomotor consequences of abrupt onsets and offsets: Onsets dominate oculomotor capture. *Perception and Psychophysics*, 67, 910–928.
- Brockmole, J. R., & Henderson, J. M. (2005a). Object appearance, disappearance, and attention prioritization in real-world scenes. *Psychonomic Bulletin and Review*, 12, 1061–1067.
- Brockmole, J. R., & Henderson, J. M. (2005b). Prioritization of new objects in real-world scenes: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 857–868.
- Brockmole, J. R., & Henderson, J. M. (2008). Prioritizing new objects for eye fixation in real-world scenes: Evidence from eye movements. *Visual Cognition*, 16(2/3), 375–390.
- Chastain, G., Cheal, M., & Kuskova, V. (2002). Inappropriate capture by diversionary dynamic elements. *Visual Cognition*, 9, 355–381.
- Droll, J., Hayhoe, M., Triesch, J., & Sullivan, B. (2004). Working memory for object features is influenced by scene context [Abstract]. *Journal of Vision*, 4(8), 152a.
- Droll, J., Hayhoe, M., Triesch, J., & Sullivan, B. (2005). Task demands control acquisition and maintenance of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1416–1438.
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 1030–1044.
- Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception and Psychophysics*, 65, 999–1010.
- Gibson, B. S., & Jiang, Y. (1998). Surprise! An unexpected color singleton does not capture attention in visual search. *Psychological Science*, 9(3), 176–182.
- Gibson, B. S., & Kelsey, E. M. (1998). Stimulus-driven attentional capture is contingent on attentional set for displaywide visual features. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 699–706.
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4), 188–193.
- Hayhoe, M., Droll, J., & Mennie, N. (2007). Learning where to look. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 641–660). Oxford, UK: Elsevier.
- Hayhoe, M. M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, 7, 43–64.
- Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1), 49–63.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, 50, 243–271.
- Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 519–537.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 113–136.
- Irwin, D. E., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance*

- XVI: Information integration in perception and communication* (pp. 125–155). Cambridge, MA: MIT Press.
- Irwin, D. E., Colcombe, A. M., Kramer, A. F., & Hahn, S. (2000). Attentional and oculomotor capture by onset, luminance and color singletons. *Vision Research*, *40*, 1443–1458.
- Itti, L., & Baldi, P. (2005). Bayesian surprise attracts human attention. In *Advances in neural information processing systems* (Vol. 19, pp. 1–8). Cambridge, MA: MIT Press.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neurosciences*, *2*, 194–203.
- Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception and Psychophysics*, *43*, 346–354.
- Jovancevic, J., Sullivan, B., & Hayhoe, M. (2006). Learning gaze allocation priorities in complex environments [Abstract]. *Journal of Vision*, *6*(6), 480a.
- Koch, C., & Ullman, S. (1985). Shifts in visual attention: Towards the underlying circuitry. *Human Neurobiology*, *4*, 219–227.
- Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. *Spatial Vision*, *13*, 201–214.
- Land, M. (2004). Eye movements in daily life. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (Vol. 2, pp. 1357–1368). Cambridge, MA: MIT Press.
- Land, M., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society. Series B: Biological Sciences*, *352*, 1231–1239.
- Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311–1328.
- Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*, 742–744.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.
- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision*, *11*, 157–178.
- Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, *6*, 8–17.
- Melcher, D., & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Research*, *41*, 3597–3611.
- O'Regan, J. K. (1992). Solving the “real” mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*(3), 461–488.
- Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, *6*, 125–154.
- 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: An International Journal for the Rapid Communication of Research in Neuroscience*, *8*, 3661–3665.
- Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, *41*, 3587–3596.
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.
- Reinagel, P., Godwin, D., Sherman, S. M., & Koch, C. (1999). Encoding of visual information by LGN bursts. *Journal of Neurophysiology*, *81*, 2558–2569.
- Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*, 17–42.
- Rothkopf, C., & Ballard, D. (in press). Task and context determine where you look. *Journal of Vision*.

- Rothkopf, C., Ballard, D., Sullivan, B., & de Barbaro, K. (2005). Bayesian modeling of task dependent visual attention strategy in a virtual reality environment. *Journal of Vision*, *5*, 920.
- Shinoda, H., Hayhoe, M., & Shrivastava, A. (2001). What controls attention in natural environments. *Vision Research*, *41*, 3535–3545.
- Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, *7*, 1–16.
- Sprague, N., & Ballard, D. (2007). Modeling embodied visual behaviors. *ACM Transactions in Applied Perception*, *4*(2), Article 11.
- Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception and Psychophysics*, *49*, 83–90.
- Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 799–806.
- Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*, 379–385.
- Torralba, A. (2003). Contextual priming for object detection. *International Journal of Computer Vision*, *53*, 169–191.
- Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, *3*(1), 86–94.
- Turano, K. A., Gerguschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research*, *43*, 333–346.
- Turatto, M., & Galfano, G. (2000). Color, form and luminance capture attention in visual search. *Vision Research*, *40*, 1639–1643.
- Wallis, G., & Bülthoff, H. (2000). What's scene and not seen: Influences of movement and task upon what we see. *Visual Cognition*, *7*, 175–190.
- Yantis, S. (1998). Control of visual attention. In H. Pashler (Ed.), *Attention* (pp. 233–256). Hove, UK: Psychology Press.
- Yantis, S., & Hillstrom, A. P. (1994). Stimulus-driven attentional capture: Evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 95–107.