

Gaze patterns in navigation: Encoding information in large-scale environments

Sahar N. Hamid

Department of Psychology, Glyndwr University,
United Kingdom



Brian Stankiewicz

Department of Psychology,
University of Texas at Austin, USA



Mary Hayhoe

University of Texas at Austin, USA



We investigated the role of gaze in encoding of object landmarks in navigation. Gaze behavior was measured while participants learnt to navigate in a virtual large-scale environment in order to understand the sampling strategies subjects use to select visual information during navigation. The results showed a consistent sampling pattern. Participants preferentially directed gaze at a subset of the available object landmarks with a preference for object landmarks at the end of hallways and T-junctions. In a subsequent test of knowledge of the environment, we removed landmarks depending on how frequently they had been viewed. Removal of infrequently viewed landmarks had little effect on performance, whereas removal of the most viewed landmarks impaired performance substantially. Thus, gaze location during learning reveals the information that is selectively encoded, and landmarks at choice points are selected in preference to less informative landmarks.

Keywords: navigation, localization, landmarks, gaze patterns, information loss

Citation: Hamid, S. N., Stankiewicz, B., & Hayhoe, M. (2010). Gaze patterns in navigation: Encoding information in large-scale environments. *Journal of Vision*, 10(12):28, 1–11, <http://www.journalofvision.org/content/10/12/28>, doi:10.1167/10.12.28.

Introduction

Large-scale environments contain abundant perceptual information such as landmarks or structural cues that are potentially available for navigation. A large-scale space is one where all the perceptual information is not visible from one vantage point. In these spaces, landmarks can aid navigation by specifying where and when an action needs to be taken (Vinson, 1999), and there is extensive evidence for the use of landmarks in a variety of different environments and situations (Tom & Denis, 2003).

A *landmark* can be defined in many different ways. It can be seen as a reference point in an environment (for example, turn right at the water fountain; Lynch, 1960) or distinctive visual features in an environment that can help identify a particular location (e.g., when you see the water fountain you are in the center of the city; Siegel & White, 1975). Landmarks can be divided into object landmarks and structural landmarks. Object landmarks are visual objects present in an environment that are not part of the environment's structure (e.g., water fountain). Structural landmarks are the geometrical visual cues found in the environment (e.g., T-junction, L-junction, dead end; Stankiewicz & Kalia, 2007). A large body of work has shown that the presence of object landmarks in an environment helps in navigation and orientation performance (e.g., Abu Ghazze, 1996; Darken & Sibert, 1996;

Foo, Warren, Duchon, & Tarr, 2005; Ruddle, Payne, & Jones, 1997; Spetch et al., 1997; Vinson, 1999).

Shortcut paradigm studies using animals or humans have also shown the significance of landmarks. Gould and Gould (1982) trained honeybees to fly consistently between their hive and two feeding points. When captured and displaced while en route to a particular feeding point, trained bees were seen to be able to take shortcuts to the right feeder location if sun-related compass information or prominent landmarks were visible from their displaced position (Dyer, 1991; Dyer, Berry, & Richard, 1993; Menzel, Chittila, Eichmuller, Pietsch, & Knoll, 1990; Wehner, Bleuler, Nievergelt, & Shah, 1990). In humans, Foo et al. (2005) looked at shortcut-taking ability in virtual reality environments. When stable landmarks were available, humans used them to make accurate novel shortcuts. When landmarks were unavailable, subjects were still able to complete novel shortcuts, but performance was comparatively poor, with less accuracy and higher variability in creating novel shortcuts. Results such as these provide confirmation that humans, bees, rodents, and other animals appear to rely heavily on landmarks when they are available for successful navigation (Dyer, 1991; Dyer et al., 1993; Menzel et al., 1990; Wehner et al., 1990). Landmark information appears to be used from the first learning trial to refine path integration. For example, Foo et al. (2005) showed, in an immersive virtual environment with colored posts as landmarks, that

reliance on visual landmarks was immediate and continued throughout the training period. A large body of additional work also shows immediate reliance on landmarks (Collet, Collet, & Wehner, 2001; Collet & Graham, 2004; Etienne, Boulens, Maurer, Rowe, & Siegrist, 2000; Etienne, Maurer, Boulens, Levy, & Rowe, 2004; Foo et al., 2005; Kearns, Warren, Duchon, & Tarr, 2002; Riecke, van Veen, & Bulthoff, 2002).

Landmarks differ in the kind of information they provide. Ruddle et al. (1997), showed that participants use highly informative object landmarks (3D models of everyday objects, e.g., cup, fork) when they are present. On the other hand, participants were seen to have difficulty in using less informative and meaningless landmarks (colored abstract paintings) as aids for successful navigation. They showed that not all landmarks are used equally when learning to navigate in a new environment. If there are any limitations to the amount of information that is encoded and stored in the brain from a visual scene that contains extensive information, landmarks may compete with one another to be stored in memory, and there must be some mechanism for selection of the information that is retained in the brain (Land & Lee, 1994; Pelz & Canosa, 2001; Turano, Geruschat, & Baker, 2003). Evidence for competition between landmarks comes from studies of rats. Once rats have learned to navigate toward a defined goal with reference to specified landmarks, they are slow to use new landmarks when they are added to the environment (Rodrigo, Chamizo, McLaren, & Mackintosh, 1997; Sanchez-Moreno, Rodrigo, Chamizo, & Mackintosh, 1999). Difficulty in replacing already stored object landmarks with newer object landmarks suggests that only a subset of the available spatial information is transferred to the stored representation of the environment.

While some selectivity of the available information is often implicit in navigational studies, direct evidence for selectivity was demonstrated by Nadeem (2008). This study measured the limitations in the transfer of information from a large-scale environment to an individual's internal cognitive representation of space, using environments of varying size and complexity. Subjects learned to navigate through large-scale environments of varying complexity and size followed by a localization task. Using an information theoretic measure (Shannon, 1993) of channel capacity, she showed that a constant proportion of information was lost, regardless of the size of the environment. Even in the simplest, smaller environments, participants did not encode all the information available to them. This leads to the question as to what available information from a large-scale space is encoded to form a mental representation of the environment. Which landmarks in a large-scale space are selected for encoding by participants or are all landmarks encoded equally efficiently? If there is selective encoding, what makes some landmarks more salient for storage compared to others?

To address these questions, we used eye movements as an indicator of what information is attended by participants learning to navigate in a new large-scale space. To date, gaze behavior has not been studied while learning to navigate a large-scale 3D space to look at landmark encoding selectivity in humans. Research in eye movements in natural environments shows that gaze patterns are overwhelmingly driven by the subject's behavioral goals (Hayhoe & Ballard, 2005; Jovancevic-Misic & Hayhoe, 2009; Land, 2004). Thus, it is likely that the gaze locations observed during navigation or while learning an environment indicate the locations that are being attended, and thus those that are most likely to be encoded on memory. Evidence from viewing 2D images of scenes also indicates that the fixated locations are preferentially stored in memory (Hollingworth & Henderson, 2002). Eye movements should indicate whether there is a selective pattern in how attention is distributed and which landmarks are attended. The fixated locations will indeed have been attended, as it is known that an attentional shift always precedes a shift in gaze location (Kowler, Anderson, & Blaser, 1995; Schneider, 1995). Fixations cannot give the complete story, however, as objects in the peripheral vision may also be attended and encoded without a direct fixation. However, the tight link between gaze locations and attention in natural tasks suggests that gaze will act as an informative indicator for much of the information being attended to. The goal of the current paper, therefore, is to explore the usefulness of gaze as an indicator of how people explore a large-scale environment, and what information is selected for memory storage.

Previous work suggests that the location of the landmark may determine whether it is selectively attended and stored. In a virtual driving environment, Aginsky, Harris, Resnick, and Beusmans (1997) made changes to landmarks showing that landmarks at decision points, that is, areas where decisions have to be made about changing direction, are important. They showed that subjects were more sensitive to changes made in these landmarks, indicating that landmarks at decision points are better remembered than those at other locations. On the other hand, Foo et al. (2005) showed that stable, reliable landmarks were used as navigational aids whether they appeared at the end, beginning, or middle of novel shortcut route. Thus, landmarks other than those at decision points also appear to be encoded. The current research therefore investigates whether subjects preferentially allocate gaze to landmarks at decision points and whether the fixated landmarks are selectively encoded in memory. To do this, we measured gaze patterns when subjects learn a new large-scale environment. We examined how a participant's localization ability in an environment is affected when half of the landmarks are removed from the scene. We removed either those landmarks that had been preferentially fixated during training, or those

that had not been fixated in order to test whether the preferentially fixated landmarks had also been preferentially encoded. If removing the fixated landmarks impairs performance, this would imply that gaze patterns reveal the encoded landmarks. It would also reveal selective encoding of object landmarks when learning to navigate in a new large-scale environment.

Methods

Participants viewed maze-like environments on a 16-inch computer screen from a first person perspective as seen in Figure 1. The display was viewed from a distance of about 100 cm and subtended approximately 20 deg of visual angle. Environments with 10 corridors were generated on a Cartesian grid such that hallways intersected at 90 degree angles (as seen in Figure 2). These environments were rendered using Vizard software. The model of the environment had 8-ft ceilings, and the length of each corridor segment was 24 ft. The viewpoint was centered laterally, from a height of 5 ft 3 in. The corridor width was slightly greater than the height (9.3 ft). When displayed on the computer monitor, however, because of the relatively small visual angle of the monitor, the scene was compressed by about a factor of five from a comparable scene in a real environment. The texture of the ceiling and walls was the same (cement) whereas the floor had a different texture (burlap) as seen in Figure 1. Participants were initially allowed to explore and learn the environment followed by a localization task in which they were tested using views from the environment.

Landmarks were placed at a uniform distance all along the corridors. Pictures of common objects were used as landmarks and none of the pictures were repeated making the object landmarks highly informative. The images used as object landmarks were chosen from a list of nouns that



Figure 1. A view of a straight corridor ending in an L-junction with landmarks from the virtual environment.

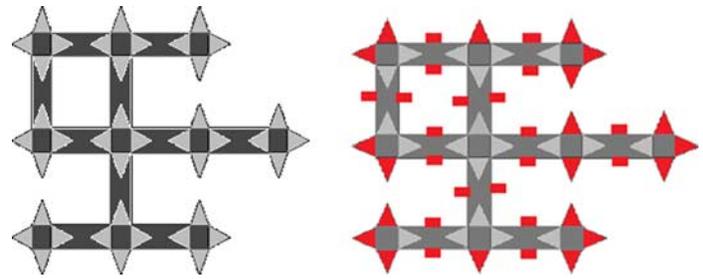


Figure 2. (a) A map of the 10-corridor environment. The corridors are in dark gray and the views are in light gray and indicate the locations and orientations tested in each environment. (b) The red marks indicate the locations of the 40 landmarks. Landmarks are placed at a uniform distance all along the corridors.¹

are easily recognized by individuals. No attempt was made to equate saliency of the individual images, but all were colored, and high contrast, and likely to be highly salient relative to the low contrast background textures in the environment. Landmarks were easily visible as they were placed in a box that protruded from the wall by a little over 25% of the width of the hallway, as seen in Figure 1. For each subject, the placement of specific pictures used as landmarks was randomized so that the properties of the picture itself did not affect data. The triangles in Figure 2 denote the views (location and orientation) that the participants were tested on. Views from all four directions were tested at each junction.

An Eyelink II eye tracker by SR Research was used to record the participant's eye movements throughout the experiment.¹ To calibrate, participants had to look at all 9 points in a 3×3 grid. The calibration process was repeated after every phase of the experiment. For improved accuracy in eye movement recording, a chin rest was used to keep the participant's head steady and at a uniform distance from the monitor screen.

Procedure

Fourteen students from the University of Texas at Austin participated in the study. They were paid \$10 for their participation. Different groups of subjects were used in the different sized environments. For each size environment, the subjects were divided into two groups for testing.

Training phase

In the *training phase*, participants started at a specified starting position in the environment. They were told to explore and try to learn the entire environment so that they could navigate easily within it. Participants were not given any details about what kind of navigating tasks would follow; they were just asked to learn the environment as they would when they are learning any new large-scale

space. The participants used keys to move through the environment. By pressing key “8,” participants could move forward for the distance of one corridor and the view for the participants changed in discrete steps. The participants could also rotate 90 degrees clockwise (key “6”) or counterclockwise (key “4”) at any position.

There were “hotspots” spread out all over the environments. The hotspots were positions in the maze that had a specific ID number and were tested during the testing phase. The hotspots were numbered randomly so as not to provide additional cues for the direction of the route. Hotspot positions were randomized and spread across the entire environment to make sure that the participants actually explored the entire environment. They were simply there to help the participants learn the environment in its entirety and they could be used as cues in the instructions when testing how well the participants had learnt the environment. Hotspots were specific locations regardless of the orientation, so a participant could have different views standing at one hotspot. The hotspots were placed so that each long hallway segment was covered, making sure that the participant explored the entire environment to cross all hotspots. When the participants moved to a hotspot, an auditory cue stated the hotspot ID (e.g., *Position One*). Each participant was allowed a specific number of forward moves during exploration after which the training phase ended. The number of forward moves allowed during exploration was dependent upon the size of the environment (5 times the number of corridors in the environment, e.g., 50 forward moves for the 10-corridor environment). Once the training phase ended, the participant went through to the *testing phase* to check how well the environment had been learned. If the participants were unable to pass the testing phase, they would return to the *training phase* for more exploration. For the 10-corridor environment, participants were able to learn the environment in one *training phase*.

Testing phase

This phase tested the participant’s knowledge of the environment. The participants started off at a random hotspot location in the environment. An auditory cue announced the starting location and also instructed the participant to go from the current position to a specified hotspot (e.g., *Position Eight, Go To Position Four*). Once the participant reached the specified target location, they were instructed to move on to another hotspot location. If a participant passed over any other hotspot on the way to the current goal, the computer provided an auditory cue stating the hotspot ID (e.g., *Position Five*). Participants were instructed to take the most direct route from hotspot to hotspot, showing their knowledge of the environment. Efficiency for reaching the target state from the starting state was measured by comparing the number of moves it took the subject to reach the goal state against the

minimum number of translations required for the navigation. The subject had to complete five consecutive tests in a row that were better than 80% efficiency.² If the participant did not pass, then a message would appear asking the person to return to the *training phase*. Otherwise, they moved on to the *experimental phase*.

Experimental phase

The *experimental phase* consisted of a localization task; that is a randomly selected view from the environment was shown on the computer screen and the location of the view had to be marked on the map of the environment. In this phase, the participants could not move in the environment; they could just view the snapshot view from the environment. The participants were not told what the task would be prior to this. During the *training phase*, participants were instructed to learn the environment so that they could complete any kind of navigation task. The *experimental phase* was divided into two different versions (*Full* and *Half*) that each participant completed. In the *Full* version, all landmarks were visible in the shown view, whereas in the *Half* version, half of the landmarks were removed based on the individual’s independent gaze patterns. Hence, different landmarks were removed for each participant based on their own particular fixation pattern. In the *Full* version, the subjects were shown a specific view (as seen in [Figure 3a](#)) from the explored environment for 1.5 s after which the map of the environment as shown in [Figure 3b](#) would appear. This is the first time the participants saw a holistic map of the environment. The map remained on the screen until the participant made a response. There were no time constraints so that the participants could match up their own mental representation of the environment with the map on the screen. On the other hand, the specific view from the environment was timed out so that participants could not use problem solving techniques based on geometric cues to localize themselves. The participant’s task was to mark the location and orientation of where the view was located. They could click on any one of the small triangles on the map, which then turned red. The map as seen in [Figure 3b](#) denotes the views being tested. The participants were only allowed to mark one spot at a time on the map after which they clicked on the “Save Data” button and moved on to the next trial. Each state was tested 5 times in a random order. All landmarks were present in the tested view.

The study was completed in two sessions on consecutive days. During the first session, the *training* and *testing phases* were completed. During the second session, the *training phase* was repeated to make sure that the participant remembered the environment and then continued on with the *experimental phase*.

Gaze data were analyzed using an in-house program that compared the dimensions for the landmark in the



Figure 3. (a) A view of a straight corridor ending in a dead end from the environment. (b) A map of the same environment. Highlighted red triangles are the possible positions simply based on geometric cues that the view could be from on the map of the environment.

environment that were matched to the coordinates received from the eye tracker. If the gaze coordinates fell within a 1 deg region centered on the landmark, this was counted as a landmark fixation. Fixations were not explicitly extracted.³ The experimenter reviewed the video records to validate the automated encoding. The total fixation time for each landmark was calculated separately for each participant during the *training* and *testing phases*. From each tested view, the total time spent looking at all visible landmarks was averaged according to the gaze movements of each participant. Each participant's gaze pattern was then used to decide which landmarks would be removed during the *experimental phase*.

Of the 40 landmarks, 20 landmarks were removed in the *Half* condition. For each subject, the median viewing time was individually calculated. The landmarks that were removed for each subject were either those that had been viewed less than the median time, or those that had been viewed more. The High-LM-Removed groups were tested in an environment where the only landmarks present were the ones that were viewed *less* than the median viewing time. Low-LM-Removed group participants were tested in an environment that had only the landmarks with *more* than the median viewing time during the *testing* and *training phases*. Both groups were also tested in an environment with all the landmarks (*Full*). The order of *Full* vs. *Half* (where *half* the landmarks were removed) was counterbalanced across subjects. The results showed no significant effect of counterbalancing the testing order.

Results

If there is selective encoding of landmark information and if fixations reveal this selection, then removing highly viewed landmarks should have a significant effect on performance. However, removing landmarks that are rarely viewed should affect performance less, as they are less likely to be encoded. A preference for any of the object landmarks in the environment will not be seen if

they are all encoded equally. In that case, removing any of the landmarks (highly fixated or least fixated) should have the same impact on performance.

Accuracy was calculated for each subject and then averaged. An accurate response consisted of marking the right position and orientation on the map for the tested view. Figure 4 shows the accuracy results for participants in the High-LM-Removed and Low-LM-Removed groups.

Figure 4 shows that there is a statistically significant difference in performance for the High-LM-Removed group, where only the low-gaze-time landmarks were present (gray marker in Figure 4) in comparison to when all of the landmarks were present, indicated by the black marker ($t(6) = 2.534, p = 0.04$). Mean accuracy differences between the *Full* and *Half* conditions are plotted in the white columns in Figure 4. We do not find a significant reduction in accuracy for the *Half* condition in comparison to the *Full* condition when the least fixated landmarks were removed ($t(6) = 0.229, p = 0.82$). Note

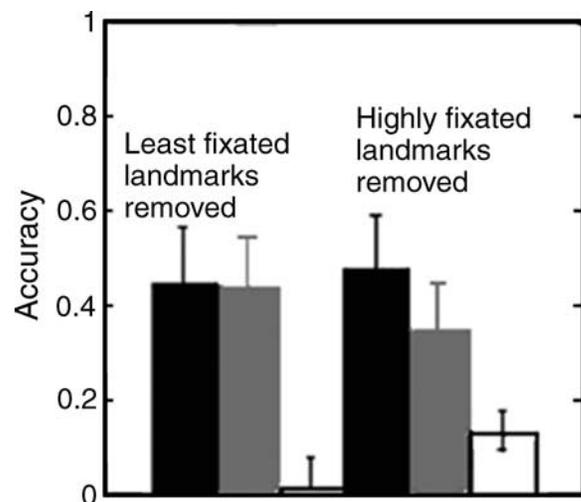


Figure 4. Plots show the mean accuracy of participants for the 10-corridor environments. The black marker shows the data for the participants in the “Full” condition and the gray marker shows the accuracy results from the “Half” condition. The white marker shows the “Difference” in performance between the “Full” and “Half” conditions.

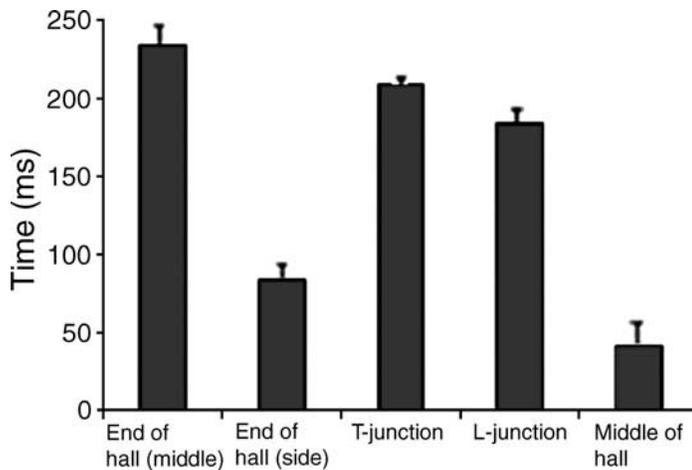


Figure 5. Plots show the averaged fixation duration on a landmark for participants. The landmarks are grouped together according to their location in the environment. “End of Hall (middle)” are the landmarks in the center, at the end of a corridor. “End of Hall (sides)” are the landmarks located on the sides at the end of a corridor. “T-junction” and “L-junction” are the landmarks placed at the T-junctions and L-junctions in the environment. “Middle” are the landmarks placed on either side in the middle of the corridor.

that when the most fixated landmarks were removed, performance was not reduced to zero. Thus, some of the least fixated landmarks were encoded to some extent, and we cannot claim from these data that fixation is a necessary condition for encoding.

Figure 5 shows the average amount of time all participants spent looking at landmarks. Gaze time was calculated by adding up the time participants spent looking at each landmark when it was visible to them in the environment. The landmarks are divided according to their position in the environment. The gray bars show the average amount of time (total time spent looking at the landmark divided by the number of instances of that particular position within the environment, e.g., T-junction) that the participants spent looking at the center of the landmark according to its position.

Figure 5 shows that gaze is not distributed equally among all object landmarks available in the environment. In views when several landmarks are available, participants tend to spend more time looking at the landmarks at the end of corridors and T- and L-junctions rather than landmarks located on the sides of corridors. Participants are likely to select some landmarks over others presumably because they are trying to preferentially encode into memory information that is most useful in navigation. Landmarks that are visible from different viewpoints can help an individual localize himself in a number of views. Landmarks at the end of corridors are visible from the greatest number of views. On the other hand, learning a landmark that is visible from several views may also be confusing as an individual can make distance or orientation errors (left or right).

Figure 6a shows the landmarks (highlighted in gray) that were removed for the participants in the *Half* experimental condition of the Least-LM-Removed Group. The landmarks in gray are the ones that the majority⁴ of the participants spent the least amount of time looking at during the *testing* and *training* phases. Figure 6b shows the location of the landmarks that had the longest gaze time for the participants during the *testing* and *training* phases and were hence removed for the majority of the participants during the *Half* experimental phase of the High-LM-Removed group. The landmarks that were removed were decided on a participant-to-participant basis, looking at their individual fixation times; however, we see a consistent pattern with the position of landmarks with the most or least gaze time being similar across participants as seen in Figures 6a and 6b. Figures 6a and 6b show the most typical layout of landmarks that were removed in the two conditions across all participants.

A very similar set of landmarks was removed for each of the subjects, indicating that for the most part subjects allocated gaze in a similar fashion. Note that landmarks are identified here by their location, not by the image, as the location of the different images was randomized between subjects and it could not be purely image characteristics that could be making some landmarks more preferred than others. There were variations in one or two landmarks that were removed for some individuals. However, there was overall consistency in the location of the object landmark used while learning the environment. The location of the landmarks appears to be an important factor in what information is encoded. Some landmarks get the majority of gaze time whereas others are barely looked at depending on their position in the environment. Figure 7 shows gaze time on landmarks for an individual participant from the experiment as an example of individual gaze distribution. This pattern of gaze was seen across all subjects: some landmarks are looked at a lot whereas others have very little to no gaze time. This adds support to the hypothesis that not all landmarks are encoded while a participant is exploring and learning a new large-scale environment. The selectivity in encoding

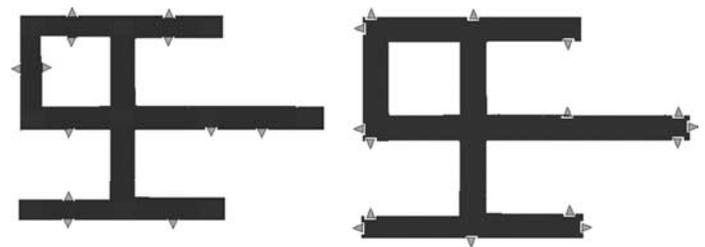


Figure 6. (a) Highlighted in gray are the landmarks removed for the participants in the Least-Removed Group “Half” Experimental Phase. (b) Highlighted in gray are the positions of the landmarks removed for the participants in the High-LM-Removed Group “Half” Experimental Phase.

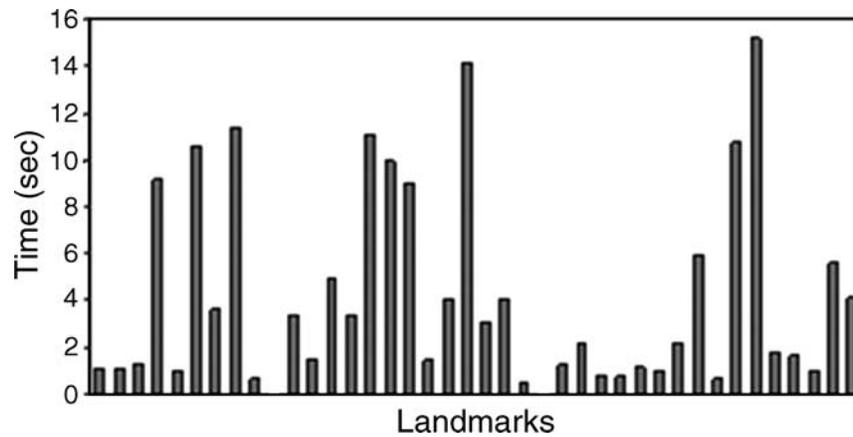


Figure 7. The amount of time (in seconds) a subject from the study spent looking at each one of the 40 landmarks in the environment. The x-axis shows each one of the object landmarks (images) and the y-axis plots time in seconds.

appears to be based on the placement of the landmarks in the environment.

The landmarks with the longest gaze times were located at the end of corridors and at T-junctions or L-junctions. Removing these landmarks impaired performance. Landmarks in the middle of the hall were visible in a scene for comparable periods as the landmarks at the end of a hall, L-junction, or T-junction, but the participants often ignored the middle of the hall completely and directed gaze to the end of the hall, despite the larger visual angle and generally small eccentricity of the landmarks in the middle.⁵ This suggests that it is the location of a landmark that is the critical factor in the allocation of attention. This is presumably an efficient use of memory resources.

Discussion

The results show that gaze patterns reveal the selective encoding of landmarks when subjects learn to navigate in a large-scale environment. The amount of time spent looking at a particular landmark is seen to be dependent on the location of the landmark in the current environments used. The participants spend more time looking at a subset of landmarks at decision points in the environment, in particular, L- and T-junctions, and at the end of halls. Landmarks placed in the middle of corridors have shorter gaze times. Subjects base their performance on the most fixated landmarks, indicating that the most fixated landmarks were encoded best. Half the landmarks could be removed without deleterious effect.

There was good agreement between subjects on which landmarks were viewed most. Those landmarks were ones that helped subjects localize from multiple views. For example, landmarks at the end of the hallways had the highest gaze times, presumably because that landmark provides information for the entire hallway, and they may

not even have to travel all the way down the hallway for successful localization (as in *experimental phase*) or wayfinding (as in *testing phase*). The object landmark at the end of the hall can tell a participant which specific hall they are in, which direction they are facing (orientation), and the specific position in the hall by calculating the distance from the landmark. This may allow efficient use of memory. For example, encoding the landmark placed at the end of the hall makes it unnecessary to also encode the landmarks placed in the middle of the hall. The middle of hall landmarks do not add new information about the participant's location within the environment; the landmark at the end of the hall is enough to identify which corridor the subject is in and their orientation. Redundant information is typically ignored by the subjects, for example, if a lot of gaze time is given to a landmark placed on the right wall, then the landmark on the left wall was barely fixated. Presumably, participants conserve memory resources by storing the minimum amount of critical information needed to complete a navigation task. Since performance was not impaired when half of the landmarks were removed (in the Least-Removed case), it is possible that even more of the landmarks could have been removed without cost. Other evidence also suggests that humans depend on the most reliable and simple navigation strategy available to them making the least demands on memory, storage, and cognitive processing (Foo et al., 2005; Lynch, 1960; Trullier, Wiener, Berthoz, & Meyer, 1997).

It is also possible that subjects choose the landmark at the end of the hall because it is at the focus of expansion. However, as can be seen in [Figure 1](#), when the subject is located at the beginning of a hallway, landmarks in the middle of the hall have an eccentricity of only a few degrees and have a relatively large visual angle (depending on the subject's simulated position) as the object landmarks in the environment are large and protrude from the wall. Nevertheless, the participants are seen to prefer to look at the landmark at the end of the hall even though

the visual angle of that landmark is much smaller. In real and virtual environments in simple hall-like configurations (where wayfinding was not required), subjects distribute gaze over a variety of locations in a scene (Jovancevic, Sullivan, & Hayhoe, 2006; Jovancevic-Misic & Hayhoe, 2009; Turano et al., 2003). In all these studies, fixations were specific to the task demands of the particular situation. For example, in Turano et al.'s (2003) study, fixations were primarily restricted to doors on the left, since subjects were required to find the 5th door on that side. In Jovancevic-Misic and Hayhoe's (2009) studies, participants walked on an oval path around a room and allocated gaze primarily to pedestrians who were likely to veer toward the walker, or on a leader pedestrian. In both cases, relatively few fixations were made straight ahead toward the focus of expansion, and fixations were primarily dedicated to the specific task-relevant locations. Other regions in the scene were largely ignored, including landmark information such as wall posters. Thus, in the current results while some fixations at the end of the hall may be for purposes of monitoring flow, it seems unlikely to be the predominant factor, particularly given the strategic value of these landmarks, as described above. Consistent with Turano et al.'s (2003) and Jovancevic et al.'s (2006; Jovancevic-Misic & Hayhoe, 2009) studies, subjects appear to actively select the specific information that is most relevant to the current behavioral needs.

Landmarks initially help participants localize, allowing an association to be formed between a specific landmark and location. This can lead to the gradual development of route knowledge, which helps in navigating from one point to the next (Ruddle et al., 1997; Vinson, 1999). Landmarks can act as navigational beacons that allow for accurate bearing estimates and allow participants to home in on the target position (Aginsky, 2001; Hamilton, Driscoll, & Sutherland, 2002; Waller, Loomis, Golledge, & Beall, 2000). Distant landmarks on known routes can enable novel shortcuts and detours, a view that has been supported by behavioral data gathered from highly skilled taxi drivers (Chase, 1983; Giraud & Peruch, 1988; Maguire, Frackowiak, & Firth, 1997; Peruch, Giraud, & Garling, 1989) and humans traveling in virtual environments (Gillner & Mallot, 1998). The current study suggests that landmarks located on major pathways or at important decision-making areas are more likely to be fixated and remembered. This finding is consistent with that of Aginsky et al. (1997) who showed that subjects are more likely to notice changes in landmarks that are located at decision points. Henderson and Hollingworth (1999) have shown that changes to objects that have been fixated are most likely to be noticed, so the landmarks in Aginsky et al.'s (1997) study were probably fixated most, although eye movements were not recorded in their study, and they did not directly relate change detection to navigation performance. The present study links the fixations to memory and to navigation performance.

One issue of concern in the current study is to what extent the results might generalize to walking in a real environment. Previous research has shown that navigational knowledge learned in virtual environments does translate well to real-world navigation (Koh, Wiegand, Garnett, Durlach, & Shinn-Cunningham, 1999; Stankiewicz & Eastman, 2006; Witmer, Bailey, & Kerr, 1996). In the current study, subjects navigated by key press and the normal proprioceptive and vestibular information from walking is not available. A similar issue arises with driving simulators. This might lead subjects to place greater reliance on visual cues. In addition, the environment was not immersive and was compressed in scale relative to a normal environment. Stankiewicz and Eastman (2006) addressed these questions and compared navigation performance in three different conditions: keyboard-press condition (like the current study), joystick condition allowing continuous movement rather than the quantized movements, and immersive condition where the participant's movements in a virtual reality arena were recorded. The environments stimulated for Stankiewicz and Eastman's (2006) study were similar to the maze-like environments used in the current studies. In each condition, the participant's efficiency in reaching the goal state was measured. The results did not reveal any significant difference in the performance across all three conditions. Thus, there is some evidence for the generality of the present findings, but it seems likely that the availability of different cues will lead to differences in strategy in different contexts. A similar concern about generality arises when considering the particular paradigm used. It is well known that gaze patterns are very sensitive to task (Hayhoe & Ballard, 2005), so a different training context could well evoke a different gaze pattern. However, subjects did not know in advance what kind of navigation task they would have to complete at each stage so it seems likely that some kind of default strategy was used. The selection of particular visual landmarks located at decision points seems like a robust strategy, and the agreement between different subjects suggests that the basic finding might be quite general.

Another issue is how much the fixations reveal about what is encoded. Although fixations are a clear indication that subjects were attending to a particular landmark, landmarks falling on the peripheral retina might also have been attended and remembered by the participants. Moreover, the encoding of global information about structure will not be indicated by a particular fixation. Since that half of the landmarks that were viewed least could be removed with little effect suggests that the fixations do in fact capture most of the landmarks that are encoded. However, removal of the most viewed landmarks resulted in a fairly modest impairment in performance, so subjects clearly have encoded the least viewed landmarks to some extent.

In summary, we investigated whether gaze patterns might reveal how subjects encode landmarks while learning to navigate in a new large-scale space. Previous

evidence suggests that landmark encoding is selective, but this has not been explicitly demonstrated. We found extensive selectivity in the gaze patterns. The landmarks subjects fixated most were consistent across subjects and appeared to be those that were most important for subsequent navigation in the environment. We also found that subjects preferentially fixated landmarks at choice points in the current maze-like environment used in the study. Thus, gaze allocation reveals the selective encoding of object landmarks in large-scale navigation tasks. Landmarks placed at specific locations within the environment, which presumably help in localizing the subject over a larger number of locations, appear to be encoded better. This may allow most effective use of memory resources in navigation. As suggested in previous research, the data support the view that gaze patterns are informative about which information is encoded. We show, in addition, that gaze is preferentially distributed in navigational tasks, that participants show a consistent pattern in how they select the data to be encoded, and that preference is given to object landmarks in specific locations that can act as decision points.

Acknowledgments

This work was supported by AFOSR FA09550-04-1-0236 AFOSR MURI FA09550-05-1-0321 and by NIH EY016089 NIH R01 EY05729.

Commercial relationships: none.

Corresponding author: Sahar N. Hamid.

Email: nadeem.sahar@gmail.com.

Address: Mold Road, Wrexham LL112AW, UK.

Footnotes

¹The figures and statistics used in the rest of the paper use the gaze data from the *training* and *testing* phases only as the *experimental* phase consisted of brief presentation of static views. The majority of the gaze data comes from the *training* phase (approximately 80%) as it took participants longer to complete the *training* phase in comparison to the *testing* phase.

²We calculated the least number of moves required traveling from the specified hotspot to the next and then we compared it to the number of moves that it took the participant to travel from the specified hotspot to the next. We compared the two values, and if the participant had less than 80% efficiency, then we counted it as a failed attempt.

³This means that minor errors are introduced since the data when the eye is in transit between fixations are not explicitly excluded, but in our experience such errors are comparable to those introduced by fixation finders.

⁴At least 5 out of the 7 participants had to have looked at the landmark for less or more than the median time for it to be plotted in [Figures 6a](#) and [6b](#). Landmarks that were least viewed or highly viewed only by 1 or 2 participants are not plotted.

⁵Observational data from when the participants were learning the environment showed that the participants tended to direct their gaze toward the end of the hall upon entering a new hallway. For middle of the hall landmarks, participants would pick one side; that is, if they looked to the right wall, then they ignored the object landmark placed on the left or vice versa.

References

- Abu-Ghazze, T. M. (1996). Movement and way finding in the King Saud University built environment: A look at freshman orientation and environmental information. *Journal of Environmental Psychology*, *16*, 303–318.
- Aginsky, V. (2001). *How visual landmarks are selected during small-scale navigation*. Providence, RI: Brown University.
- Aginsky, V., Harris, C., Resnick, R., & Beusmans, J. (1997). Two strategies for learning a route in a driving simulator. *Journal of Environmental Psychology*, *17*, 317–331.
- Chase, W. G. (1983). Spatial representation of taxi drivers. In D. Rogers & J. A. Sloboda (Eds.), *The acquisition of symbolic skill* (pp. 391–405). New York: Plenum Press.
- Collett, T. S., Collett, M., & Wehner, R. (2001). The guidance of desert ants by extended landmarks. *Journal of Experimental Biology*, *204*, 1635–1639.
- Collett, T. S., & Graham, P. (2004). Animal navigation: Path integration, visual landmarks and cognitive maps. *Current Biology*, *14*, R475–R477.
- Darken, R. P., & Sibert, J. L. (1996). Navigating in large virtual worlds. *International Journal of Human-Computer Interaction*, *8*, 49–72.
- Dyer, F. C. (1991). Bees acquire route-based memories but not cognitive maps in a familiar landscape. *Animal Behavior*, *41*, 239–246.
- Dyer, F. C., Berry, N. A., & Richard, A. S. (1993). Honey bee spatial memory: Use of route-based memories after displacement. *Animal Behavior*, *45*, 1028–1030.
- Etienne, A. S., Boulens, V., Maurer, R., Rowe, T., & Siegrist, C. (2000). A brief view of known landmarks reorients path integration in hamsters. *Naturwissenschaften*, *87*, 494–498.
- Etienne, A. S., Maurer, R., Boulens, V., Levy, A., & Rowe, T. (2004). Resetting the path integrator: A

- basic condition for route-based navigation. *Journal of Experimental Biology*, *207*, 1491–1508.
- Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 195–215.
- Gillner, S., & Mallot, H. A. (1998). Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of Cognitive Neuroscience*, *10*, 445–463.
- Giraud, M. D., & Peruch, P. (1988). Spatio-temporal aspects of the mental representation of urban space. *Journal of Environmental Psychology*, *8*, 9–17.
- Gould, J. L., & Gould, C. G. (1982). The insect mind: Physics or metaphysics? In D. R. Griffin (Ed.), *Animal mind–human mind* (pp. 269–298). Berlin, Germany: Springer-Verlag.
- Hamilton, D. A., Driscoll, I., & Sutherland, R. J. (2002). Human place learning in a virtual Morris water task: Some important constraints on the flexibility of place navigation. *Behavioral Brain Research*, *1*, 159–170.
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*, 188–194.
- Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, *10*, 438–443.
- Hollingworth, & 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.
- Jovancevic, J., Sullivan, B., & Hayhoe, M. (2006). Control of attention and gaze in complex environments. *Journal of Vision*, *6*(12):9, 1431–1450, <http://www.journalofvision.org/content/6/12/9>, doi:10.1167/6.12.9. [PubMed] [Article]
- Jovancevic-Misic, J., & Hayhoe, M. (2009). Adaptive gaze control in natural environments. *Journal of Neuroscience*, *29*, 6234–6238.
- Kearns, M. J., Warren, W. H., Duchon, A. P., & Tarr, M. J. (2002). Path integration from optic flow and body senses in a homing task. *Perception*, *31*, 349–374.
- Koh, G., von Wiegand, T., Garnett, R., Durlach, N., & Shinn-Cunningham, B. (1999). Use of virtual environments for acquiring configurational knowledge about specific real-world spaces: I. Preliminary experiment. *Presence: Teleoperators and Virtual Environments*, *8*, 632–656.
- Kowler, E., Anderson, E., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*, 1897–1916.
- Land, M. F. (2004). Eye movements in daily life. In I. Chalupa & J. Werner (Eds.), *The visual neurosciences* (vol. 2, pp. 1357–1368). Cambridge, MA: MIT.
- Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*, 742–744.
- Lynch (1960). *The image of the city*. Cambridge, MA: MIT Press.
- Maguire, E. A., Frackowiak, R. S., & Frith, C. D. (1997). Recalling routes around London: Activation of the right hippocampus in taxi drivers. *Journal of Neuroscience*, *17*, 7103–7110.
- Menzel, R., Chittka, L., Eichmüller, S., Peitsch, D., & Knoll, P. (1990). Dominance of celestial cues over landmarks disproves map-like orientation in honey bees. *Zeitschrift für Naturforschung C*, *45*, 723–726.
- Nadeem, S. (2008). *Information acquisition in navigation*. Ph.D. thesis, Psychology Department, University of Texas at Austin.
- Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, *41*, 3587–3596.
- Peruch, P., Giraud, M. D., & Garling, T. (1989). Distance cognition by taxi drivers and the general public. *Journal of Environmental Psychology*, *9*, 233–239.
- Riecke, B. E., van Veen, H., & Bulthoff, H. H. (2002). Visual homing is possible without landmarks: A path integration study in virtual reality. *Presence: Teleoperators and Virtual Environments*, *11*, 443–473.
- Rodrigo, T., Chamizo, V. D., McLaren, I. P. L., & Mackintosh, N. J. (1997). Blocking in the spatial domain. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 110–118.
- Ruddle, R. A., Payne, S. A., & Jones, D. M. (1997). Navigating buildings in “desk-top” virtual environments: Experimental investigations using extended navigational experience. *Journal of Experimental Psychology: Applied*, *3*, 143–159.
- Sanchez-Moreno, J., Rodrigo, T., Chamizo, V. D., & Mackintosh, N. J. (1999). Overshadowing in the spatial domain. *Animal Learning and Behavior*, *27*, 391–398.
- Schneider, W. X. (1995). VAM: A neuro-cognitive model for visual attention control of segmentation, object recognition, and space-based motor action. *Visual Cognition*, *2*, 331–375.
- Shannon, C. E. (1993). The best detection of pulses. In N. J. A. Sloane & A. D. Wyner (Eds.), *Collected papers* (pp. 148–150). New York: IEEE Press.
- Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. In

- H. W. Reese (Ed.), *Advances in child development and behavior* (vol. 10, pp. 9–55). New York: Academic Press.
- Spetch, M. L., Cheng, K., MacDonald, S. E., Linkenhoker, B. A., Kelly, D. M., & Doerkson, S. R. (1997). Use of landmark configuration in pigeons and humans. Generality across search tasks. *Journal of Comparative Psychology*, *111*, 14–24.
- Stankiewicz, B. J., & Eastman, K. M. (2006). Lost in virtual space II: The role of proprioception and discrete actions when navigating with uncertainty. *Environment and Behavior*.
- Stankiewicz, B. J., & Kalia, A. (2007). Acquisition of structural versus object landmark knowledge. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 378–390.
- Tom, A., & Denis, M. (2003). Referring to landmark or street information in route directions: What difference does it make? In W. Kuhn, M. F. Worboys, & S. Timpf (Eds.), *Lecture notes in computer science* (vol. 2825, pp. 362–374). Heidelberg/Berlin, Germany: Springer-Verlag.
- Trullier, O., Wiener, S. I., Berthoz, A., & Meyer, J. A. (1997). Biologically based artificial navigation systems: Review and prospects. *Progress in Neurobiology*, *51*, 483–544.
- Turano, K. A., Geruschat, D. R., & Baker, F. M. (2003). Oculomotor strategies for the direction of gaze tested with a real world activity. *Vision Research*, *43*, 333–346.
- Vinson, G. N. (1999). Design guidelines for landmarks to support navigation in virtual environments. In *Proceedings of the Annual ACM SIGCHI 1999 Conference* (pp. 278–285).
- Waller, D., Loomis, J., Golledge, R., & Beall, A. C. (2000). Place learning in humans: The role of distance and direction information. *Spatial Cognition and Computation*, *2*, 333–354.
- Wehner, R., Bleuler, S., Nievergelt, C., & Shah, D. (1990). Bees navigate by using vectors and routes rather than maps. *Naturwissenschaften*, *77*, 479–482.
- Witmer, B. G., Bailey, J. H., & Kerr, B. W. (1996). Virtual spaces and real world places: Transfer of route knowledge. *International Journal of Human-Computer Studies*, *45*, 413–428.