Eye gaze during route learning in a virtual task

Elizabeth H. Hall[1,2], Martha R. Forloines[3], John M. Henderson[1,2], Joy J. Geng[1,2]

[1]Center for Mind and Brain, University of California, Davis

267 Cousteau Place 267, Davis, CA 95618 USA

[2]Department of Psychology, University of California, Davis

1 Shields Ave, Davis, CA 95616 USA

[3]Department of Neurology, University of California, Davis

4800 Y Street, Sacramento, CA 95817 USA

*Address correspondence to:

Elizabeth H. Hall, ehhall@ucdavis.edu

Joy J. Geng, jgeng@ucdavis.edu

Abstract

Looking behavior during route learning is uniquely attuned to information relevant for navigation, which unfolds continuously over time. To better understand how visual information is used to learn routes, we tracked eye-position in participants on a guided walk through a simulated urban environment. The route was presented twice to participants in either a natural sequential order or a randomly scrambled order. After viewing the route, participants were presented with images of the intersections and asked to report the correct direction to continue travel along the route. Predictably, the sequential group performed much better on memory for the route than the scrambled group. Analyses of the eye-data revealed reliable differences between groups during route learning but no differences during scene viewing in the memory test trials. Specifically, while both groups looked at objects such as people and cars during route learning, looking at the path ahead was highly predictive of being in the sequential group, whereas looking at buildings ahead was predictive of being in the scrambled group. The results suggest that route learning involves encoding, and anticipating, information about the path to-be-traveled, perhaps reflecting formation of a temporally continuous representation of the route. In contrast, those in the scrambled group appeared to rely on building landmarks to attempt to create a map of the route, albeit unsuccessfully.

Keywords: route learning, navigation, predictive looking, eye movements, eye tracking

Introduction

Route learning is a ubiquitous skill among animals (Brown & Cook, 2006; Kelly & Gibson, 2007), but the information used to learn and remember routes varies across species. For example, birds can use an internal magnetic compass (Able & Able, 1990), ants have been shown to count steps (Wittlinger, Wehner, & Wolf, 2006), and honeybees use landmarks to guide navigation (Cartwright & Collett, 1982). In humans, route learning is based on spatial representations in the hippocampus that are built using a network of brain regions receiving multisensory inputs while they traverse the route (Maguire, et al., 2003; 2006; Moser, et al., 2008; Schedlbauer, et al., 2014; Stachenfeld, et al., 2017; van Asselen, et al., 2006). These studies have found that the spatial maps are built from a combination of allocentric and egocentric information (Newcombe, 2018; Zhang, et al., 2012). Despite a rich literature on the representation of spatial maps learned from navigation, relatively little is known about what information is sampled through attention and eye-movements during route learning. Eye gaze during route learning is important because fixations indicate what information is selected to build mental representations of a route.

Detailed models of how eye gaze is controlled during free viewing and during task have been developed within the scene viewing literature (Torralba, Olivia, Castlehano, & Henderson, 2006). Gaze allocation during free viewing of natural scenes is known to be biased towards objects that inform the viewer about the context and gist of the scene (Peacock, Hayes, & Henderson, 2019; Neider & Zelinsky, 2006; Resnick, O'Regan, & Clark, 1997), are visually salient (Itti, Koch & Nieber, 1998), or have meaning (Henderson & Hayes, 2017; Hwang, Wang, & Pomplun, 2011). However, when a task is introduced, looking behaviors tend to be driven by specific information related to the goal. For example, when searching for a trashcan, participants look at the bottom third of an office scene containing floorspace rather than the middle or top portions (Castelhano & Krzys, 2020; Pereria and Castelhano, 2019; Torralba, Oliva, Castelhano, & Henderson, 2006; see also Malcolm & Henderson, 2010). Likewise, search for a teddy bear in a bedroom begins with eye-movements towards the bed whereas search for a computer in the same room begins with looks to a desk (Võ, Boettcher, and Draschkow, 2019). These studies demonstrate that learned expectations guide observer looking behaviors.

Similarly, task requirements drive gaze behavior in real world tasks such as object sorting (Triesch, et al., 2003), sandwich making (Ballard, Hayhoe, & Pelz, 1995), driving (Shinoda, Hayhoe, & Shrivastava, 2001), or walking (Turano, Geruschat, & Baker, 2003; for a review see Henderson, 2017). In these tasks, the participant is an active agent within a continuous environment and results show that eye-movements not only seek out task-relevant information but do so in an anticipatory manner. For example, when making a sandwich, individuals look at the peanut butter before actually reaching for the jar (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003). When driving a car, drivers look toward the tangent of the curve to predict and prepare for the upcoming turn (Land & Lee, 1994). Cricket batsmen look to where the ball will bounce just prior to its arrival rather than tracking the ball from the bowler's release (Land & McLeod, 2000). Finally, walkers told to turn left at a specific room will focus gaze largely on the left side of the hallway, especially at doors, until the correct turn is detected. Together, these results indicate that in real-world tasks, eye gaze reflects anticipatory information sampling: observers selectively look at information that is necessary for active behavior in the next moment of time.

The pattern of looking during natural walking also depends on physical and cognitive demands. For example, if the terrain is rocky, walkers look at the next immediate foothold, indicating a need to extract more precise information to modify the current step; alternatively, if the terrain is smooth, participants look farther ahead on the path or at environmental surroundings (Hayhoe & Matthis, 2018). When learning a new route through a cityscape, participants look selectively at landmarks located at intersections where turns are made. For example, on a righthand turn, walkers trying to learn the route focus on the landmark on their right side leading up to, and throughout, the turn. They appear to do this to encode the landmark and its navigationally relevant position for future recall of the route. Participants that walked the route without the intention of future recall, did not look at landmarks at intersections differently based on whether a turn was made there or not (Wenczel, Hepperle & von Stülpnagel, 2017).

Landmarks are known to be particularly important for route learning; humans have been shown to use landmarks in Morris Water Mazes (Bullens, et al., 2010), can triangulate precise goal locations using only landmarks (Forloines, Bodily, & Sturz, 2015), and at their simplest can aid in learning and recalling new environments (Spiers & Maguire, 2004; for a review of the non-human animal literature on landmark learning see Brown & Cook, 2006; Kelly & Gibson, 2007) but there is still much unknown about how gaze can drive attention to gather information that contributes to building memories for new routes. In this study, we asked what features of the environment are looked at when learning a new route. To answer these questions, we asked participants to view a route with the intention of encoding it for later recall by following a "guide" through a virtual city via a series of screenshot images taken from Grand Theft Auto V (Rockstar Games), an immersive 3D third-person perspective video game. We split participants into two groups that differed in how the images were presented; either in natural temporal order or in a scrambled order. Both groups viewed the route twice to provide two chances for learning. We queried their knowledge of the direction to proceed along the route and their confidence in that response by presenting a single image taken from each intersection.

We hypothesized that when provided with the temporal context of a continuous walk, participants would encode more information about the spatial structure of the route, over and above just the landmarks in a static scene, in order to create a spatial map. Consistent with the findings from Hayhoe et al., (2003) and Land et al., (2000; 2004), we expected the ordered group to look more at the path ahead, particularly on the second walk, in anticipation of where to go next. Further, when provided with temporal context, we expected there to be more looks to irrelevant, yet salient, features of the environment such as people and cars along the route. We predicted that if participants are not provided with temporal context (i.e., view the route in a scrambled order) they will employ an alternative learning strategy, relying on landmarks such as the buildings directly ahead of the route to serve as the main source of information to build a representation of the route (Wenczel, et al., 2017). Finally, we predicted that there would be differences in the participants provided with temporal context based on the memory requirements and location along the route.

Method

Participants. Sixty undergraduate participants (mean age = 20.25, SD = 1.74, 16 males) from UC Davis volunteered for this study. Participants were compensated with course credit for their participation. All participants provided informed consent in accordance with UC Davis' Institutional Review Board.

Stimuli and Apparatus. Participants were seated in a sound attenuated and dimly lit room in front of a 27-in Asus LCD monitor (2560 x 1440 pixels) with a refresh rate of 60 Hz. Participants' eyes were tracked throughout the entire experiment using an EyeLink 1000 Desktop mount eye-tracker sampling at 500 Hz (SR Research, Mississuagua, Ontario, Canada). All participants completed 13-point eye-tracking calibration, and participants' eyes were recalibrated between repetitions if needed. Recalibration was performed for all participants prior to the final test phase. The experiment was constructed using PsychoPy (https://www.psychopy.org/) and consisted of showing participants two repetitions of the route either in natural temporal order or in a scrambled order followed by a memory test of the route.

Three identical routes were recorded from a path through the virtual city center from Grand Theft Auto V (Rockstar Games). The path followed a guide (the player in the center portion of the screen) through the city and crossed 10 intersections consisting of 3 right hand turns, 3 left hand turns, and 4 straightaways. For 7 of these intersections, there were 3 potential directions to travel (straight, left, and right) and for the remaining 3 intersections there were only 2 potential directions to travel (2 with only right and left, and 1 with only straight and right). The only difference between the 3 routes was the time of day of the recordings. This ensured participants would experience the route similarly to how it would be experienced in real life with varying unstable features such as cars, bystanders, shadows, or sun-glares during each walk. As such, participants could not rely on these inconsistent features to recall the route. Once the routes were recorded, screenshots were taken at 1s intervals for use in the task from the first (569 images) and second (557 images) repetitions, while the third route screenshots were used for the test images (10 images). The slight difference in the number of images in each repetition was due to slightly different paths taken along the route (e.g., avoiding a bystander along the route), but the overall route walked was identical. Each image was presented to participants for 1s with no interstimulus interval for the most seamless image presentation possible. While viewing the two repetitions, participants were not required to respond. If participants' fixations moved off the screen (e.g., they closed their eyes, looked away from the computer, etc.) the images paused until a fixation was present for 1s to ensure participants did not miss any images (see Figure 1 top panel for the trial progression of each group).

Memory test images were taken from a third time of day in the game and were composed of 10 images that aligned most closely to the turning point of each intersection prior to any indication from the guide regarding the turn (e.g., a turn in body posture; Figure 1 bottom panel). Test images were presented to all participants in a random order and were shown for up to 2 minutes or until participants responded with which direction the route proceeded from that location (straightaway, turn left, or turn right). Following each image participants were asked to rate their confidence in the direction they chose. The confidence rating was on a sliding scale and ranged from “not at all” to “very”.

Procedure. Participants were randomly split into two groups of 30, an ordered group and a scrambled group. The ordered group viewed the images in a natural temporal order while the scrambled group viewed the images in a randomly presented order (Figure 1 top panel). All participants were shown the route twice and were told to learn the route as best they could. For each repetition, participants were shown the on-screen instructions presented below. The bolded portion was added for the scrambled group and was not included for the ordered group. “You will be shown images taken from a route through a virtual city, **but the images will be scrambled**. Please pay close attention to the images you are shown because you will be asked questions about the route at the conclusion of the experiment.”

Between repetitions, participants were told to close their eyes and rest for a minute. If needed, their eyes were recalibrated. The instructions were shown again, and participants viewed the second repetition. Following the second repetition, participants were again given a break and recalibration was completed. They were then shown the following instructions. “Now you will be shown a series of images. Each of these images is of an intersection from the route you just learned. You will have to report which direction you should travel at each intersection to continue along the route. **Even though the images you saw were out of order, please try your best to decide which direction you should go**. Please press the arrow keys to input which direction is the correct direction to continue along the route. Press LEFT if you should turn left. Press RIGHT if you should turn right. Press UP if you should go straight. Following this, you will be asked about your confidence in your answer. Use the ARROW KEYS to move on the scale and ENTER to input your rating. You will only have 2 minutes to answer each question, so please answer quickly, but try to be as accurate as possible." The bolded portion was added for the scrambled group and was not included for the ordered group. Participants made route decisions for each test image and rated their confidence in their answer on a sliding scale rating from 1 (not confident) – 10 (confident). Once they completed these test questions the experiment was finished.

[Insert Figure 1 here]

Data Analysis. The behavioral measure of recall via accuracy and confidence rating was collected during the test trials. As the two repetitions involved only passive viewing, there was no behavioral measure collected during the route learning portions of the experiment.

The eye data was the measure of interest during route learning in both groups and was collected continuously throughout the experiment. The primary analyses of interest were based on the proportion of fixation duration to each object of interest as a function of group. Fixations shorter than 100ms were removed from the analyses as noise. Each image during route learning was only presented for 1s with an average of 2.32 fixations per participant per image (*SD* = 1.18, range = 1 to 10, median = 2).

Areas of interest (AOIs) were drawn for the 5 images prior to and the 5 images following the intersection and for each test image, resulting in 11 images per intersection (110 intersection images total; Figure 2 bottom panel). In addition to the images pre- and post-intersection, 3 sets of 10 sequential non-intersection images (30 images total) were taken from portions along the route that contained no visible intersections. The non-intersection data was used as a comparison to determine how gaze differs when participants are in vs. out of an intersection. AOIs were determined a priori to be potentially relevant and were bounded by rectangles fitted to the height and width of each object. The AOIs were created around the following in all images: the guide, the building ahead, building signs, the path ahead, the path behind, the incorrect path, people, and vehicles (Figure 2 top panel for an example). The AOIs encompassing the people, vehicles, and building signs were drawn over each object as wholly as possible, with a new AOI for each object. The "path ahead" was defined as the area encompassing the to-be-walked path by the guide. The "incorrect path" was defined similarly to the path ahead but covered the area the guide would travel if they were to go along any of the alternate paths. The "buildings ahead" were defined as the building directly in front of the guide. Fixations to areas not included in the AOIs consisted of trees, the sky, distant buildings, etc. and were classified as "other".

[Insert Figure 2 here]

Each of the AOIs contained varying levels of information for the participants. For example, the buildings ahead could help to unite the spatial representation and act as visual landmarks during the recall test. Alternately, the vehicles and people varied across repetition and as such carried little information value. For the majority of the intersection and non-intersection images, all AOIs were included, however two intersections had no building signs, one intersection had no incorrect path or building ahead. Of note, the non-intersection images did not contain an incorrect path as there was only one direction the guide could potentially travel. For the intersection memory test images, there were two intersections without a building ahead, two intersections without other people, one without a building ahead or vehicles, and one without building signs or a path behind.

Results

Behavioral Results. First, in order to determine if the ordered group indeed learned the route better than the scrambled group, we compared performance on the memory test between groups. We expected that performance would be much higher for the ordered group compared to the scrambled group. A one-way ANOVA was conducted on the average accuracy during test trials with group (ordered, scrambled) as a factor. Results revealed a main effect of group F (1, 58) = 88.845, $p < .001$, $\eta^2 = .605$. Post hoc comparisons of group showed that the ordered group ($M = .84$, $SD = .125$) chose the correct direction to go more often than the scrambled group ($M = .5$, $SD = .153$), $p < .001$, Cohen's $d = 2.434$ (Figure 3, left panel). To test if participants in either group were performing to a level greater than chance, separate one-sample t-tests were conducted. Of note, chance was based on the total number possible directions at that location. For 7 of the 10 intersections, chance was set at .33, however, for three intersections chance was .5 as there was no third path to choose. For the ordered group, all turns ($ps < .05$) except intersection 9 had greater than chance performance ($p > .05$). The scrambled group only performed above chance on 4 intersections (intersection, 4, 6, 7, 10). Of note, these intersections on which the scrambled group performed above chance were all straightaways. To test if participants in the scrambled group had a straightaway response bias (i.e., guessing "straightaway" when they did not know), we completed a one-sample t-test comparing the proportion of responses to each direction against chance (.33). There were more responses to the straightaway direction than would be expected by chance alone t (9) = 3.264, $p < .01$. The responses to the left and right direction were not greater than chance (*ps* > .05). Taken together, the behavioral results from test trials confirm that participants in the ordered group were able to learn the route well with only two walk repetitions, and the scrambled group was not.

A one-way ANOVA was conducted on the average confidence rating during test trials with group (ordered, scrambled) as factors. Results revealed a marginal but significant main effect of group F (1, 58) = 4.028, $p < .05$, $\eta^2 = .065$. The average rating of the ordered group (M = 7.767, SD = 1.412) was greater than that of the scrambled group (M = 7.05, SD = 1.354), $p <$ .05, Cohen's d = .518 (Figure 3, right panel). These confidence rating results show a marginal difference in confidence across groups, but as can be seen in Figure 3, this is due to confidence ratings being high regardless of performance accuracy. This is not unexpected given participants' tendency to overestimate their performance in testing situations (Foster, Was, Dunlosky, & Isaacson, 2017; Händel & Dresel, 2018; Rahnev, et al., 2020).

[Insert Figure 3 here]

AOI Fixation Results. Fixations for each trial were transformed into the proportion of total time spent fixating on each AOI, and then averaged across the 11 images surrounding the intersection to create the average proportion of time spent fixating on each AOI. This duration metric was used for all of the following fixation analyses. Our primary goal was to test our hypothesis that the ordered group would show anticipatory gaze toward the path ahead (similar to and Hayhoe et al., 2003; Land et al., 2000; 2004), particularly more so on the second repetition when participants might have already learned parts of the route, and that the *scrambled group* would attend more to landmarks (e.g., the building ahead, and building signs). In addition, we tested for group differences in looking time to people and vehicles, which are known to draw attention during free-viewing of static images. To do this, separate linear mixed effects ANOVAs with subjects as a random effect were conducted on the proportion of fixation duration on each of these select AOIs (building ahead, building signs, path ahead, people, vehicles), using the factors of group (ordered, scrambled), and repetition (1, 2; Figure 4).

[Insert Figure 4 here]

First, to test our main hypothesis that the ordered group would spend a greater proportion of fixation time on the path ahead, we conducted a linear mixed effects ANOVA which revealed a significant main effect of group F (1, 57.45) = 25.713, p < .001, but no effect of repetition F (1, 55.75) = 1.938, p > .05. There was however, an interaction between group and repetition F (1, 55.75) = 11.303, p < .01. As predicted, the ordered group fixated on the path ahead to a greater extent than the scrambled group, *t* (57.454) = 5.071, *p* < .001. The interaction was driven by the ordered group fixating on the path ahead more than the scrambled group in the second repetition *t* (55.751) = -3.362, p < .01. This difference in each group's fixations on the path ahead can be seen in sample intersection heatmaps shown in Figure 5. To further examine if the ordered group was fixating the path ahead as a function of having *learned* the route (e.g., more fixations in the second repetition than the first) or if they were fixating on the route as a function of *learning* the route (e.g., no difference in fixations across repetitions) we separated the ordered group and conducted a paired samples *t*-test on fixations by repetition in the ordered group only which showed a difference between the first and second repetition *t* (13383) = -3.909, *p* < .001. In order to gauge the strength of this difference, we conducted a Bayesian paired samples *t*-test. The analysis revealed a strong difference between the repetitions with a BF_{10} of 20.185. The proportion of fixation duration was shorter during the first repetition (*M* = .011, *SD* = .061, CI [.010 - .012]) than in the second repetition (*M* = .014, *SD* = .069, CI [.013 - .015]). In order to determine if these differences were simply due to the participant attending to the guide's body motion cues (i.e., physically turning) during the second half of the turn following pivot point, we conducted a similar set of *t*-tests on the fixations during the first five images leading up to the tangent and the five images following the tangent. The paired samples *t*-test revealed a significant difference in these two image sets, t (9197) = -2.580, $p < .05$, however the Bayesian paired samples *t*-test revealed this difference was not reliable with a BF_{10} of .327, suggesting this result should be interpreted with caution. Fixations during the first five images (M = .009, SD = .056, CI [.008 - .010]) were not reliably different from the second five images (M = .011, SD = .063, CI [.010 - .013]). Combined these results suggests that the anticipatory gaze in the ordered group is a function of *learning* the route as much as it is a function of having *learned* the route. We found differences in fixations once the participants knew which direction the path would continue (*learned*: the second repetition), but showed no reliable differences based on where in the intersection they were positioned (*learning*: first half vs second half).

[Insert Figure 5 here]

To test our hypothesis that the scrambled group would show greater fixation time on the landmarks (i.e., buildings and their features), we conducted a linear mixed ANOVA on fixations to the building ahead which revealed a significant main effect of group F (1, 57.91) = 18.61, $p <$ .01 and of repetition F (1, 58.09) = 5.753, $p < .05$. There was an interaction between group and repetition F (1, 58.09) = 4.72, $p < .05$. As predicted, the scrambled group fixated more on the building ahead than the ordered group, t (57.913) = -4.314, $p < .001$. There were more fixations during the second repetition compared to the first t (58.093) = -2.398, $p < .05$. The interaction was driven by an increase in the scrambled group's fixations on the building ahead in the second repetition compared with no change across repetitions in the ordered group. Both groups had similar fixation times on building signs $F(1, 58.09) = .079$, $p > .05$ and showed no differences in these fixation patterns across repetitions $F(1, 58.78) = .07$, $p > .05$. To determine if there were differences in how the groups fixated on the unstable and navigationally irrelevant environmental features (i.e., people and vehicles), we conducted a linear mixed ANOVA on the fixations to people which revealed a main effect of group $F(1, 57.91) = 12.815$, $p < .001$, and of repetition $F(1, 57.57) = 163.431$, $p < .001$. There was an interaction between group and repetition $F(1, 57.57) = 26.925$, $p < .001$. The ordered group fixated on people to a greater extent than the scrambled group, $t(57.906) = 3.58$, p < .001. There were more fixations to people in the second repetition than the first $t(57.575) = -12.784$, $p < .001$. The interaction was driven by a greater increase in fixations to people in the ordered group in the second repetition than the scrambled group, $t(57.575) = -5.189$, $p < .001$. When looking at fixations to vehicles, there was a main effect of group $F(1, 58.02) = 8.363$, $p < .01$, and of repetition $F(1, 56.07) = 88.318$, $p < .001$. There was no interaction $F(1, 56.07) = 3.881$, $p > .05$.

Classifier results. The previous results focused on AOIs for which we had *a priori* hypotheses for differences between groups. Next, we explored looking behavior to other AOIs in order to better evaluate the relative contribution of each type of information on route learning in the two groups. We trained a classifier to decode from the pattern of fixations whether a participant had studied the image in the scrambled or ordered condition. A linear support vector machine was trained on the total fixation time to each of the nine AOIs (Figure 2) in an image in a given trial. As participants had only 1s to study an image, they often fixated on only a few of the nine AOIs. This led to a large, but sparse dataset, with only 110,098 non-zero feature values out of 551,120 feature values, or 19.98% (*SD* = .48) of all feature values. To address this we used techniques standard to text classification which also deals with datasets with many zero feature values; namely we normalized each trial independently of other trials so that its norm equaled one, and we employed a liblinear solver in our support vector machine as it is optimized for the classification of large, sparse datasets (Fan et al. 2008). The classifier was iteratively trained and tested on all trials using leave-one-out cross validation and accuracy was estimated as the proportion of correctly classified trials.

[Insert Figure 6 here]

The classifier was able to determine whether the trial was viewed in the ordered or scrambled condition at an accuracy level significantly above-chance (.56 correct; binomial test, $p < .001$). A linear classifier creates a hyperplane that uses support vectors to maximize the distance between the two conditions and the coefficients obtained from the model represent the vector coordinates orthogonal to the hyperplane. The direction of the coefficients obtained from the model indicates the predicted group, and the magnitude of the coefficients in relation to each other can be used to determine the importance of fixation duration to each AOI when classifying viewing group. As the classifier was iteratively trained and tested on all trials using leave-one-out cross validation, coefficient magnitudes are averaged across all trials. Higher fixation time on path ahead ($M = 1.352$, $SD = .0004$), path incorrect ($M = 1.0$, $SD = .0007$), people ($M = 0.705$, $SD = .0003$), vehicle ($M = .618$, $SD = .0003$), guide ($M = .293$, $SD = .0001$) and path behind ($M = .222$, $SD = .0029$) led the model to classify the trial as one viewed in the ordered condition, while higher fixation time on building sign ($M = -.004$, $SD = .0003$), other ($M = -.182$, $SD = .0003$), and building ahead ($M = -.998$, $SD = .0003$) led the model to classify the trial as viewed in the scrambled condition.

We also trained and tested the model separately on trials by repetition. Trials in which the images were studied in the first repetition (.56 correct; binomial test, $p < .001$), and second repetition (.58 correct; binomial test, $p < .001$) were both classified significantly above-chance. However, there was a difference in classification accuracy by repetition, with trials from the second repetition classified at a significantly higher rate (Welch's t-test: t (55105.33) = 8.734, $p < .001$). There was also a range of classification accuracy across images. The image with the highest average classification accuracy was classified at an average .72 accuracy (binomial test, $p < .001$), significantly above chance. The image with the lowest average classification accuracy across trials was classified significantly below chance (.41 correct, binomial test, $p < .01$), suggesting that the pattern of fixation time to AOIs in that image for the ordered group was more similar to that of the scrambled group and therefore more likely to be misclassified.

Discussion

We sought to determine how gaze varies as a function of route learning in the presence or absence of temporal context. The behavioral results indicated that the route was well learned in only two repetitions with temporal context but was challenging without it. However, as both groups were actively trying to learn the route, albeit with one group at a disadvantage, we analyzed gaze patterns to determine how looking was different across groups to get an idea of what people determined as navigationally important throughout various sections of the route.

The primary prediction we tested was that the ordered group would use temporal context to anticipate the path ahead. Consistent with this prediction, the ordered group showed anticipatory gaze toward the path and the scrambled group focused more on the most reliable landmarks (i.e., the buildings ahead of them). We hypothesize that these looking patterns reflect different strategies used by each group to stitch together a spatial representation of the route. Additionally, we also found that the ordered group looked more at navigationally irrelevant objects (e.g., people and vehicles), perhaps because tracking the route was less demanding, allowing participants to explore other objects of general interest. That is, because the ordered group could use temporal context to anticipate the route, they may have freed cognitive resources to explore interesting, but unstable objects regardless of their usefulness. These results suggest that individuals use temporal order in route learning to create and test predictions about where to go next along the route. Presumably, when the prediction is not confirmed, an error signal is generated that updates their route representation (Bestmann et al., 2008; Blakemore et al., 1998; Franklin & Wolpert, 2011; den Ouden, Kok, & de Lange, 2012). The use of error signals to update representations is ubiquitous in human learning (Bar, 2009; Friston, 2005; Hawkins, 2004) and suggests route learning employs similar mechanisms.

Next, to explore the relative contribution of information from the *a priori* AOIs compared to other possible sources of information, we entered all of the object AOIs into an SVM classifier to determine what information was most important for each group in building a spatial representation of the route[1]. The AOIs that were most heavily weighted in the classification were the path ahead for the ordered group and the building ahead for the scrambled group. Of note, the path ahead and building ahead, were also statistically significant in the group comparison, but they were not actually fixated for very long (approximately 5% of the total time). This finding indicates that although these were important sources of information in differentiating between route learning in the two groups, the absolute amount of time spent sampling this information was rather small. This suggests that a lot of time was spent looking at objects that did not specifically contribute to route learning in the two groups. In this study, one of those objects was

[1] We also considered fixations to the regions outside of the preassigned AOIs (termed "other" in the classifier analyses). These regions contained trees, the sky, distant buildings, and the areas between the preassigned AOIs. A one-way ANOVA revealed a main effect of group F (1, 556) = 88.854, $p < .001$, $\eta^2 = .132$, and repetition F (1, 556) = 24.090, $p < .001$, $\eta^2 = .036$. The scrambled group spent a greater proportion of time fixating on other areas compared to the ordered group ($p_{\text{tukey}} < .001$, Cohen's $d = -.779$). There was a greater proportion of fixations to other areas in the second repetition than the first ($p_{\text{tukey}} < .001$, Cohen's $d = -.385$). There were no interactions.

the guide. Although a large proportion of time was spent looking at the guide, the guide provided very little information about the path and did not predict learning in either group. Similarly, there was substantial looking at other objects such as passerbys, vehicles, other objects (deHaas et al., 2019), as might be expected from static scene viewing, but these objects were not as predictive of route learning.

Overall, the results suggest that scene viewing during a route learning is influenced by many factors, but only a small subset of information is relevant to route learning, *per se*. Specifically, we found key differences in looking behavior during route learning when temporal order was present versus absent such that the path ahead was looked at more frequently by the ordered group and differentiated the ordered from the scrambled group. In contrast, the scrambled group relied more on landmarks such as buildings ahead, but they nevertheless performed poorly on the memory task. This suggested that route learning is indicated by anticipatory looking towards the expected path, even when this anticipatory looking involves relatively little absolute time.

Acknowledgements: We would like to thank Michelle Occhipinti for completing the data collection and Bo-Yeong Won, Philip Witkowski, Xinger Yu, and Zhiheng Zhou for their thoughtful comments and guidance. This research was supported by NIH T32AG050061 to MRF and NIH MH113855 to JJG.

References

Able, K. P., & Able, M. A. (1990). Calibration of the magnetic compass of a migratory bird by celestial rotation. *Nature, 347,* 378-380.

Aginsky, V., Harris, C., Rensink, R., & Beusmans, J. (1997). Two strategies for learning a route in a driving simulator. *Journal of Experimental Psychology, 17,* 317-331.

Agisnky, V., & Tarr, M. J. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition, 7,* 147-162.

Aivar, M. P., Hayhoe, M. M., Chizk, C. L., & Mruczek, R. E. B. (2005). Spatial memory and saccadic targeting in a natural task. *Journal of Vision, 5,* 177-193. doi:10.1167/5.3.3

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7,* 66-80.

Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. *Introduction. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364*, 1181–1182. 10.1098/rstb.2008.0321

Bestmann, S., Harrison, L. M., Blankenburg, F., Mars, R. B., Haggard, P., Friston, K. J., et al. (2008). Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. *Curr. Biol. 18,* 775–780. doi: 10.1016/j.cub.2008.04.051

Blakemore, S. J., Goodbody, S. J., and Wolpert, D. M. (1998). Predicting the consequences of our own actions: the role of sensorimotor context estimation. J. Neurosci. 18, 7511–7518. doi: 10.1523/JNEUROSCI.18-18-07511.1998

Brown, M.F., & Cook, R.G. (Eds.). (2006). *Animal Spatial Cognition: Comparative, Neural, and Computational Approaches*. [On-line]. Available: pigeon.psy.tufts.edu/asc/

Bullens, J., Nardini, M., Doeller, C. F., Braddick, O., …, Burgess, N. (2010). The role of landmarks and boundaries in the development of spatial memory. *Developmental Science, 13,* 170-180. doi: 10.1111/j.1467-7687.2009.00870.x

Cartwright, B.A., & Collett, T.S. (1982). How honeybees use landmarks to guide their return to food. *Nature, 295,* 560-564.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9,* 1-15. doi:10.1167/9.3.6

Cristino, F., & Baddeley, R. (2009). The nature of the visual representations involved in eye movements when walking down the street. *Visual Cognition, 17,* 880-903. doi:10.1080/13506280902834696

de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *PNAS, 116,* 11687-11692. doi: 10.1073/pnas.1820553116

den Ouden, H. E. M., Kok. P., & de Lange, F. (2012). How prediction errors shape prediction, attention, and motivation. *Frontiers in Psychology, 3,* Article 548. doi: 10.3389/fpsyg.2012.00548

Ekstrom, AD, Huffman, DJ, & Starrett, M. (2017). Interacting networks of brain regions underlie human spatial navigation: A review and novel synthesis of the literature. *Journal of Neurophysiology, 188*(6), 3328-3344. doi:10.1152/jn.00531.2017

Fan, RE, Chang, KW, Hsieh, CJ, Wang, XR, & Lin, CJ. (2008). LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research*, *9*, 1871-1874.

Forloines, M. R., Bodily, K. D., & Sturz B. R. (2015). Evidence consistent with the multiple-bearings hypothesis from human virtual landmark-based navigation. *Frontiers in Psychology, 6:488.* doi: 10.3389/fpsyg.2015.00488

Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2017). Even after thirteen class exams, students are still overconfident: The role of memory for past exam performance in student predictions. *Metacognition Learning, 12*, 1-19. doi: 10.1007/s11409-0`6-9158-6

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research, 51,* 1920-1931. doi:10.1016/j.visres.2011.07.002

Franklin, D. W., and Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron, 72,* 425–442. doi: 10.1016/j.neuron.2011.10.006

Friston, K. J., Shiner, T., Fitzgerald, T., Galea, J. M., Adams, R., Brown, H., et al. (2012). Dopamine, affordance and active inference. *PLoS Comput. Biol. 8*:e1002327. doi: 10.1371/journal.pcbi.1002327

Händel, M., & Dresel, M. (2018). Confidence in performance judgement accuracy: The unskilled and unaware effect revisited. *Metacognition and Learning, 13,* 265-285. doi: 2-/2—7/s11409-018-9185-6

Hawkins, J. (2004). On Intelligence. New York: Times Books.

Hayhoe, M. M., & Ballard, D. (2005). Eye movements in natural behavior. *TRENDS in Cognitive Sciences, 9,* 188-194. doi:10.1016/j.tics.2005.02.009

Hayhoe, M. M., & Matthis, J. S. (2018). Control of gaze in natural environments: Effects of rewards and cost, uncertainty and memory in target selection. *Interface Focis, 8,* 2018009. doi:10.1098/rsfs.2018.0009

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3,* 49-63. doi: 10:1167/3.1.6

Heft, H. (1979). The role of environmental features in route-learning: Two exploratory studies of wayfinding. *Environmental Psychology and Nonverbal Behavior, 3,* 172-185.

Henderson, J. M. (2017). Gaze control as prediction. *TiCS, 21,* 15-23. doi: 10.1016/j.tics.2016.11.003

Henderson, J. M., & Haynes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behavior, 1,* 743-747. doi: 10.1038/s41562-017-0208-0

Hwang, A. D., Wang, H-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research, 51,* 1192-1205. doi:10.1016/j.visres.2011.03.010

Itti, L. Koch, C., & Nieber, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE, 20,* 1254-1259

Kelly, D. M., & Gibson, B. M. (2007). Spatial navigation: Spatial learning in real and virtual environments. *Comparative Cognition & Behavior Reviews, 2,* 111-124.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature, 369,* 742-744.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience, 3,* 1340-1345

Maguire, E. A., Nannery, R., & Spiers, H. J. (2006). Navigation around London by a taxi driver with bilateral hippocampal lesions. *Brain, 129,* 2894-2907. doi:10.1093/brain/awl286

Maguire. E. A., Spiers, H. J., Good, C. D., Hartley, T., Frackowiak, R. S. J., Burgess, N. (2003). Navigation expertise and the human hippocampus: A structural brain imaging analyses. *Hippocampus, 13,* 208-217. doi:10.1002/hipo.10087

Malcom, G. L., & Shomstein, S. (2015). Object-based attention in real-world scenes. *Journal of Experimental Psychology: General, 144,* 257-263. doi: 10.1037/xge0000060

Meister, M. L. R., & Buffalo, E. A. (2016). Getting directions from the hippocampus: The neural connection between looking and memory. *Neurobiology of Learning and Memory, 134,* 135-144. doi:10.1016/j.nlm.2015.12.004

Mennie, N., Hayhoe, M., & Sullican B. (2007). Look ahead fixations: Anticipatory eye movements in natural tasks. *Exp Brain Res, 179,* 427-442. doi: 10.1007/s00221-006-0804-0

Moser, E. I., Kropff, E., & Moser, M-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci. 31,* 69-89. doi: 10.1146/annrev.neuro.31.061307.090723.

Peacock, C., Haynes, T., R., & Henderson, J. M. (2019) Meaning guides attention during scene viewing, even when it is irrelevant. *Attention, Perception, & Psycholphysics, 81,* 20-34. doi: 10.3758/s13414-018-1607-7

Pereira, E. J., & Castelhano, M. S. (2019). Attentional capture is contingent on scene region: Using surface guidance framework to explore attentional mechanisms during search. *Psychonomic Bulletin & Review, 26,* 1273-1281. doi: 10.3758/s13423-019-01610-z

Péruch, P. & Wilson, P. N. (2004). Active versus passive learning and testing in a complex outside built environment. *Cognitive Processes, 5,* 218 – 227. doi: 10.1007/s10339-004-0027-x

Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., …, Zylberberg, A. (2020). The confidence database. *Nature Human Behavior, 4,* 317-325. doi: 10.1038/s41562-019-0813-1

Rensink, R. A, O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. Psychological Science, 8, 368-373.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7,* 1-20. doi:10.1167/7.14.16

Schedlbauer, A. M., Copara, M. S., Watrous, A. J., Ekstrom, A. D. (2014). Multiple interacting brain areas underlie successful spatiotemporal memory retrieval in humans. *Scientific Reports, 4,* 6431.

Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments? *Vision Research, 41,* 3535-3545.

Smith, C. N., Hopkins, R. O., & Squire, L. R. (2006). Experience-dependent eye movements, awareness, and hippocampus-dependent memory. *The Journal of Neuroscience, 26,* 11304-11312. doi:10.1523/jneurosci.3071-06.2006

Spiers, H. J., & Maguire, E. A. (2004). A 'landmark' study on the neural basis of navigation. *Nature Neuroscience, 7, 572-574.*

Stachenfeld, K. L., Botvikick, M. M., Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience, 20,* 1643-1653. doi:10.1038/nn.4650

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*, 1-23. doi:10.1167/11.5.5

Tom, A., & Denis, M. (2004). Language and spatial cognition: Comparing the roles of landmarks and street names in route instructions. *Applied Cognitive Psychology, 18,* 1213-1230. doi: 10.1002/acp.1045

Tom, A. C., & Tversky, B. (2012). Remembering route: Streets and landmarks. *Applied Cognitive Psychology, 26,* 182-193. doi: 10.1002/acp.1805

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113,* 766-786. doi: 10.1037/0033-295X.113.4.766

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision, 3,* 86-94. doi:10.1167/3.1.9

Turano, K., Geruschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research, 43,* 333-346.

van Asselen, M., Kessels, R. P. C., Neggers, S. F. W., Kappelle, L. J., Frijns, C. J. M., & Postma, A. (2006). Brain areas involved in spatial working memory. *Neuropsychologia, 44,* 1185-1194. doi: 10.1016/j.neuropsyhologia.2005.10.005

Viaene, P., Vansteenkiste, P., Lenior, M., De Wulf, A., & De Maeyer, P. (2016). Examining the validity of the total swell time of eye fixations to identify landmarks in a building. *Journal of Eye Movement Research, 9,* 1-11. doi:10.16910/jemr.9.3.4

Võ, M., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology, 29,* 205-210. doi: 10.1016/j.copsyc.2019.03.009

Wenczel, F., Hepperle, L., & von Stülpnagel, R. (2017). Gaze behavior during incidental and intentional navigation in an outdoor environment. *Spatial Cognition & Computation, 17,* 121-142. doi:10.1080/13875868.2016.1226838

Wetherell, A. (1979). Short-term memory for verbal and graphic route information. *Preceedings of the Human Factors Society Annual Meeting, 23,* 464-468. doi: 10.1177/1071181379023001117

Wiener, J. M., Hölscher, C., Büchner, S., & Konieczny, L. (2011). Gaze behavior during space perception and spatial decision making. *Psychological Research, 76,* 713-729. doi:10.1007/s00426-011-0397-5

Wittlinger, M., Wehner, R., & Wolf, H. (2006). The ant odometer: Stepping on stilts and stumps. *Science, 312,* 1965-1967. doi: 10.11126/science.1126912

Wu, C-C., Wang, H-C., & Pomplun, M. (2014). The roles of scene gist and dependency among objects in the semantic guidance of attention in real-world scenes. *Vision Research, 105,* 10-20. doi:10.1016/j.visres.2014.08.019

Yarbus, A. L. Eye Movements and Vision. New York: Plenumm Press; 1967.

Zhang, H., Copara, M., & Ekstrom, A. D. (2012). Differential recruitment of brain networks following route and cartographic map learning of spatial environments. *PloS ONE, 7,* e44886, doi:10.1371/journal.pone.0044886

Figure 1.

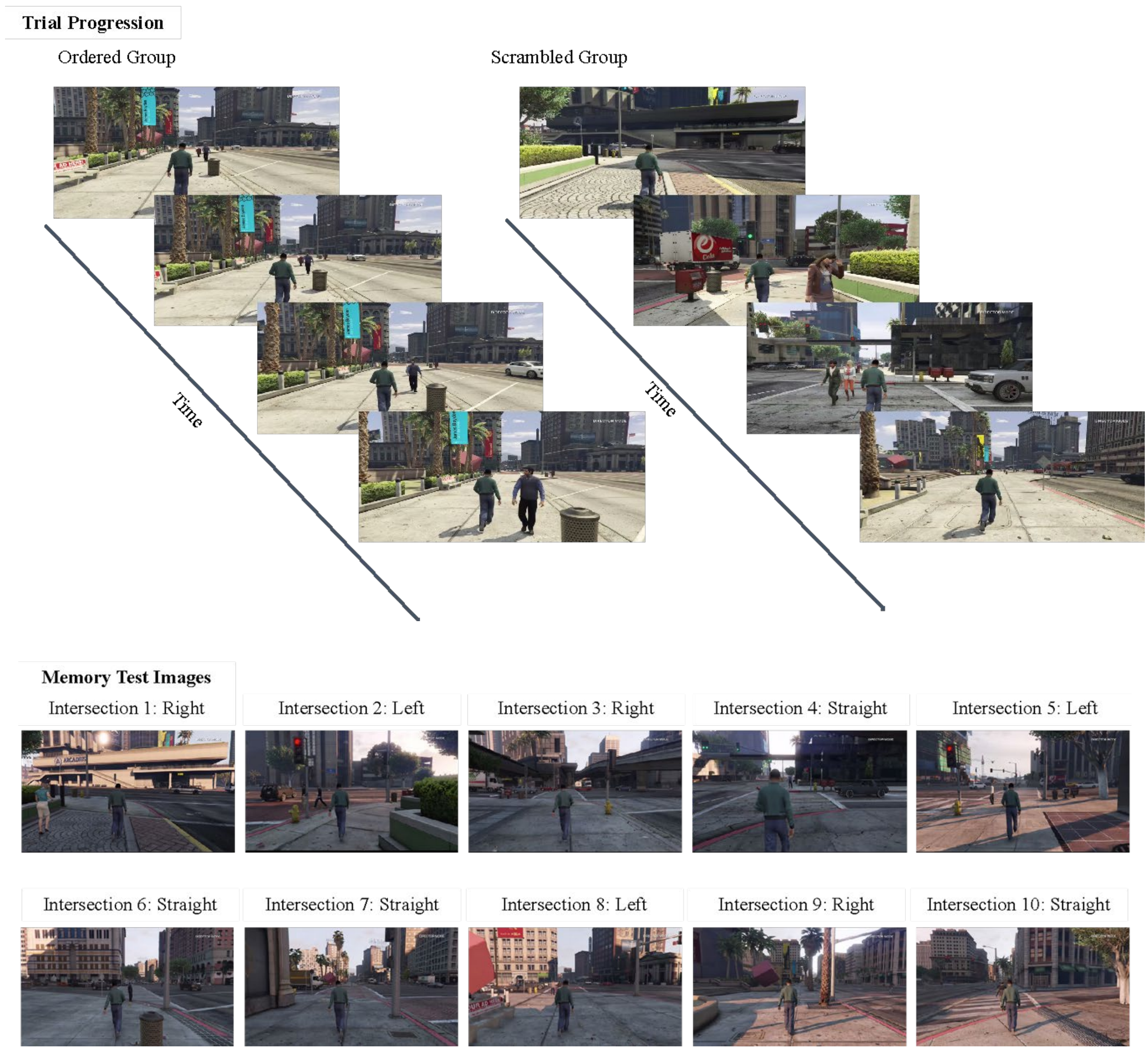

Trial Progression
Ordered Group
Scrambled Group
Time
Time
Memory Test Images
Intersection 1: Right
Intersection 2: Left
Intersection 3: Right
Intersection 4: Straight
Intersection 5: Left
Intersection 6: Straight
Intersection 7: Straight
Intersection 8: Left
Intersection 9: Right
Intersection 10: Straight

Figure 2.

AOI Example

Intersection Image Example

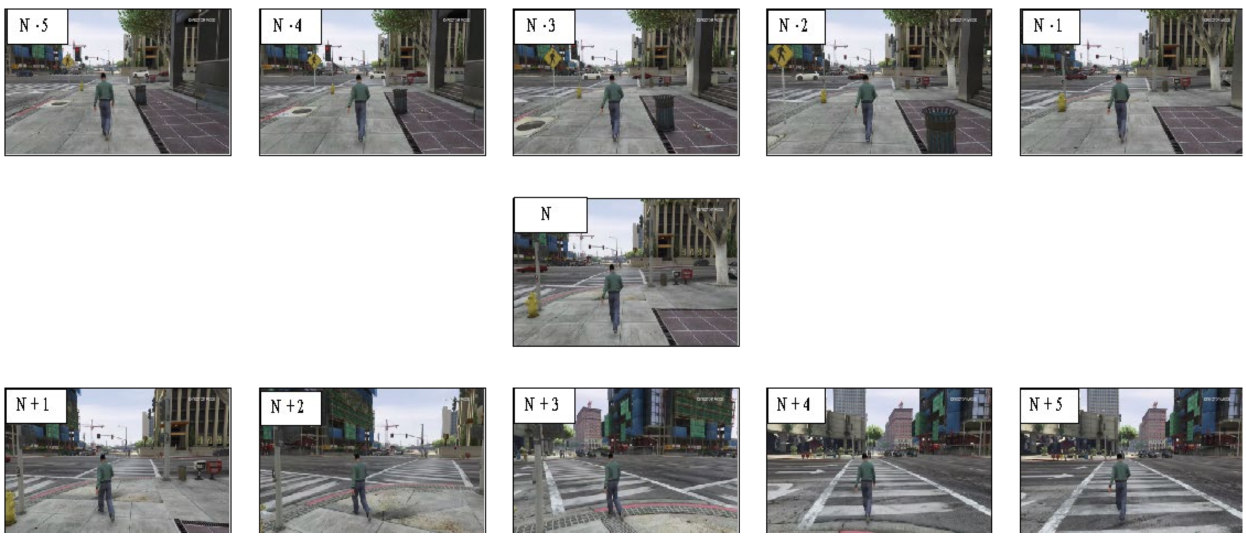
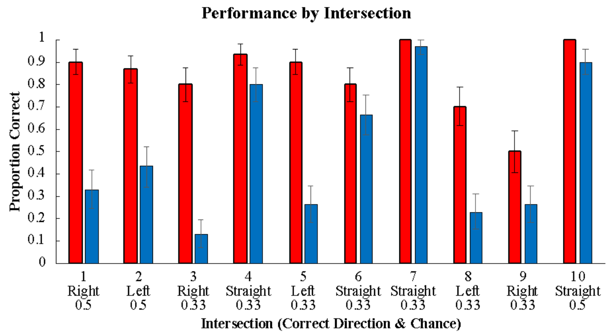


Figure 3.


Performance by Intersection
Proportion Correct
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 Right 0.5
2 Left 0.5
3 Right 0.33
4 Straight 0.33
5 Left 0.33
6 Straight 0.33
7 Straight 0.33
8 Left 0.33
9 Right 0.33
10 Straight 0.5
Intersection (Correct Direction & Chance)

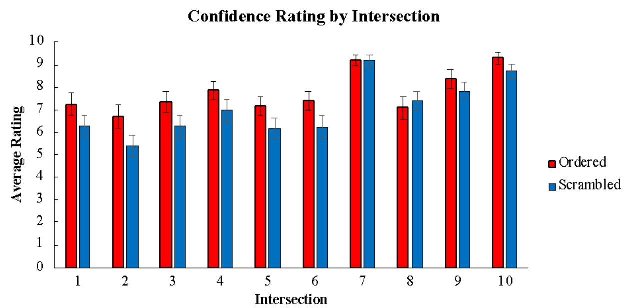

Confidence Rating by Intersection
Average Rating
10
9
8
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10
Intersection
Ordered
Scrambled

Figure 4.

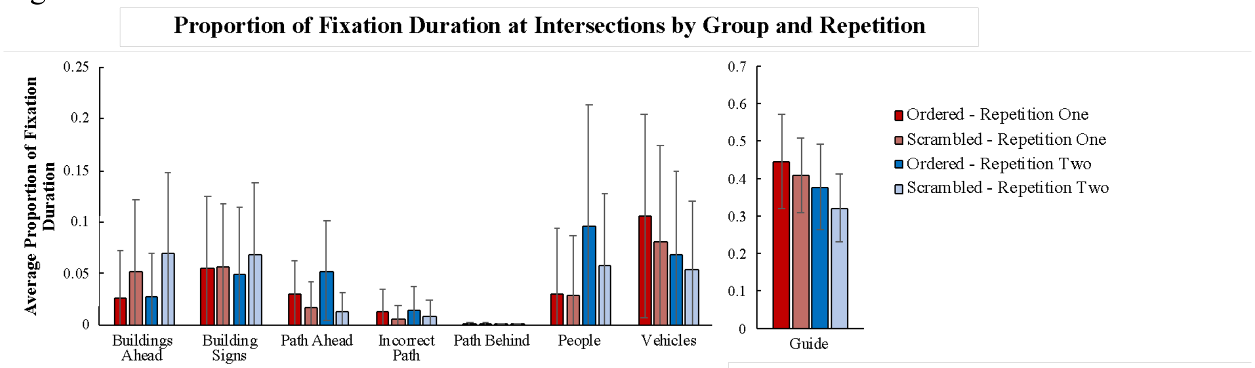
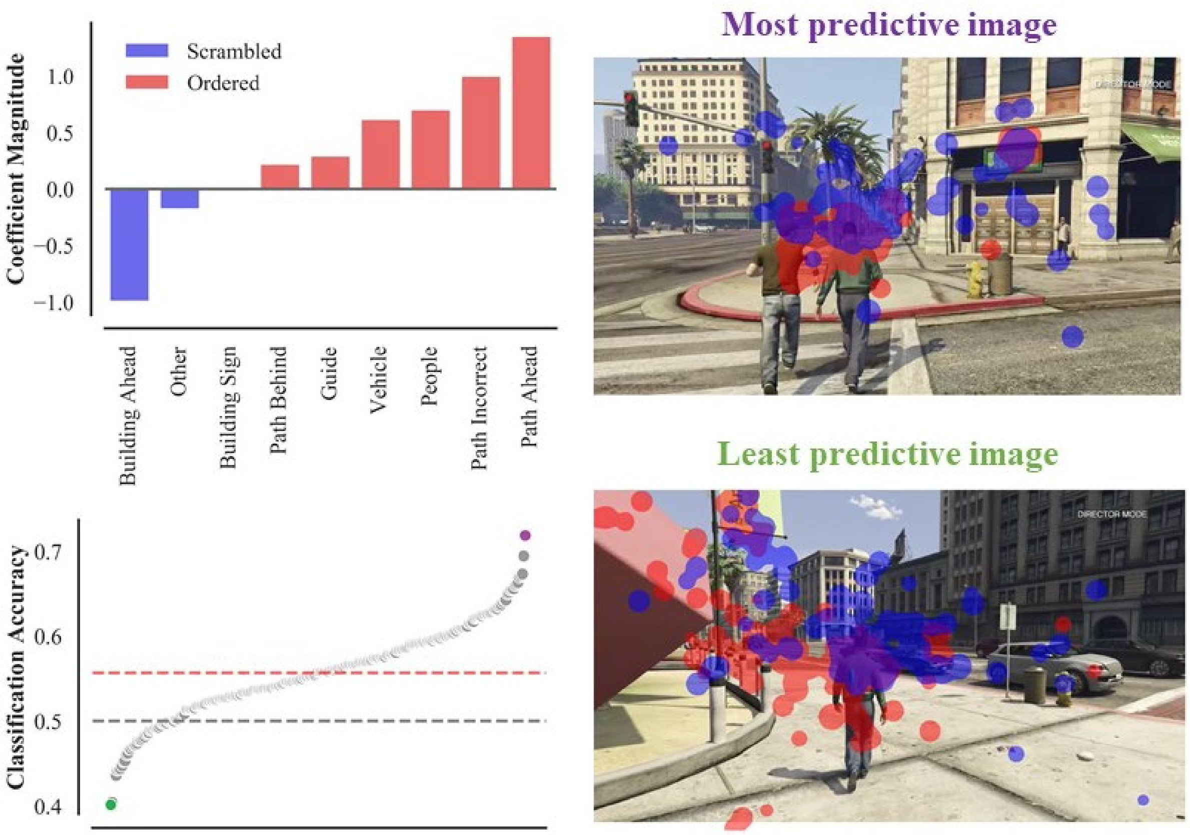

Proportion of Fixation Duration at Intersections by Group and Repetition
Average Proportion of Fixation Duration
0.25
0.2
0.15
0.1
0.05
0
Buildings Ahead
Building Signs
Path Ahead
Incorrect Path
Path Behind
People
Vehicles
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Guide
Ordered - Repetition One
Scrambled - Repetition One
Ordered - Repetition Two
Scrambled - Repetition Two

Figure 5.

Difference in Gaze Patterns Across Groups

Figure 6.



Most predictive image
Least predictive image
Scrambled
Ordered
Coefficient Magnitude
1.0
0.5
0.0
−0.5
−1.0
Building Ahead
Other
Building Sign
Path Behind
Guide
Vehicle
People
Path Incorrect
Path Ahead
Classification Accuracy
0.7
0.6
0.5
0.4
DIRECTOR MODE

Figure Captions

Figure 1. A representation of the experimental images. The top panel shows the trial progression across both groups as a function of time. The bottom panel shows the memory test images. The intersection is labeled above each image with the corresponding correct direction after the colon.

Figure 2. The top panel shows an example of the AOIs drawn on a sample image. The bottom panel shows an illustration of an image of an intersection (the N^{th} image), the 5 images prior to, and the 5 images following that were included within the analyses.

Figure 3. Bar graphs showing the behavioral results from the memory test. The red bars represent the ordered group and the blue bars represent the scrambled group. The graph on the left shows the proportion of correct responses by intersection. The x-axis labels contain the intersection number, the correct direction, and the chance level. For 3 of the intersections (1, 2, and 10) there were only 2 potential turning directions resulting in a .5 chance level. The remaining intersections had 3 potential turning directions resulting in a .33 chance level. The graph on the right shows the average rating on the confidence rating scale by intersection. Error bars represent the standard error of the mean.

Figure 4. A bar graph showing the average proportion of fixation duration to each of the AOIs. The dark red and light red bars represent the first repetition, and the dark blue and light blue bars represent the second repetition. The dark red and dark blue bars represent the ordered group, and the light red and light blue bars represent the scrambled group. The graph on the left and the graph on the right show the same information, but the axes are different to highlight the differences in fixations to each object. Error bars represent standard deviations.

Figure 5. Two heatmaps of gaze patterns across groups. The left two images are from the ordered group and the right two are from the scrambled group. The top two images are from a right-hand turn intersection, and the bottom two are from a left-hand turn intersection.

Figure 6. The bar graph shows the coefficient magnitudes for the linear support vector machine that was trained on the total fixation duration to each of the nine AOIs in an image. The blue bars represent higher fixation time on AOIs that led the model to classify the trial as one viewed in the scrambled condition and the red bars represent higher fixation duration on AOIs that led the model to classify the trial as one viewed in the ordered condition. In the plot below each point represents the average classification accuracy for a single image in a leave-one-out cross-validation. The purple point represents the image with the highest classification accuracy, and the green point represents the image with the lowest classification accuracy. The red line is plotted at the mean classification accuracy across images, while the gray line is plotted at chance classification accuracy. Both images are shown to the right with eye-movements from participants in the two conditions overlaid. In the most predictive image, the path ahead is heading straight and is one of the final images of the sequence. In the least predictive image, the guide has passed the tangent of the intersection and is walking along the path ahead.