Eye movements reveal spatiotemporal dynamics of visually-informed planning in navigation

Seren L. Zhu[1,5,*], Kaushik J. Lakshminarasimhan[2,5], Nastaran Arfaei[3], and Dora E. Angelaki[1,4]

[1]Center for Neural Science, New York University, New York
[2]Center for Theoretical Neuroscience, Columbia University, New York
[3]Department of Psychology, New York University, New York
[4]Department of Mechanical and Aerospace Engineering, New York University, New York
[5]These authors contributed equally
[*]Corresponding author

Contact:
lt1686@nyu.edu
jl5649@columbia.edu
na2481@nyu.edu
da93@nyu.edu

Abstract

Goal-oriented navigation is widely understood to depend upon internal maps. Although this may be the case in many settings, humans tend to rely on vision in complex, unfamiliar environments. To study the nature of gaze during visually-guided navigation, we tasked humans to navigate to transiently visible goals in virtual mazes of varying levels of difficulty, observing that they took near-optimal trajectories in all arenas. By analyzing participants' eye movements, we gained insights into how they performed visually-informed planning. The spatial distribution of gaze revealed that environmental complexity mediated a striking trade-off in the extent to which attention was directed towards two complimentary aspects of the world model: the reward location and task-relevant transitions. The temporal evolution of gaze revealed rapid, sequential prospection of the future path, evocative of neural replay. These findings suggest that the spatiotemporal characteristics of gaze during navigation are significantly shaped by the unique cognitive computations underlying real-world, sequential decision making.

Introduction

Planning, the evaluation of prospective future actions using a model of the environment, plays a critical role in sequential decision making [1, 2]. Two-step choice tasks have revealed quantitative evidence that humans are capable of flexible planning [3–5]. Under unfamiliar or uncertain task conditions, planning may depend upon and occur in conjunction with active sensing, the cognitively motivated process of gathering information from the environment [6]. After all, one cannot make decisions about the future without knowing what options are available. Humans and animals perform near-optimal active sensing via eye movements during binary decision making tasks [7, 8] and visual search tasks [9–11]. Such tasks are typically characterized by an observation model, a mapping between states and observations, and visual information serves to reduce uncertainty about the state. In contrast, sequential decision-making tasks require knowledge about the structure of the environment characterized by state transitions, and visual information can additionally contribute to reducing uncertainty about this structure. Therefore, the principles of visually-informed decision-making uncovered in simplified, discrete settings may not generalize to natural behaviors like real-world navigation, which entails planning a sequence of actions rather than a binary choice. How can we study visually-informed planning in structured, naturalistic sequential decision-making ventures such as navigation?

Theoretical work suggests that information acquisition and navigational planning can be simultaneously achieved through active inference – orienting the sensory apparati to reduce uncertainty about task variables in the service of decision making [6]. Humans are fortuitously equipped with a highly evolved visual system to perform goal-oriented inference. By swiftly parsing a large, complex scene on a millisecond timescale, the eyes actively interrogate and efficiently gather information from different regions of space to facilitate complex computations [12, 13]. At the same time, eye movements are influenced by the contents of internal deliberation and the prioritization of goals in real-time, providing a faithful readout of important cognitive variables [14–17]. Thus, eye tracking lends itself as a valuable tool for investigating how humans and animals gather information to plan action trajectories [14, 18–20].

Over the past few decades, research on eye movements has led to a growing consensus that the oculomotor system has evolved to prioritize top-down, cognitive guidance over image salience [21–23]. During routine activities such as making tea, we tend to foveate specifically upon objects relevant to the task being performed (e.g. boiling water) while ignoring salient distractors [22, 24]. A current consensus about active sensing is that gaze elucidates how humans mitigate uncertainty in a goal-oriented manner [14]. Therefore, we hypothesize that in the context of navigation, gaze will be directed towards the most informative regions of space, depending upon the specific relationship between the participant's position and their goal. A candidate framework to formalize this hypothesis is Reinforcement Learning (RL), whereby the goal of behavior is cast in terms of maximizing total long-term reward [25]. For example, this framework has been previously used to provide a principled account of why neuronal responses in the hippocampal formation depend upon behavioral policies and environmental geometries [26, 27], as well as a unifying account of how the hippocampus samples memories to replay [28]. Incidentally, RL provides a formal interpretation of active sensing, which can be understood as optimizing information sampling for the purpose of improving knowledge about the environment, allowing for better planning and ultimately greater long-term reward [14]. Here, we invoke the RL framework and hypothesize that eye movements should be directed towards spatial locations where small changes in the local structure of the environment can drastically alter the expected reward.

While active sensing manifests as overt behavior, the planning algorithms which underlie action selection are thought to be more covert [1]. Researchers have proposed that certain neural codes, such as hierarchical representations, would support efficient navigational planning by exploiting structural redundancies in the environment [29, 30]. There is also evidence for predictive sequential neural activations during sequence learning and visual motion viewing tasks [31–33]. This is reminiscent of replay, a well-documented phenomenon in rodents during navigation [34–37]. The mechanism by which humans perform visually-informed planning may similarly involve simulating sequences of actions that chart out potential trajectories,

and chunking a chosen trajectory into subgoals for efficient implementation. As recent evidence shows that eye movements reflect the dynamics of internal beliefs during sensorimotor tasks [38], we hypothesize that participants' gaze dynamics would also reveal sequential trajectory simulation, and thus reveal the strategies by which humans plan during navigation.

To test both hypotheses mentioned above, we designed a virtual reality navigation task where participants were asked to navigate to transiently visible targets using a joystick in unfamiliar arenas of varying degrees of complexity. We found that human participants balanced foveating the hidden reward location with viewing highly task-consequential regions of space both prior to and during active navigation, and that environmental complexity mediated a trade-off between the two modes of information sampling. The experiment also revealed that participants' eyes indeed rapidly traced the trajectories which they subsequently embarked upon, with such sweeps being more prevalent in complex environments. Furthermore, participants seemed to decompose convoluted trajectories by focusing their gaze on one turn at a time until they reached their goal. Taken together, these results suggest that the spatiotemporal dynamics of gaze are significantly shaped by cognitive computations underlying sequential decision making tasks like navigation.

Results

Humans use vision to efficiently navigate to hidden goals in virtual arenas

To study human eye movements during naturalistic navigation, we designed a virtual reality (VR) task in which participants navigated to hidden goals in hexagonal arenas. As we desired to elicit the most naturally occurring eye movements, we used a head-mounted VR system with a built-in eye tracker to provide a full immersion navigation experience with few artificial constraints. Participants freely rotated in a swivel chair and used an analog joystick to control their forward and backward motion along the direction in which they were facing (Figure 1a). The environment was viewed from a first-person perspective through an HTC Vive Pro headset with a wide field of view, and several eye movement parameters were recorded using built-in software.

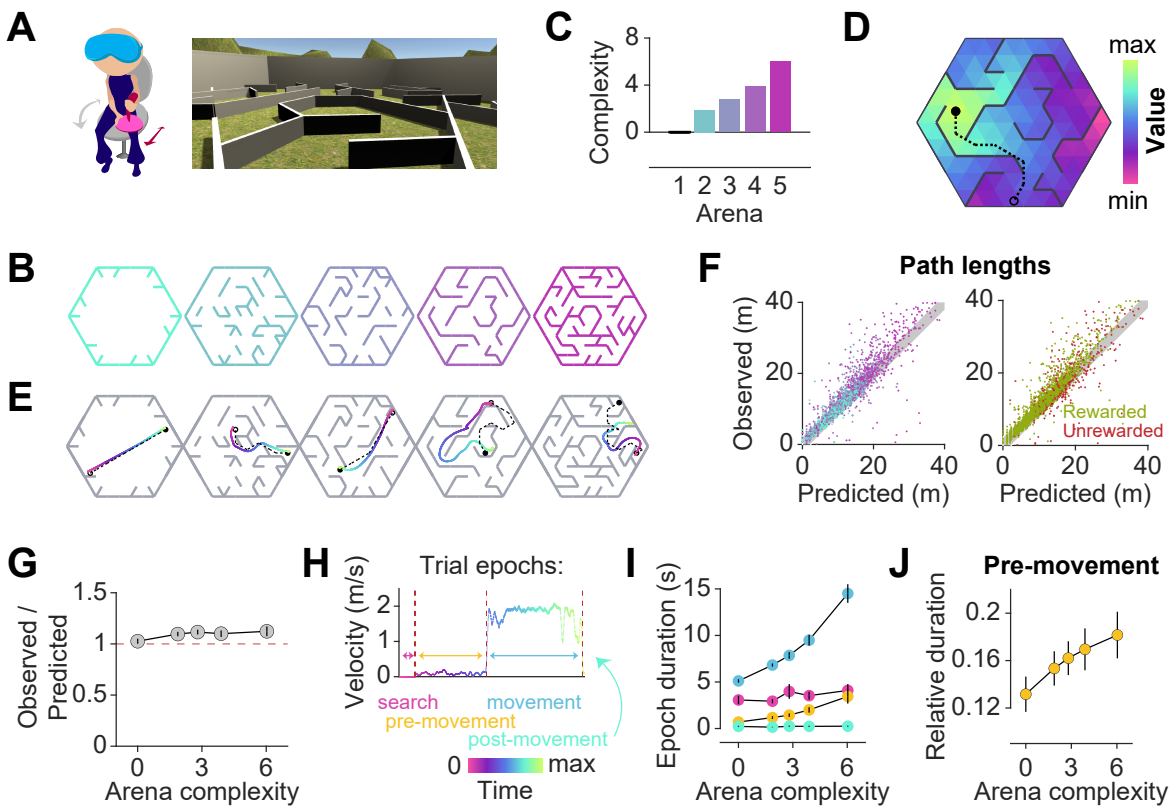

Figure 1: **Participants exhibit near-optimal navigation performance across multiple environments. A.** Left: Human participants wore a VR headset and executed turns by rotating in a swivel chair, while translating forwards or backwards using an analog joystick. Right: A screenshot of the first-person view of the display. The headset conferred an immersive field of view of 110°. **B.** Aerial view showing the layout of the arenas. **C.** Arenas ranged in complexity, which is related to negative mean state closeness centrality. **D.** Heatmap showing the value function corresponding to an arbitrary goal state (closed circle) in one of the arenas. The value of each state is related to the geodesic distance between that state and the goal. Dashed line denotes the optimal trajectory from an example starting state (open circle). **E.** Trajectories from an example trial in each arena, executed by one participant. The optimal trajectory is superimposed in black (dashed line). Time is color-coded. **F.** Comparison of the empirical path length against the path length predicted by the optimal policy. The gray shaded region denotes the width of the outer reward zone (see Figure S1a). Left: Data points are colored in accordance to the colors of each arena as depicted in **B**. Right: Unrewarded trials (red) vs. rewarded trials (green) had similar path lengths. For both plots, all trials for all participants and all arenas are superimposed. **G.** Across participants, the average ratio of observed vs. optimal (predicted) trajectory lengths is consistently around 1 in all arenas. **H.** The search epoch was defined as the period between goal stimulus appearance and goal stimulus foveation. A threshold applied on the filtered joystick input (movement velocity) was used to delineate the pre-movement and movement epochs. **I.** The average duration of the pre-movement (orange) and movement epochs (blue; colored according to the scheme in **H**) increased with arena complexity, in conjunction with the trial-level effects exerted by path lengths (Figure S11a). **J.** The relative planning time, calculated as the ratio of pre-movement to total trial time after goal foveation, was higher for more complex arenas. For **G**, **I**, and **J**, error bars denote ± 1 SEM.

Facilitating quantitative analyses, we designed arenas with a hidden underlying triangular tessellation, where each triangular unit (covering 0.67% of the total area) constituted a ***state*** in a discrete state space (Figure S1a). A fraction of the edges of the tessellation was chosen to be impassable barriers, defined as obstacles. Participants could take ***actions*** to achieve ***transitions*** between adjacent states which were not separated by obstacles. As participants were free to rotate and/or translate, the space of possible actions was continuous such that participants did not report knowledge about the tessellation. Furthermore, participants experienced a relatively high vantage point and were able to gaze over the tops of all of the obstacles (Figure 1a).

On each trial, participants were tasked to collect a ***reward*** by navigating to a random goal location drawn uniformly from all states in the arena. The goal was a realistic banana which the participants had to locate and foveate in order to unlock the joystick. The banana disappeared 200 ms after foveation, as we wanted to discourage beaconing and encourage active sensing behaviors. Participants were instructed to press a button when they believed that they have arrived at the remembered goal location. Then, feedback was immediately displayed on the screen, showing participants that they had received either two points for stopping within the goal state, one point for stopping in a state sharing a border with the goal state (up to three possible), or zero for stopping in any other state. While participants viewed the feedback, a new goal for the next trial was spawned without breaking the continuity of the task. In separate blocks, participants navigated to fifty goals in each of five different arenas (Figure 1b). All five arenas were designed by defining the obstacle configurations such that the arenas varied in the average path length between two states, as quantified by the average state closeness centrality (Methods – Eq 2; S1b) [39]. One of the blocks involved an entirely open arena that contained only a few obstacles at the perimeter, such that on most trials, participants could travel in straight lines to all goal locations (Figure 1b – leftmost). On the other extreme was a maze arena in which most pairs of states were connected by only one viable path (Figure 1b – rightmost). Because lower centrality values correspond to more complex arenas, negative centrality can be interpreted as a measure of arena complexity. For simplicity, we defined complexity using a linear transformation of negative centrality such that the open arena had a complexity value of zero, and we used this scale throughout the paper (see Methods; Figure 1c). We captured both within-participant and between-participant variability by fitting linear mixed effects (LME) models with random slopes and intercepts to predict trial-specific outcomes, and found consistent effects across participants. Therefore, we primarily show average trends in the main text, but participant-specific effects are included in the supplementary figures and tables.

To quantify behavioral performance, we first computed the optimal trajectory for each trial using dynamic programming, an efficient algorithm with guaranteed convergence. This technique uses two pieces of information — the goal location (***reward function***) and the obstacle configuration (***transition structure***) — to find an optimal ***value function*** over all states such that the value of each state is equal to the (negative) length of the shortest path between that state and the goal state (Figures 1d, S1c). The ***optimal policy***

requires that participants select actions to climb the value function along the direction of steepest ascent, which would naturally bring them to the goal state while minimizing the total distance traveled. Figure 1e shows optimal (dashed) as well as behavioral (colored) trajectories from an example trial in each arena. Behavioral path lengths were computed by integrating changes in the participants' position in each trial.

Although participants occasionally took a suboptimal route (Figure 1e – second from right), they took near-optimal paths (i.e. optimal to within the width of the reward zone) on most trials (Figure 1f), scoring (mean $\pm$ SD across participants) 72$\pm$7% of the points across all arenas and stopping within the reward zone on 85$\pm$6% of all trials (Figures S1d-e). We quantified the degree of optimality by computing the ratio of observed vs. optimal path lengths to the participants' stopping location. Across all rewarded trials, this ratio was close to unity (1.1$\pm$0.1), suggesting that participants were able to navigate efficiently in all arenas (Figure 1g). Navigational performance was near-optimal from the beginning, such that there was no visible improvement with experience (Figure S2a). Even on unrewarded trials, participants took a trajectory that is, on average, only 1.2$\pm$0.1 times the optimal path length from the participant's initial state to their stopping location (Figures 1e – rightmost, S2b). This suggests that remembering the goal location was not straightforward. In fact, the fraction of rewarded trials decreased with increasing arena complexity (Pearson's r(63) = -0.64, p = 8 x10^{-9}), suggesting that the ability to remember the goal location is compromised in challenging environments (Figures S2c). Each trial poses unique challenges for the participant, such as the number of turns in the trajectory, the length of the trajectory, and the angle between the initial direction of heading and the direction of target approach (relative bearing). Among these variables, the length of the trajectory best predicts the error in the participants' stopping position (Figure S1f).

In order to understand how participants tackled the computational demands of the task, it is critical to break down each trial into three main epochs: ***search*** – when participants sought to locate the goal, ***pre-movement*** – when participants surveyed their route prior to utilizing the joystick, and ***movement*** – when participants actively navigated to the remembered goal location (Figure 1h). On some trials, participants did not end the trial via button press immediately after stopping, but this post-movement period constituted a negligible proportion of the total trial time. Although participants spent a major portion of each trial navigating to the target, the relative duration of other epochs was not negligible (mean fraction $\pm$ SD – search: 0.27$\pm$0.05, pre-movement: 0.11$\pm$0.03, movement: 0.60$\pm$0.06; Figure S2d). There was considerable variability across participants in the fraction of time spent in the pre-movement phase (coefficient of variation (CV) – search: 0.18, pre-movement: 0.31, movement: 0.10), although this did not translate to a significant difference in navigational precision (Figure S2e). One possible explanation is that some participants were simply more efficient planners or were more skilled at planning on the move. While the duration of the search epoch was similar across arenas, the movement epoch duration increased drastically with increasing arena complexity (Figure 1i). This was understandable as the more complex arenas posed, on average, longer trajectories and more winding paths by virtue of their lower centrality values. Notably, the pre-movement duration was also higher in more complex arenas, reflecting the participant's commitment to meet the increased planning demands in those arenas (Figures 1j). Nonetheless, the relative pre-movement duration was similar for rewarded and unrewarded trials (Figures S2f). This suggests that the participants' performance is limited by their success in remembering the reward location, rather than in meeting planning demands. On a finer scale, the duration of the pre-movement and movement epochs were both strongly influenced by the path length and the number of turns, but not by the bearing angle (Figure S2g). Overall, these results suggest that in the presence of unambiguous visual information, humans are capable of adapting their behavior to efficiently solve navigation problems in complex, unfamiliar environments.

A computational analysis supports that human eye movements are task relevant

Aiming to gain insights from participants' eye movements, we begin by examining the spatial distribution of gaze positions during different trial epochs (Figure 2a). Within each trial, the spatial spread of the gaze position was much larger during visual search than during the other epochs (mean spread $\pm$ SD across participants – search: 5.8$\pm$1.1 m, pre-movement: 2.0$\pm$0.2 m, movement: 3.0$\pm$0.6 m; Figure 2b – left). This pattern was reversed when examining the spatial spread across trials (mean spread $\pm$ SD

– search: 5.8±0.7 m, pre-movement: 6.5±0.3 m, movement: 6.4±0.4; Figure 2b – right). This suggests that participants' eye movements during pre-movement and movement were chiefly dictated by trial-to-trial fluctuations in task demands. Furthermore, the variance of gaze positions within a trial, both prior to and during movement, was largely driven by the path length (Figure 2c) and therefore increased with arena complexity (Figure S3a).

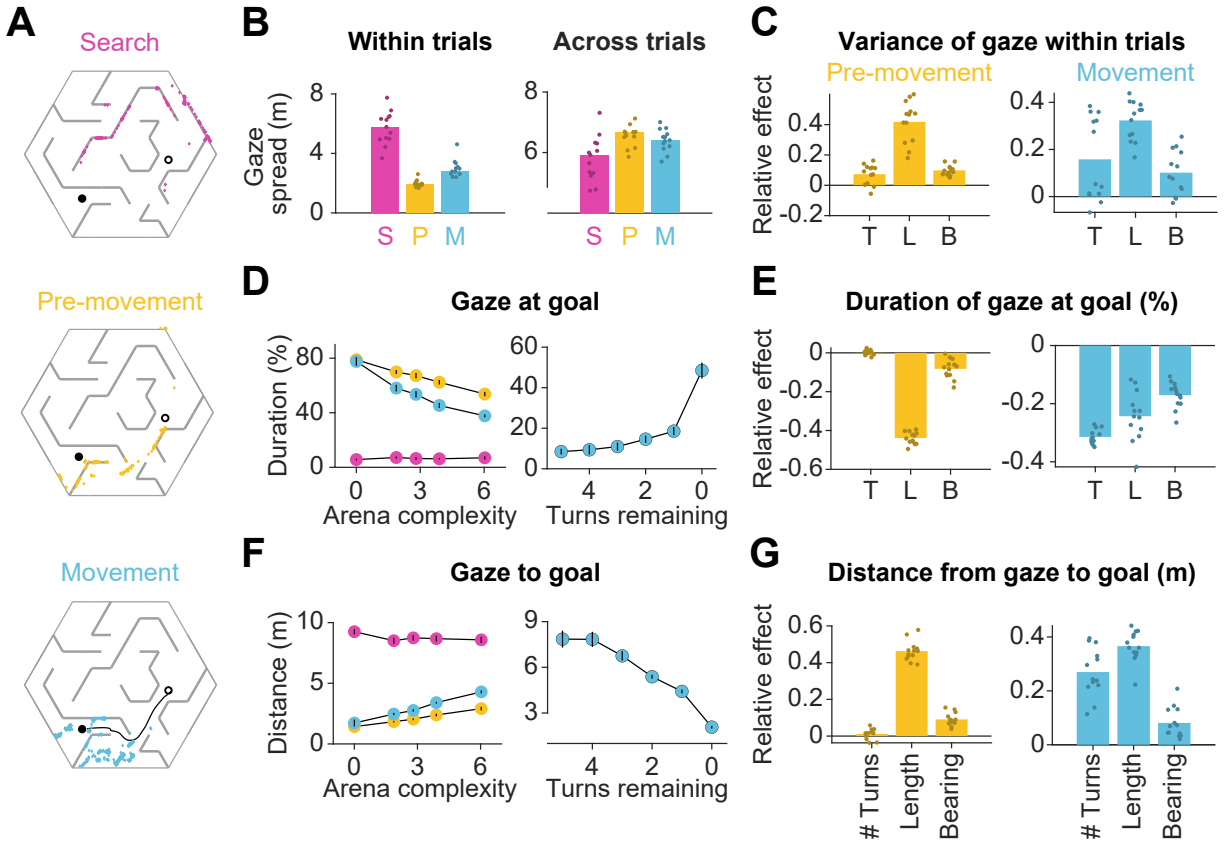

Figure 2: **Eye movements are modulated by goal location and environment complexity. A.** Eye positions on a representative trial for one participant during the three main trial epochs. Each datapoint corresponds to one frame. An open black circle denotes the start location, while a closed black circle denotes the goal. The color scheme applies to all plots in this figure. **B.** Left: The median spatial spread of gaze within trial epochs (averaged across trials and arenas) was higher during search than during pre-movement and movement. Right: In contrast, the median spread of the average gaze positions across trials was higher during the pre-movement and movement epochs. Individual participant data are overlaid on top of the bars. **C.** Left: A linear mixed model for the effect of trial-specific variables (number of turns, length of optimal trajectory, relative bearing) on the variance of gaze within the pre-movement epoch reveals that the expected path length has the greatest effect on gaze spread. The overlaid scatter shows fixed effect slope + participant-specific random effect slope. Right: Similar result for gaze spread within the movement epoch. **D.** Left: Across participants, the average fraction of time for which gaze was near (within 2 m of) the center of the goal state decreased with arena complexity. The arena-level variable (complexity) and the trial-level equivalent (path length) both independently exert effects on the amount of time subjects looked at the goal (Figure S11b). Right: Participants spent more time looking near the goal location when fewer turns separated them from the goal. **E.** Left: A linear mixed model reveals that expected path length had the greatest negative effect on the fraction of time that participants spent gazing at the goal location prior to movement. Right: During movement, all measures of trial difficulty decreased goal-fixation behavior, especially the number of turns. **F.** Left: The average distance between the gaze position and the goal state increased with arena complexity during pre-movement and movement. Right: The average distance of the point of gaze from the goal location decreases as the participant approaches the target. **G.** Left: Expected path length best predicted the average distance of gaze to the goal prior to movement. Right: During movement, the number of turns and the expected path length most positively affected this statistic. All error bars denote ±1 SEM, and all variables were z-scored prior to model fitting.

How did the task demands constrain human eye movements? Studies have shown that reward circuitry tends to orient the eyes toward the most valuable locations in space [40, 41]. Moreover, when the goal

is hidden, it has been argued that fixating the hidden reward zone may allow for the oculomotor circuitry to carry the burden of remembering the latent goal location [38, 42]. Consistent with this, participants spent a large fraction of time looking at the reward zone, and this statistic was interestingly higher during pre-movement (66±10%) than during movement (54±6%). However, goal fixation decreased with arena complexity (Figures 2d – left, S4a – left), resulting in a larger mean distance between the gaze and the goal in more complex arenas (Figures 2f – left, S4a – right). This effect could not be attributed to participants forgetting the goal location in more complex arenas, as we found a similar trend when analyzing gaze in relation to the eventual stopping location (which could be different from the goal location; Figure S3b). A more plausible explanation is that looking solely at the goal might prevent participants from efficiently learning the task-relevant transition structure of the environment, as the structure is both more instrumental to solving the task and harder to comprehend in more challenging arenas. If central vision is attracted to the remembered goal location only when planning demands are low, this tendency should become more prevalent as participants approach the target. Indeed, participants spend significantly more time looking at the goal when there is a straight path to the goal than when the obstacle configuration requires that they make at least one turn prior to arriving upon such a straight path (Figures 2d, f – right). Also in alignment with this explanation, trial-level analyses revealed that during pre-movement, the tendency to look at the goal substantially decreased with greater path lengths (Figures 2e, g – left). During movement, goal fixation has more diverse influences from affordances linked to navigation, especially as a greater amount of turns also decreased the amount of time participants dedicated to looking at the remembered goal location (Figures 2e, g – right).

As mentioned earlier, computing the optimal trajectory requires precisely knowing both the reward function as well as the transition structure. While examining the proximity of gaze to the goal reveals the extent to which eye movements are dedicated to encoding the reward function, how may we assess the effectiveness with which participants interrogate the transition structure of the environment to solve the task of navigating from point A to point B? In the case that a participant has a precise model of the transition structure of the environment, they would theoretically be capable of planning trajectories to the remembered goal location without vision. However, in this experiment, the arena configurations were unfamiliar to the participants, such that they would be quite uncertain about the transition structure. The finding that participants achieved near-optimal performance on even the first few trials in each arena (Figure S2a) indicates that humans are capable of using vision to rapidly reduce their uncertainty about the aspects of the model needed to solve the task. This reduction in uncertainty could be accomplished in two ways: (i) by actively sampling visual information about the structure of the arena in the first few trials and then relying largely on the internal model later on after this information is consolidated, or (ii) by actively gathering visual samples throughout the experiment on the basis of immediate task demands on each trial. We found evidence in support of the second possibility — arena-specific pre-movement epoch durations did not significantly decrease across trials (Figure S4b).

Did participants look at the most informative locations? Depending on the goal location, misremembering the location of certain obstacles would have a greater effect on the subjective value of actions than for other obstacles (Figure S5a, "Relevance simulations" in Supplementary Materials). We leveraged this insight and defined a metric to quantify the task-***relevance*** of each transition by computing the magnitude of the change in value of the participant's current state, for a given goal location, if the status of that transition was misremembered:

$$\Omega_k(s_0, s_G) = [V(s_0|T_k = 1) - V(s_0|T_k = 0)]^2 \tag{1}$$

where $\Omega_k(s_0, s_G)$ denotes the relevance of the kth transition for navigating from state s_0 to the goal state s_G, T_k denotes the status of that transition (1 if it is passable and 0 if it is an obstacle), and $V(s_0|T_k = 1)$ denotes the value of state s_0 computed with respect to the goal state s_G by setting T_k to 1. It turns out that this measure of relevance is directly related to the magnitude of expected change in subjective value of the current state when looking at the kth transition, provided that the transitions are stationary and the participant's uncertainty is uniform across transitions (Supplementary Notes). Thus, maximally relevant transitions identified by Equation 1 are precisely those which may engender the greatest changes of mind about the current return (i.e. utility) of the current state. On each trial, the transitions with the highest relevance

strikingly correspond to bottleneck transitions that bridge clusters of interconnected states (Figures S5b). Relevance was also high for obstacles that precluded a straight path to the goal, as well as for transitions along the optimal trajectory (Figures S5c). By defining relevance of transitions according to Equation 1, we can thus capture multiple task-relevant attributes in a succinct manner. In the supplementary notes, we point to a generalization of this relevance measure for settings in which the transition structure is stochastic (e.g. in volatile environments) and the subjective uncertainty is heterogeneous (i.e. the participant is more certain about some transitions than others).

We quantified the usefulness of participants' eye position on each frame as the relevance of the transition closest to the point of gaze, normalized by that of the most relevant transition in the entire arena given the goal state on that trial. Then, we constructed a distribution of shuffled relevance values by analyzing gaze with respect to a random goal location. Figure 3a shows the resulting cumulative distributions across trials for the average participant during the three epochs in an example arena. As expected, the relevance of participants' gaze was not significantly different from chance during the search epoch, as the participant had not yet determined the goal location. However, relevance values were significantly greater than chance both during pre-movement and movement (median relevance for the most complex arena, pre-movement – true: 0.14, shuffled: 0.006; movement – true: 0.20, shuffled: 0.06; see Figure S7, Table S2 for other arenas).

To concisely describe participants' tendency to orient their gaze toward relevant transitions in a scale-free manner, we constructed receiver operating characteristic (ROC) curves by plotting the cumulative probability of shuffled gaze relevances against the cumulative probability of true relevances (Figure 3a – rightmost). An area under the ROC curve (AUC) greater (less) than 0.5 would indicate that the gaze relevance was significantly above (below) what is expected from a random gaze strategy. Across all arenas, the AUC was highest during the pre-movement epoch (Figure 3b; mean AUC $\pm$ SD – search: 0.52 ± 0.03, pre-movement: 0.77 ± 0.03, movement: 0.68 ± 0.07). This suggests that participants were most likely to attend to relevant transitions when contemplating potential actions before embarking upon the trajectory.

As the most relevant transitions can sometimes be found near the goal (e.g. Figure S5b – left), we investigated whether our evaluation of gaze relevance was confounded by the observation that participants spent a considerable amount of time looking at the goal location (Figure 2c). Therefore, we first quantified the tendency to look at the goal location in a manner analogous to the analysis of gaze relevance (Figures 3a-b) by computing the area under the ROC curves (AUC) constructed using the true vs. shuffled distributions of the duration spent foveating the goal in each epoch (Figures 3c). Across all arenas, AUCs were high during the pre-movement and movement epochs, confirming that there was a strong tendency for participants to look at the goal location (Figure 3d). When we excluded gaze positions that fell within the reward zone while computing relevance, we found that the degree to which participants looked at task-relevant transitions outside of the reward zone increased with arena complexity: the tendency to look at relevant transitions was greater in more complex arenas, falling to chance for the easiest arena (Figure 3e; Pearson's $r(63)$ – pre-movement: 0.40, $p = 0.004$; movement: 0.28, $p = 0.05$). In contrast, the tendency to look at the goal location followed the opposite trend and was greater in easier arenas (Figure 3f; Pearson's $r(63)$ – pre-movement: -0.46, $p = 0.0003$; movement: -0.33, $p = 0.001$). These analyses reveal a striking trade-off in the allocation of gaze between encoding the reward function and transition structure that closely mirrors the cognitive requirements of the task. This trade-off is not simply a consequence of directing gaze more/less often at the reward, as such temporal statistics are preserved while shuffling. Instead, it points to a strategy of directing attention away from the reward and towards *task-relevant* transitions in complex arenas. This compromise allowed participants to dedicate more time to surveying the task-relevant structure in complex environments and likely underlies their ability to take near-optimal paths in all environments, albeit at the cost of an increased tendency to forget the precise goal location in complex environments (Figure S2c). The trade-off reported here is roughly analogous to the trade-off between looking ahead towards where you're going and having to pay attention to signposts or traffic lights. One could get away with the former strategy while driving on rural highways whereas city streets would warrant paying attention to many other aspects of the environment to get to the destination.

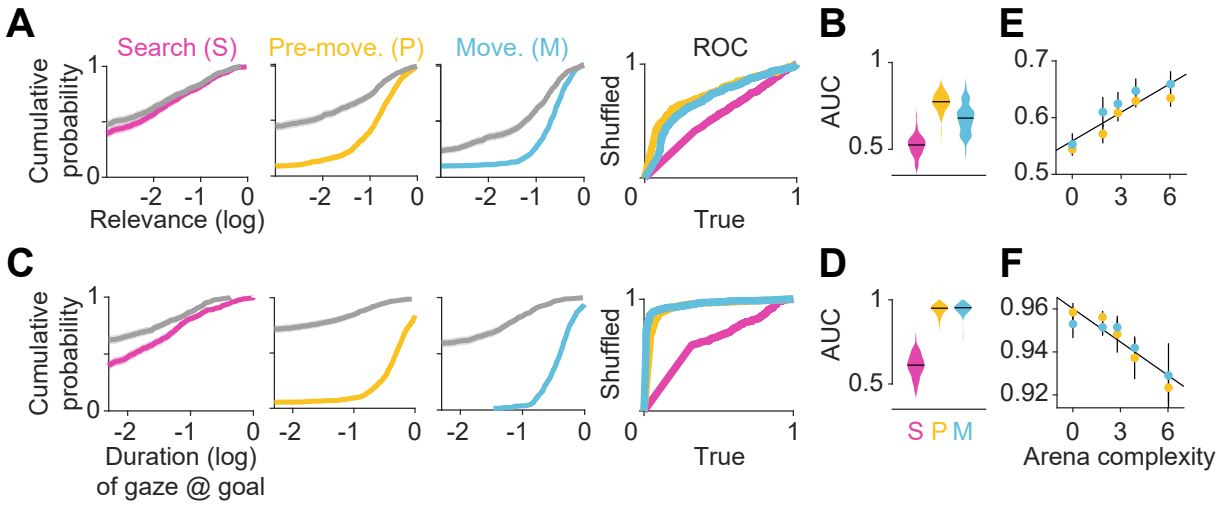

Figure 3: **Eye movements reveal a cognitive trade-off between reward and transition encoding. A.** Left: Cumulative distribution (computed by pooling trials from all participants) of average log (normalized) relevance values (colored line) and the corresponding shuffled distribution (gray) during search (left), pre-movement (center), and movement (right) epochs (data for the most complex arena is shown). Shaded regions denote 95% confidence bounds computed using Greenwood's Formula. Rightmost: ROC curves characterizing the gaze relevance during the three epochs. **B.** Area under the ROC curves (AUC) for different epochs, colored according to the color scheme in **A**. **C-D**: Similar plots as **A-B**, but for the distributions of the log fraction of the duration in each epoch spent gazing near (within 2m of) the eventual stopping position (which was assumed to be the participants' believed goal location). **E.** AUC values of gaze relevance computed for the distributions of trial-averaged relevances, after excluding fixations within the reward zone, during the pre-movement (orange) and movement (blue) epochs. Black line represents best-fit linear regression model. **F.** Similar to **E**, but showing the AUC values of gaze durations within the reward zone. All error bars were computed using bootstrapping.

The temporal evolution of gaze includes distinct periods of sequential prospection

So far, we have shown that the spatial distribution of eye movements adapts to trial-by-trial fluctuations in task demands induced by changing the goal location and/or the environment. However, planning and executing optimal actions in this task requires dynamic cognitive computations within each trial. To gain insights into this process, we examined the temporal dynamics of gaze. Figure 4a shows a participant's gaze in an example trial which has been broken down into nine epochs (pre-movement: I-VI, movement: VII-IX) for illustrative purposes (see Video S1 for more examples). The participant initially foveated the goal location (epoch I), and their gaze subsequently traced a trajectory *backwards* from the goal state towards their starting position (II) roughly along a path which they subsequently traversed on that trial (dotted line). This sequential gaze pattern was repeated shortly thereafter (IV), interspersed by periods of non-sequential eye movements (III and V). Just before embarking on their trajectory, the gaze traced the trajectory, now in the *forward* direction, until the end of the first turn (VI). Upon reaching the first turning point in their trajectory (VII), they executed a similar pattern of sequential gaze from their current position toward the goal (VIII), tracing out the path which they navigated thereafter (IX). We refer to the sequential eye movements along the future trajectory in the backwards and forwards direction as ***backward sweeps*** and ***forward sweeps***, respectively. During such sweeps, participants seemed to rapidly navigate their future paths with their eyes, and all participants exhibited sweeping eye movements without being explicitly instructed to plan their trajectories prior to navigating. The fraction of time that participants looked near the trajectories which they subsequently embarked upon increased with arena and trial difficulty (Figure S8a). To algorithmically detect periods of sweeps, gaze positions on each trial were projected onto the trajectory taken by the participant by locating the positions along the trajectory closest to the point of gaze on each frame (Methods). On each frame, the length of the trajectory up until the point of the gaze projection was divided by the total trajectory length, and this ratio was defined as the "fraction of trajectory". We used the increase/decrease

of this variable to determine the start and end times of periods when the gaze traveled sequentially along the trajectory in the forward/backward directions (sweeps) for longer than chance (Figure 4b; see Methods).

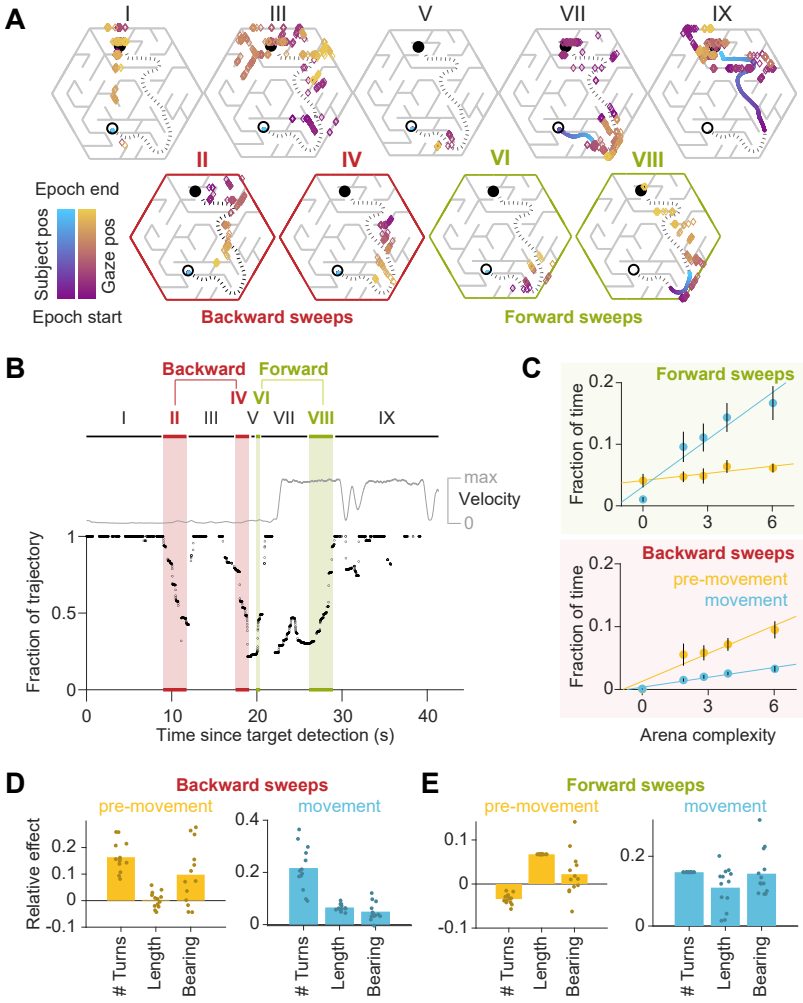

Figure 4: **Gaze traveled forwards and backwards along the intended trajectory. A.** Spatial locations of gaze positions (the arrow of relative time within each window increases from violet to orange) and participant positions (violet to blue) during individual time windows demarcated in panel B. Panels in the bottom row correspond to time periods corresponding to sweeps. The participant's trajectory from the starting location (open black circle) to the goal (closed black circle) is denoted by a black dashed line. **B.** Time-series of the points on the trajectory that were closest to the participant's gaze on each frame, expressed as a fraction (0: start of trajectory, 1: end of trajectory) during one example trial. Only frames during which the gaze position fell within 2 m of the trajectory are plotted. The gray trace shows the movement velocity of the participant during this trial. Red and green shaded regions highlight time windows during which the sweep classification algorithm detected backward and forward sweeps, respectively. In this trial, there were two backward sweeps before movement, and one forward sweep each before and during movement. **C.** Across all participants, the fraction of time spent sweeping in the forward and backward directions within each epoch reveals an antiparallel effect: more time was spent sweeping forwards during movement than during pre-movement (top), whereas more time was spent sweeping backwards during pre-movement than during movement (bottom). Generally, the arena complexity as well as the trial-specific path lengths, both increase the fraction of time sweeping (Figure S11c). Error bars denote ± 1 SEM. **D.** Linear mixed models with random intercepts and slopes for the effect of trial-specific variables (number of turns, length of optimal trajectory, relative bearing) on the fraction of time that participants spent sweeping their trajectory in the backward direction, separated for pre-movement and movement epochs. The overlaid scatter shows fixed effect slope + participant-specific random effect slope. **E.** Similar analysis as **D**, but for forward sweeps. All variables were z-scored prior to model fitting.

The environmental structure exerted a strong influence on the probability of sweeping: the fraction of trials in

which this phenomenon occurred was significantly correlated with arena complexity (Pearson's r(63) = 0.73, p = 5e-12; Figure S8b). This suggests that sweeping eye movements could be integral to trajectory planning. Most notably, on average, backward sweeps occupied a greater fraction of time during pre-movement than during movement, but forward sweeps predominantly occurred during movement (backward sweeps – pre-movement: 5.6±3.3%, movement: 1.9±0.6%; forward sweeps – pre-movement: 5.2±3.0%, movement: 10.5±6.3%; Figure 4c). This suggests that the initial planning is primarily carried out by sweeping backwards from the goal. Furthermore, trials with a greater number of turns and those in which participants initially move away from the direction of the target tend to have more backward sweeps during pre-movement (Figure 4d – left). In contrast, the number of turns inhibited forward sweeps during pre-movement, which were instead driven largely by the length of the trajectory (Figure 4e – left). During movement on the other hand, multiple measures of trial difficulty increased the likelihood of forward sweeps (Figure 4e – right), whereas backward sweeps depended primarily on the number of turns (Figure 4d – right). Together, these dependencies explain why backward sweeps are more common during pre-movement but forward sweeps dominate during movement.

Aside from gazing upon the target or the trajectory on each trial, about 20% of eye movements were made to other locations in space (Figure S8c). Besides task-relevant locations such as bottleneck transitions, we wanted to know whether these other locations also comprised alternative trajectories to the goal. To test this, we identified all trajectories whose path lengths were comparable (within about 1 SD; Methods) to the chosen trajectory on each trial. The fraction of time spent looking at alternative trajectories and the chosen trajectory both increased with the number of alternatives (Pearson's r(14) = 0.91, p = 1e-6 pre-movement; r = 0.90, p = 2e-6 movement), suggesting that participants engage in some form of deliberation (Figure S8d). This deliberation happens at the expense of reducing the fraction of time spent looking at the hidden goal location (r(14) = -0.85, p = 3e-5 pre-movement; r = -0.64, p = 7e-3 movement), revealing that the planning algorithm is likely subject to a cognitive trade-off mediated by the number of available options, in addition to that mediated by model complexity demonstrated earlier.

When looking at the trajectory, the mean speed of backward sweeps was greater than the speed of forward sweeps across all arenas (backward sweeps: 26±4 m/s, forward sweeps: 21±3 m/s; Figure S9a). Notably, sweep velocities were more than 10x greater than the average participant velocity during the movement epoch (1.4±0.1 m/s). This is reminiscent of the hippocampal replay of trajectories through space, as such sequential neural events are also known to be compressed in time (around 2-20x the speed of neural sequence activation during navigation) [43]. Both sweep speeds and durations slightly increased with arena complexity (Figure S9a). This is because peripheral vision processing must lead the control of central vision to allow for sequential eye movements to trace a viable path [44, 45]. In more complex arenas such as the maze where the search tree is narrow and deep, the obstacle configuration is more structured and presents numerous constraints, and thus path tracing computations might occur more quickly. Accordingly, longer path lengths best predicted sweep speeds (Figure S9b). However, due to the lengthier and more convoluted trajectories in those arenas, the gaze must cover greater distances and make more turns, resulting in sweeps which last longer (Figures S9a, S9c). Another property of sweeps is that they comprised more saccades in more difficult trials (Figures S9a, S9d), and saccade rates were higher during sweeps than at other times after goal detection (Figure S9f). This suggests that either visual processing during sweeps was expedited compared to average, or sweeps resulted from eye movements which followed a pre-planned saccade sequence.

If the first sweep on a trial occurred during pre-movement, the direction of the sweep was more likely to be backwards, while if the first sweep occurred during movement, it was more likely to be in the forwards direction (Figures S10a, b). The latency between goal detection and the first sweep increased with arena difficulty (Figure S10c – top row), and more specifically the number of turns and expected path length (Figure S10c – bottom row), suggesting that sweep initiation is preceded by brief processing of the arena, and more complex tasks elicited longer processing. While the sequential nature of eye movements could constitute a swift and efficient way to perform instrumental sampling, we found that task-relevant eye movements were not necessarily sequential. When we reanalyzed the spatial distribution of gaze positions by removing periods of sweeping, the resulting relevance values remained greater than chance (Figure S10d).

What task conditions promote sequential eye movements? To find out, we computed the probability that the participants engaged in sweeping behavior as a function of time and position, during the pre-movement and movement epochs respectively, and focusing on the predominant type of sweep during those periods (backward and forward sweeps respectively; Figure 4c). During pre-movement, we found that the probability of sweeping gradually increased over time, suggesting that backward sweeps during the initial stages of planning are separated from the time of target foveation by a brief pause, during which participants may be gathering some preliminary information about the environment (Figure 5a – left). During movement, on the other hand, the probability of sweeping is strongly influenced by whether participants are executing a turn in their trajectory. Obstacles often preclude a straight path to the remembered goal location, and thus participants typically find themselves making multiple turns while actively navigating. Consequently, a trajectory may be divided into a series of straight segments separated by brief periods of elevated angular velocity. We isolated such periods by applying a threshold on angular velocity, designating the periods of turns as ***subgoals***, and aligned the participant's position in all trials with respect to the subgoals. The likelihood of sweeping the trajectory in the forward direction tended to spike precisely when participants reached a subgoal (Figure 5a – right). There was a concomitant decrease in the average distance of the point of gaze from the goal location in a step-like manner with each subgoal achieved (Figure 5b – right). In contrast to backward sweeps, which were made predominantly to the most proximal subgoal prior to navigating (Figure 5c – left), forward sweeps that occurred during movement were not regularly directed toward one particular location. Instead, in a strikingly stereotyped manner, participants appeared to lock their gaze upon the *upcoming* subgoal when rounding each bend in the trajectory (Figure 5c – right). This suggests that participants likely represented their plan by decomposing it into a series of subgoals, focusing on one subgoal at a time until they reached the final goal location. In contrast to sweeping eye movements, the likelihood of gazing alternative trajectories peaked much earlier during pre-movement (Figure 5d – left). Likewise, during movement, participants tend to look briefly at alternative trajectories shortly before approaching a subgoal (Figure 5d – right) which might constitute a form of vicarious trial and error behavior at choice points [46].

To summarize, we found participants made sequential eye movements sweeping forward and/or backward along the intended trajectory, and the likelihood of sweeping increased with environmental complexity. During the pre-movement phase, participants gathered visual information about the arena, evaluated alternative trajectories, and typically traced the chosen trajectory backwards from the goal to the first subgoal (Figure 5e – orange). While moving through the arena, they tended to lock their gaze upon the upcoming subgoal until shortly before a turn, at which point they exhibit a higher tendency of gazing upon alternative trajectories. During turns, participants often sweep their gaze forward to the next subgoal (Figure 5e – blue). Via eye movements, navigators could construct well-informed plans to make sequential actions that would most efficiently lead to rewards (Figure 5f).

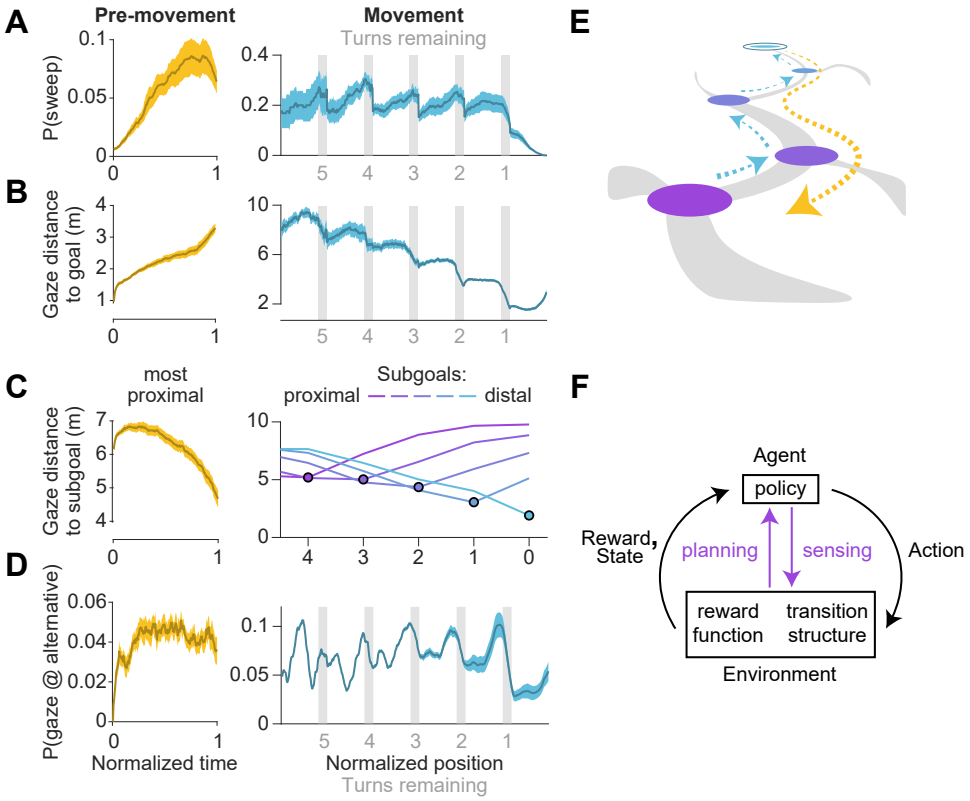

Figure 5: **Timing of sweeps reveals task decomposition.** Trials across all arenas and all participants were aligned and scaled for the purpose of trial-averaging. This process was carried out separately for the pre-movement and movement epochs. **A.** Left: Prior to movement, the probability of (backward) sweeps increased with time. Right: During movement, the probability of (forward) sweeps transiently increased at the precise moments when participants reached each subgoal. Participant position is defined in relation to the location of subgoals. Subgoals are designated as numbers starting from the goal (subgoal 0) and counting backwards along the trajectory (subgoals 1, 2, 3 etc.) such that greater values correspond to more proximal subgoals. **B.** Left: Gaze traveled away from the goal location prior to movement. Right: The average distance of gaze from the goal decreased in steps, with steps occurring at each subgoal. **C.** Distance of gaze from individual subgoals (most proximal in yellow, most distal in cyan). Left: Gaze traveled towards the most proximal subgoal prior to movement, consistent with the increased probability of backward sweeps during this epoch. Right: The average distance of gaze to each individual subgoal (colored lines) was minimized precisely when participants approached that subgoal. **D.** Left: The probability of gazing at alternative trajectories is relatively constant throughout the pre-movement epoch. Right: Participants gaze at alternative trajectories more frequently when approaching turns. **E.** A graphical summary of the spatiotemporal dynamics of eye movements in this task. Subgoals are depicted in the same color scheme used in **C**. **F.** Diagram of a standard Markov Decision Process, augmented with an additional pathway for agent-environment interaction through eye movements (colored arrows). Dashed arrows denote sweeps, and possible paths throughout the arena are depicted in gray. Darker bounds in **A-C** denote ± 1 SEM.

Discussion

In this study, we highlight the crucial role of eye movements for flexible navigation. We found that humans took trajectories nearly optimal in length through unfamiliar arenas, and spent more time planning prior to navigation in more complex environments. The spatial distribution of gaze was largely concentrated at the hidden goal location in the simplest environment, but participants increasingly interrogated the task-relevant structure of the environment as the arena complexity increased. In the temporal domain, participants often rapidly traced their future trajectory to and from the goal with their eyes (sweeping), and generally concentrated their gaze upon one subgoal (turn) at a time until they reached their destination. In summary, we found evidence that the neural circuitry governing the oculomotor system optimally schedules and allocates resources to tackle the diverse cognitive demands of navigation, producing efficient eye movements through space and time.

Eye movements provide a natural means for researchers to understand information seeking strategies, in both experimental and real-world settings [47, 48]. Past studies using simple decision making tasks probed whether active sensing, specifically via eye movements, reduces uncertainty about the *state* of the environment (such as whether a change in an image has occurred) [7, 8, 49]. But in sequential decision making tasks such as navigation, there is added uncertainty about the task-contingent causal structure (*model*) of the environment [2, 6]. Common paradigms for goal-oriented navigation occlude large portions of the environment from view, usually in the interest of distinguishing model-based strategies from conditioned responses [50–52] or allow very restricted fields of view where eye movements have limited potential for sampling information [39, 53]. Occlusions eliminate the possibility of gathering information about the structure of the environment using active sensing. By removing such constraints, we allowed participants to acquire a model of the environment without physically navigating through it, which yielded new insights about how humans perform visually-informed planning.

In particular, we found that the gaze is distributed between the two components of the model required to plan a path —- the transition function, which describes the relationship between states, and the reward function, which describes the relationship between the states and the reward —- with the distribution skewed in favor of the former in more complex environments. When alternative paths were available, gaze tended to be directed towards them at the expense of looking at the hidden reward location. These findings suggest a context-dependent mechanism which dictates the dynamic arbitration between competing controllers of the oculomotor system that seek information about complementary aspects of the task. Neurally, this could be implemented by circuits that exert executive control over voluntary eye movements. Candidate substrates include the dorsolateral prefrontal cortex, which is known to be important for contextual information processing and memory-guided saccades [54–56], and the anterior cingulate cortex, which is known to be involved in evaluating alternative strategies [57, 58]. To better understand the precise neural mechanisms underlying the spatial gaze patterns we observed, it would be instructive to examine the direction of information flow between the oculomotor circuitry and brain regions with strong spatial and value representations during this task in animal models. Future research may also investigate multi-regional interactions in humans by building on recent advances in data analysis that allow for eye movements to be studied in fMRI scanners [59]. Our analysis of spatial gaze patterns is grounded in the RL framework, which provides an objective way to measure the utility of sampling information from different locations. However, this measure was agnostic to the temporal ordering of those samples. Given that previous work demonstrated evidence for the planning of multiple saccades during simple tasks like visual search [11], incorporating chronology into a normative theory of eye movements in sequential decision-making tasks presents an excellent opportunity for future studies.

Meanwhile, we found that the temporal pattern of eye movements revealed a fine-grained view of how planning computations unfold in time. In particular, participants made sequential eye movements sweeping forward and/or backward along the future trajectory, evocative of forward and reverse replay by place cells in the hippocampus [35, 60]. Shortly after fixating on the goal, participants' gaze often swept backwards along their future trajectory, mimicking reverse replay. Because these sweeps predominantly occurred be-

fore movement, they may reflect depth-first tree search, a model-based algorithm for path discovery [61]. Then, during movement, participants were more likely to make forward sweeps when momentarily slowing down at turning points, analogous to finding that neural replay mainly occurs during periods of relative immobility [62]. Several recent studies have also supported that replay serves to consolidate memory and generalize information about rewards [63–65]. In light of the similarities between sweeps and sequential hippocampal activations, we predict that direct or indirect hippocampal projections to higher oculomotor controllers (e.g. the supplemental eye fields through the orbitofrontal cortex) may allow eye movements to embody the underlying activations of state representations [66–68]. This would allow replays to influence the active gathering of information. Alternatively, active sensing could be a result of rapid peripheral vision processing which drives saccade generation, such that the eye movements reflect the outcome of sensory processing rather than prior experience. Consistent with this idea, past studies have demonstrated that humans can smoothly trace paths through entirely novel 2D mazes [45, 69]. Interestingly, neural modulation does occur in this direction — the contents of gaze have been found to influence activity in the hippocampus and entorhinal cortex [70–75]. Therefore, it is conceivable that sequential neural activity could emerge from consolidating temporally extended eye movements such as sweeps. We hope that in future, simultaneous recordings from brain areas involved in visual processing, eye movement control, and the hippocampal formation would uncover the mechanisms underlying trajectory sweeping eye movements and their relationship to perception and memory.

Value-based decisions are known to involve lengthy deliberation between similar alternatives [76, 77]. Participants exhibited a greater tendency to deliberate between viable alternative trajectories at the expense of looking at the reward location. Likelihood of deliberation was especially high when approaching a turn, suggesting that some aspects of path planning could also be performed on the fly. More structured arena designs with carefully incorporated trajectory options could help shed light on how participants discover a near-optimal path among alternatives. However, we emphasize that deliberative processing accounted for less than one-fifth of the spatial variability in eye movements, such that planning largely involved searching for a viable trajectory.

Although we have analyzed strategies of active sensing and planning separately, these computations must occur simultaneously and influence each other. This is formalized by the framework of active inference that unifies planning and information seeking by integrating the RL framework, which describes exploiting rewards for their extrinsic value, and the information theoretic framework, which describes exploring new information for its epistemic value [6]. Using this framework to simulate eye movements in a spatial navigation task, Kaplan and Friston found that gaze is dominated by epistemic (curiosity) rather than pragmatic (reward) considerations in the first few trials, a prediction that is not supported by our results. However, it is possible that participants were able to rapidly resolve uncertainty about the arena structure in our experiments. Future studies must identify the constraints under which active inference models can provide quantitatively good fits to our data. In another highly relevant theoretical work, Mattar and Daw proposed that path planning and structure learning are variants of the same operation, namely the spatiotemporal propagation of memory [28]. The authors show that prioritization of reactivating memories about reward encounters and imminent choices depends upon its utility for future task performance. Through this formulation, the authors provided a normative explanation for the idiosyncrasies of forward and backward replay, the overrepresentation of reward locations and turning points in replayed trajectories, and many other experimental findings in the hippocampus literature. Given the parallels between eye movements and patterns of hippocampal activity, it is conceivable that gaze patterns can be parsimoniously explained as an outcome of such a prioritization scheme. But interpreting eye movements observed in our task in the context of the prioritization theory requires a few assumptions. First, we must assume that traversing a state space using vision yields information that has the same effect on the computation of utility as does information acquired through physical navigation. Second, peripheral vision allows participants to form a good model of the arena such that there is little need for active sensing. In other words, eye movements merely reflect memory access and have no computational role. Finally, long-term statistics of sweeps gradually evolve with exposure, similar to hippocampal replays. These assumptions can be tested in future studies by titrating the precise amount of visual information available to the participants, and by titrating their experience and characterizing gaze over longer exposures. We suspect that a pure prioritization-based account might be

sufficient to explain eye movements in relatively uncluttered environments, whereas navigation in complex environments would engage mechanisms involving active inference. Developing an integrative model that features both prioritized memory-access as well as active sensing to refine the contents of memory, would facilitate further understanding of computations underlying sequential decision-making in the presence of uncertainty.

The tendency of humans to break larger problems into smaller, more tractable subtasks has been previously established in domains outside of navigation [29, 78–81]. However, theoretical insights on clustered representations of space have not been empirically validated in the context of navigation [82, 83], primarily due to the difficulty in distinguishing between flat and hierarchical representations from behavior alone. Our observation that participants often gazed upon the upcoming turn during movement supports that participants viewed turns as subgoals of an overall plan. Future work could focus on designing more structured arenas to experimentally separate the effects of path length, number of subgoals, and environmental complexity on participants' eye movement patterns.

We hope that the study of visually-informed planning during navigation will eventually generalize to understanding how humans accomplish a variety of sequential decision-making tasks. A major goal in the study of neuroscience is to elucidate the principles of biological computations which allow humans to effortlessly exceed the capabilities of machines. Such computations allow animals to learn environmental contingencies and flexibly achieve goals in the face of uncertainty. However, one of the main barriers to the rigorous study of active, goal-oriented behaviors is the complexity in estimating the participant's prior knowledge, intentions, and internal deliberations which lead to the actions that they take. Luckily, eye movements reveal a wealth of information about ongoing cognitive processes during tasks as complex and naturalistic as spatial navigation.

Methods

Experimental Model and Participant Details. Thirteen human participants (all >18 years old, ten males) participated in the experiments. All but two participants (S6 and S9) were unaware of the purpose of the study. Four of the participants, including S6 and S9, were exposed to the study earlier than the rest of the participants, and part of the official dataset for two of these participants (S4 and S8) was collected two months prior to the rest of data collection as a safety precaution during the COVID-19 pandemic. Eight additional human participant recruits (all >18 years old, four males) were disqualified due to experiencing motion sickness while in the VR environment and not completing a majority of trials. All experimental procedures were approved by the Institutional Review Board at New York University and all participants signed an informed consent form.

Stimulus. Participants were seated on a swivel chair with 360° of freedom in physical rotation and navigated in a full-immersion hexagonal virtual arena with several obstacles. The stimulus was rendered at a frame rate of 90 Hz using the Unity game engine v2019.3.0a7 (programmed in `C#`) and was viewed through an HTC VIVE Pro virtual reality headset. The subjective vantage point (height of the point between the participants' eyes with respect to the ground plane) was 1.72 meters. The participant had a field of view of 110.1° of visual angle. Forward and backward translation was enabled via a continuous control CTI Electronics M20U9T-N82 joystick with a maximum speed recorded at 4.75 m/s. Participants executed angular rotations inside the arena by turning their head, while the joystick input enabled translation in the direction in which the participant's head was facing. Obstacles and arena boundaries appeared as gray, rectangular slabs of concrete. The ground plane was grassy, and the area outside of the arena consisted of a mountainous background. Peaks were visible above the outer boundary of the arena to provide crude orientation landmarks. Clear blue skies with a single light source appeared overhead.

State space geometry. The arena was a rectangular hexagon enclosing an area of approximately 260 m^2 of navigable space. For ease of simulation and data analyses, the arena was imparted with a hidden triangular tessellation (*deltille*) composed of $6n^2$ equilateral triangles, where n determines the state space granularity. We chose n = 5, resulting in triangles with a side length of 2 meters, each of which constituted a state in the discrete state space (Figure S1a). The arena contained several obstacles in the form of unjumpable obstacles (0.4 meters high) located along the edges between certain triangles (states). Obstacle locations were predetermined offline using MATLAB by either randomly selecting a chosen number of edges of the tessellation or by using a graphical user interface (GUI) to manually select edges of the tessellation; these locations were loaded into Unity. Outer boundary walls of height 2.5 m enclosed the arena. We chose five arenas spanning a large range in *average* state closeness centrality $\langle C(s) \rangle$ (Eq 2), where $C(s)$ is defined as the inverse average path length d from state s to every other state s' (N states in total). On average, arenas with lower centrality will impose greater path lengths between two given states, making them more complex to navigate. We defined a measure of arena complexity by adding an offset to mean centrality and then scaling it, such that the simplest arena had a complexity value of zero ($-100 \times (\bar{C} - \max[\bar{C}])$) where $\bar{C}$ denotes mean centrality across states and max is taken over arenas. Such a transformation would preserve the correlation and p-values between dependent variables and arena centrality, while the new metric would allow for the graphic representation of arenas in an intuitive order (Figure 1b). A complexity value of zero corresponds to the simplest arena that we designed. The order of arenas presented to each participant was randomly permuted but not entirely counterbalanced due to the large number of permutations (Table S1).

$$C(s) = \frac{N-1}{\Sigma_{s'} d(s', s)} \tag{2}$$

Eye tracking. At the beginning of each block of trials, participants calibrated the VIVE Pro eye tracker using inbuilt Tobii software which prompted participants to foveate several points tiling a 2D plane in the VR environment. Both eyes were tracked, and the participant's point of foveation (*x-y* coordinates), object of foveation (ground, obstacles, boundaries, etc.), eye openness, and other variables of interest were recorded

on each frame using the inbuilt software. Sipatchin et. al. (2020) reported that during free head movements, point-of-gaze measurements using the VIVE Pro eye tracker has a spread of 1.15° ± 0.69° (SE) [84]. This means that when the participant fixates a point on the ground five meters away, the 95% confidence interval (CI) for the measurement error in the reported gaze location would be 0-23 cm (roughly one-tenth of the length of one transition or obstacle) and 0-67 cm (one-third of a transition length) for points fifteen meters away. While machine precision was not factored into the analyses, the fraction of eye positions that may have been misclassified due to hardware and software limitations is likely very small. Furthermore, Sipatchin et. al. reported that the system latency was 58.1 ms. While there is reason to suspect that the participant's position was recorded with a similar latency of around five frames, even if the gaze data lagged the position data, the participant would only have moved 28.5 cm if they were translating at the maximum possible velocity over this interval.

Behavioral task. At the beginning of each trial, a target in the form of a realistic banana from the Unity Asset store appeared hovering 0.4 meters over a state randomly drawn from a uniform distribution over all possible states. The joystick input was disabled until the participant foveated the target, but the participant was free to scan the environment by rotating in the swivel chair during the visual search period. Two hundred milliseconds after target foveation, the banana disappeared and participants were tasked with navigating to the remembered target location without time constraints. Participants were not given instructions on what strategy to use to complete the task. After reaching the target, participants pressed a button on the joystick to indicate that they have completed the trial. Alternatively, they could press another button to indicate that they wished to skip the trial. Feedback was displayed immediately after pressing either button (see section below). Skipping trials was discouraged except when participants did not remember seeing the target before it disappeared, and these trials were recorded and excluded from the analyses (< 1%).

Reward. If participants stopped within the triangular state which contained the target, they were rewarded with two points. If they stopped in a state sharing a border with the target state, they were rewarded with one point. After the participant's button press, the number of points earned on the current trial was displayed for one second at the center of the screen. The message displayed was 'You earned p points!'; the font color was blue if p = 1 or p = 2, and red if p = 0. On skipped trials, the screen displayed 'You passed the trial' in red. In each experimental session, after familiarizing themselves with the movement controls by completing ten trials in a simplistic six-compartment arena (granularity n = 1), participants completed one block of fifty trials in each of five arenas (Figure 1b). At the end of each block, a blue message stating 'You have completed all trials!' prompted them to prepare for the next block. Session durations were determined by the participant's speed and the length of the breaks that they needed from the virtual environment, ranging from 1.5-2 hours, sometimes spread across more than one day. Participants were paid $0.02/point for a maximum of 5 arenas x 50 trials/arena x 2 points/trial x $0.02/point = $10, in addition to a base pay of $10/hour for their time (the average payment was $27.55).

RL formulation. Navigation can be formulated as a Markov Decision Process (MDP) described by the tuple $< S, A, P, R, \gamma >$ whose elements denote, respectively, a finite state space S, a finite action space A, a state transition distribution P, a reward function R, and a temporal discount factor γ that captures the relative preference of distal over proximal rewards [85]. Given that an agent is in state $s \in S$, the agent may execute an action $a \in A$ in order to bring about a change in state $s \to s'$ with probability $P(s'|s, a)$ and harvest a reward $R(s, a)$. To relate this formalism to the structure of the arena, it is instructive to consider the *possibility* of traversal from state s to any state s' in a single time step, as described by the adjacency matrix T: $T(s, s') = 1$ if there exists an available action which would bring about the change in state $s \to s'$ with a non-zero probability, and $T(s, s') = 0$ otherwise. By definition, $T(s, s') = 0$ if there is an obstacle between s and s'. A state not bordered by any obstacle would have three non-zero entries in the corresponding row of T. Thus, the arena structure is fully encapsulated in the adjacency matrix.

In the case that an agent is tasked with navigating to a goal location s_G where the agent would receive a reward, the reward function $R(s, a) > 0$ if and only if the action a allows for the transition $s \to s_G$ in one time step, and $R(s, a) = 0$ otherwise. Given this formulation, we may compute the optimal policy $\pi^*(a|s)$,

which describes the actions that an agent should take from each state in order to reach the target state in the fewest possible number of time steps. The optimal policy may be derived by computing optimal state values $V^*(s)$, defined as the expected future rewards to be earned when an agent begins in state s and acts in accordance with the policy π^*. The optimal value function can be computed by solving the Bellman Equation (Eq 3) via dynamic programming (specifically value iteration) — an efficient algorithm for path-finding — which iteratively unrolls the recursion in this equation [86]. The optimal policy is given by the argument a that maximizes the right-hand side of (3). Intuitively, following the optimal policy requires that agents take actions to ascend the value function where the value gradient is most steep (Figure 1d).

$$V^*(s) = \max_a \left[R(s,a) + \gamma \sum_{s'} P(s'|s,a)V^*(s')\right] \tag{3}$$

For the purposes of computing the optimal trajectory, we considered twelve possible degrees of freedom in the action space, such that one-step transitions could result in relocating to a state that is 0°, 30°, 60°, . . . , 300°, or 330° with respect to the previous state. However, the center-to-center distances between states for a given transition depends on the angle of transition. Specifically, as shown in Figure S1c, if a step in the 0° direction requires translating 1 m, then a step in the 60°, 120°, 180°, 240°, and 300° directions would also require translating 1 m, but a step in the 30°, 150°, and 270° directions would require translating $2\sqrt{3}/3$ m, and a step in the 90°, 210°, and 330° directions would require translating $\sqrt{3}/3$ m. Therefore, in Eq 3, $R(s,a) = -1$, $-2\sqrt{3}/3$, or $-\sqrt{3}/3$, depending on the step size required in taking an action a. The value of the goal state s_G was set to zero on each iteration. Value functions were computed for each goal location, and the relative value of states describes the relative minimum number of time steps required to reach s_G from each state. The lower the value of a state, the greater the geodesic separation between the state and the goal state. We set $\gamma = 1$ during all simulations and performed 100 iterations before calculating optimal trajectory lengths from an initial state s_i to the target state s_G, as this number of iterations allowed for the algorithm to converge.

Relevance computation. To compute the relevance $\Omega_k(s_0, s_G)$ of the kth transition to the task of navigating from a specific initial state s_0 to a specific goal s_G, we calculated the absolute change induced in the optimal value of the initial state after toggling the navigability of that transition by changing the corresponding element in the adjacency matrix from 1 to 0 or from 0 to 1 (Eq 4). For the simulations described below, we also tested a non-myopic, path-dependent metric $\Omega_k(s_0, s_G; \pi^*)$ defined as the sum of squared differences induced in the values of all states along the optimal path (Eq 5). Furthermore, we tested the robustness of the measure to the precise algorithm used to compute state values by computing value functions using the successor representation (SR) algorithm, which caches future state occupancy probabilities learned with a specific policy [27]. (While SR is more efficient than value iteration, it is less precise.) As we used a random walk policy, we computed the matrix of probabilities M analytically by temporally abstracting a one-step transition matrix T: $M = (I - \gamma T)^{-1}$. The cached probabilities can then be combined with a one-hot reward vector $R(s) = \mathbb{1}(s = s_G)$ to yield state values $V = MR$. We set the temporal discount factor $\gamma = 1$ and integrated over 100 time steps.

$$\Omega_k(s_0, s_G) = [V(s_0|T_k = 1) - V(s_0|T_k = 0)]^2 \tag{4}$$

$$\Omega_k(s_0, s_G; \pi^*) = \sum\nolimits_{s=s_0}^{s_G} [V(s|T_k = 1) - V(s|T_k = 0)]^2 \tag{5}$$

Relation to bottlenecks. In order to assess whether the relevance metric is predictive of the degree to which transitions are bottlenecks in the environment, we correlated normalized relevance values (averaged across all target locations and normalized via dividing by the maximum relevance value across all transitions for each target location) with the average betweenness centrality G of the two states on either side of a transition (Eq 6). Betweenness centrality essentially calculates the degree to which a state controls the traffic flowing through the arena. σ_{ij} represents the number of shortest paths between states i and j, and $\sigma_{ij}(s)$ represents the number of such paths which pass through state s. For this analysis, transitions within 1 m from the goal state were excluded due to their chance of having spuriously high relevance values.

$$G(s) = \sum_{i \neq s \neq j} \frac{\sigma_{ij}(s)}{\sigma_{ij}} \tag{6}$$

Simulations. Behavior of three artificial agents with qualitatively different planning capacities was simulated. All agents were initialized with a noisy model of the environment. Representational noise was simulated by toggling 50% of randomly selected unavailable transitions from $T(s, s') = 0$ to 1, and the equivalent number of randomly selected available transitions from $T(s, s') = 1$ to 0. This is analogous to the agents misplacing obstacles in their memories, or equivalently, a subjective-objective model mismatch induced by volatility in the environment. The blind agent was unable to correct its model during a planning period. On each trial, eight transitions (out of 210 available) were drawn for each sighted agent; the agent's model was compared with the true arena structure at these transitions and, if applicable, corrected prior to navigation. Visual samples were drawn uniformly from all possible transitions (without replacement) for the random exploration agent. For the goalward looking agent, the probability of drawing a transition was determined by a circular normal (von Mises) distribution with $\mu = \theta_G$ (where θ_G is the angle of the goal w.r.t the agent's heading), $\sigma = 1$, and concentration parameter $\kappa = 5$. In contrast, the directed sampling agent gathered information specifically about the eight transitions that were calculated to be most relevant for that trial. After the model updates, if any, the agents' subjective value functions were recomputed, and agents took actions according to the resulting policies. When an agent encountered a situation in which no action was subjectively available, they attempted a random action. In the case that a new action is discovered, the agents temporarily updated $T(s, s')$ from 0 to 1 for that action. Conversely, in the case that an agent attempted to take an action but discovered that it was not actually feasible, they temporarily updated their subjective models to account for the transition block which they had just learned about. In both cases, value functions were recomputed using the updated model. Simulations were conducted with 25 arenas of granularity n = 3 (state space size = 54 for computational tractability) and 100 trials per arena. Furthermore, we tested the agents' performance using a range of gaze samples evenly spaced between 2 and 14 foveations.

Data processing. In order to identify moving and non-moving epochs within each trial, movement onset and offset times were detected by applying a moving average filter of window size 5 frames on the absolute value of the joystick input function. When the smoothed joystick input exceeded the threshold of 0.2 m/s (approx. 10% of the maximum velocity), the participant was deemed to be moving, and when the input fell below this threshold for the last time on each trial, the participant was deemed to have stopped moving. Participants' relative planning time was defined as the ratio of pre-movement time to the total trial duration, minus the search period (which was roughly constant across arenas). Prior to any eye movement analyses, blinks were filtered from the eye movements by detecting when the fraction of the pupil visible dipped below 0.8. The spread in the (x, y) gaze positions within trials was calculated as the expectation of variance, $\mathbb{E}_n[\sqrt{\text{Var}_t[x] + \text{Var}_t[y]}\ ]$, where $\text{Var}_t[\cdot]$ denotes the variance across time t within a trial and $\mathbb{E}_n[\cdot]$ denotes expectation across trials denoted by n. The spread across trials was calculated as the variance of the expectation, $\sqrt{\text{Var}_n[\mathbb{E}_t[x]] + \text{Var}_n[\mathbb{E}_t[y]]}$, where $\mathbb{E}_t[\cdot]$ denotes expectation across time and $\text{Var}_n[\cdot]$ denotes the variance across trials.

For Figures 1f, 1g, S2a, S2b, and S2e, the first trial of each run was removed from the analyses due to an occasional rapid teleportation of the participant to a random starting location associated with the software starting up. While there were a few instances where more than one run occurred per block due to participants adjusting the headset, at least 51 trials were actually collected during each block such that most blocks consisted of 50 trials when the first trial of each run was omitted. For analyses such as epoch duration, gaze distribution, relevance, sweep detection, and subgoal detection, the first trial was not discarded since the teleportation only affected the recorded path length, but as the teleportation was virtually instantaneous, the new starting locations on such trials could be used for analyses which do not depend upon the path length variable.

Linear mixed effects models. To separately examine how various aspects of the navigation task contribute to the behavioral and eye movement patterns that we observed, we fit linear mixed effects models

of the form Eq 7, where each datapoint either corresponds to one trial (for trial-level analyses) or one participant in one arena (for analyses like % trials). In the former case, the number of predictor variables was $J = 4$, and a fixed effect slope β_j and participant-specific random effect slope β_{ij}, quantified the effect of each variable in a linear combination with a fixed intercept β_0 and a participant-specific random intercept β_{i0} to produce an estimate of the dependent variable y_i on that trial for participant i. The three variables were the number of turns (described in the "Subgoal analysis" section), the length of the optimal trajectory (described in the "RL formulation" section), and the unsigned angle between the direction of participants' initial heading vs. the optimal direction of target approach (relative bearing). The ranges of the predictors were – number of turns: 0-17, length of optimal trajectory: 0-45 m, relative bearing: 0-90 degrees. Eq 8 describes the specific form used to examine trial-specific effects on the output. For the case where analyses required pooling trials for each participant in each arena, $J = 1$ and the single set of fixed/random slopes correspond to the arena complexity variable. All variables were z-scored for each participant prior to model fittings, so the intercepts were close to zero for most cases and not shown in the bar plots.

$$y_i = \beta_0 + \beta_{i0} + \sum_{j=1}^{J} (\beta_j x_j + \beta_{ij} x_j) \tag{7}$$

$$\begin{gathered}\mathrm{Output} \sim (1|\mathrm{Participant}) + \mathrm{NTurns} + (\mathrm{NTurns\text{-}1}|\mathrm{Participant}) + \mathrm{PathLength} + (\mathrm{PathLength\text{-}1}|\mathrm{Participant}) + \\ \mathrm{Bearing} + (\mathrm{Bearing\text{-}1}|\mathrm{Participant})\end{gathered} \tag{8}$$

Relevance estimation. Prior to estimating the task-relevance of the participants' gaze positions at each time point, the closest transition k to the participant's point of gaze was identified and the effect of toggling the transition on the value function was computed as $\Omega_k(s(t), s_G)$. In order to construct a null distribution of relevance values, we paired the eye movements on each trial with the goal location for a random trial, given the participant's position in the current trial. This shuffled average is not task-specific, and therefore may be compared with the true Ω values to probe whether the spatial distribution of gaze positions was sensitive to the goal location on each trial. Similarly, the shuffled fraction of time looking at the goal was computed with a goal state randomly chosen from all states.

Sweep classification. Forward and backward eye movements (sweeps) along the intended trajectory were classified by first calculating the point (x, y) on the trajectory closest to the location of gaze in each frame. For each trial, the fraction of the total trajectory length corresponding to each point was stored as a variable f, and periods when $f(t)$ consecutively ascended or descended were identified. For each period, we determined m, an integer whose magnitude denoted the sequence length and whose sign denoted the sequence direction (+/- for ascending/descending sequences). We then constructed a null distribution $p(m)$ describing the chance-level frequency of m by selecting 20 random trials and recomputing f based on the participant's trajectories on those trials. Sequential eye movements of length m where the CDF of $p(m)$ was less than $\alpha/2$ or greater than $1 - \alpha/2$ were classified as backward and forward sweeps, respectively. The significance threshold α was chosen to be 0.02. Compensating for noise in the gaze position, we applied a median filter of length 20 frames to both the true and shuffled f functions. During post-processing, sweeps in the same direction that were separated by less than 25 frames were merged, and sweeps for which the gaze fell outside of 2 meters from the intended trajectory on >30% of the frames pertaining to the sweep were eliminated. Sweeps were required to be at least 25 frames in length. To remove periods of fixation, the minimum variance in $f(t)$ values for all time points corresponding to the sweep was required to be 0.001. Finally, sweeps which did not cover at least 20% of the total trajectory length were removed from the analyses. This algorithm allowed for the automated detection of sequential eye movements pertaining to the prospective evaluation of trajectories which participants subsequently took.

Alternative trajectories. To find the number of trajectory options for each trial, we identified all paths between the initial and goal states that were comparable within a factor of 1.25 to the optimal trajectory length and shared no more than 50% of the states with each other. The factor of 1.25 ensured that the trajectories were within 1 standard deviation of the trajectory chosen by the participants. The gaze was

classified to be exploring an alternative trajectory only when it was disjoint from the trajectory that the participant executed.

Saccade detection. Saccade times were classified to be the times at which eye movement speeds v crossed a threshold of 50°/s from below, where speeds were computed using Eq 9, where x, y, and z correspond to the coordinates of the point of gaze (averaged across both eyes), and α and β respectively correspond to the lateral and vertical displacement of the pupil in degrees.

$$\alpha(t) = \tan^{-1}\left(\frac{x(t)}{\sqrt{y^2(t) + z^2(t)}}\right),\ \beta(t) = \tan^{-1}\left(\frac{z(t)}{\sqrt{y^2(t) + x^2(t)}}\right),\ v(t) = \sqrt{\left(\frac{d}{dt}\alpha(t)\right)^2 + \left(\frac{d}{dt}\beta(t)\right)^2} \quad (9)$$

Subgoal analysis. Turns in the participants' trajectories (defined as subgoals) were isolated by applying a threshold of 60 deg/s on their angular velocity (smoothed with a median filter; window size = 8 frames). The first and last frames for periods of elevated angular velocity were recorded. For the purposes of the analysis in Figure 5, all trials (for all participants in all arenas) were broken down into periods of turns vs. periods of navigating straight segments. Starting from the stopping location, these periods were independently interpolated to fit an arbitrarily defined common timeline of 25 time points per turn and 100 time points per straight segment. For example, if a trial had two turns, then eye movement variables from the last turn to the stopping location were interpolated to the points -100 to -1. The last turn was represented by -125 to -101, and the second last turn was represented by -250 to -226. The segment between the two turns was represented by -225 to -126. Finally, the segment between the participant's starting position and the first turn was represented by -350 to -251. Note that as a consequence, the number of trials for which there were (for example) more than four turns in the trajectory was substantially fewer than the number of trials for which there were one or no turns, such that the quantity of raw data contributing to each normalized position value in Figure 5 increases from left to right.

Data and code availability. The dataset is available at https://gin.g-node.org/neuro-sci/gaze-navigation, and the MATLAB code used to produce the analysis figures has been published at https://github.com/neuro-sci/gaze-navigation.

Acknowledgements

We thank the members of the Angelaki Lab and Professor Wei Ji Ma for insightful discussions. This work was supported by the NIH (1U19-NS118246 – BRAIN Initiative, 1R01-EY022538), the NSF NeuroNex Award (DBI-1707398) and the Gatsby Charitable Foundation.

References

[1] Hunt LT, Daw ND, Kaanders P, MacIver MA, Mugan U, Procyk E, Redish AD, Russo E, Scholl J, Stachenfeld K, Wilson CR, Kolling N. Formalizing planning and information search in naturalistic decision-making. *Nature Neuroscience*, 24(8):1051–1064, 2021. ISSN 15461726.

[2] Mattar MG, Lengyel M. Planning in the brain. *Neuron*, 2022. ISSN 08966273.

[3] Momennejad I, Russek EM, Cheong JH, Botvinick MM, Daw ND, Gershman SJ. The successor representation in human reinforcement learning. *Nature Human Behaviour*, 1(9):680–692, 2017. ISSN 23973374.

[4] Miller KJ, Venditto SJC. Multi-step planning in the brain, 2021.

[5] Wunderlich K, Dayan P, Dolan RJ. Mapping value based planning and extensively trained choice in the human brain. *Nature Neuroscience*, 15(5):786–791, 2012. ISSN 10976256.

[6] Kaplan R, Friston KJ. Planning and navigation as active inference. *Biological Cybernetics*, 112(4):323–343, 2018. ISSN 14320770.

[7] Yang SCH, Lengyel M, Wolpert DM. Active sensing in the categorization of visual patterns. *eLife*, 5(FEBRUARY2016):1–22, 2016. ISSN 2050084X.

[8] Renninger LW, Verghese P, Coughlan J. Where to look next? Eye movements reduce local uncertainty. *Journal of Vision*, 7(3):1–17, 2007. ISSN 15347362.

[9] Najemnik J, Geisler WS. Optimal eye movement strategies in visual search. *Nature*, 434(7031):387–391, 2005. ISSN 00280836.

[10] Ma WJ, Navalpakkam V, Beck JM, Berg RVD, Pouget A. Behavior and neural basis of near-optimal visual search. *Nature Neuroscience*, 14(6):783–790, 2011. ISSN 10976256.

[11] Hoppe D, Rothkopf CA. Multi-step planning of eye movements in visual search. *Scientific Reports*, 9(1):1–12, 2019. ISSN 20452322.

[12] Leigh RJ, Kennard C. Using saccades as a research tool in the clinical neurosciences. *Brain*, 127(3):460–477, 2004. ISSN 00068950.

[13] Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P. Dynamics of Active Sensing and Perceptual Selection. *Curr Opin Neurobiol*, 20(2):172–176, 2010.

[14] Yang SCH, Wolpert DM, Lengyel M. Theoretical perspectives on active sensing. *Current Opinion in Behavioral Sciences*, 11:100–108, 2016. ISSN 23521546.

[15] Gottlieb J, Oudeyer PY. Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, 19(12):758–770, 2018. ISSN 14710048.

[16] Hutton SB. Cognitive control of saccadic eye movements. *Brain and Cognition*, 68(3):327–340, 2008. ISSN 02782626.

[17] Ryan JD, Shen K. The eyes are a window into memory. *Current Opinion in Behavioral Sciences*, 32:1–6, 2020. ISSN 23521546.

[18] Hoppe S, Loetscher T, Morey SA, Bulling A. Eye movements during everyday behavior predict personality traits. *Frontiers in Human Neuroscience*, 12(April):1–8, 2018. ISSN 16625161.

[19] Henderson JM, Shinkareva SV, Wang J, Luke SG, Olejarczyk J. Predicting Cognitive State from Eye Movements. *PLoS ONE*, 8(5):1–6, 2013. ISSN 19326203.

[20] Eckstein MK, Guerra-Carrillo B, Miller Singley AT, Bunge SA. Beyond eye gaze: What else can eye-tracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, 25:69–91, 2017. ISSN 18789307.

[21] Henderson JM, Hayes TR. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, 1(10):743–747, 2017. ISSN 23973374.

[22] Hayhoe M, Ballard D. Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4):188–194, 2005. ISSN 13646613.

[23] Schütt HH, Rothkegel LO, Trukenbrod HA, Engbert R, Wichmann FA. Disentangling bottom-up vs. top-down and low-level vs. high-level influences on eye movements over time. *arXiv*, 19:1–23, 2018.

[24] Kowler E. Eye movements: The past 25 years. *Vision Research*, 51(13):1457–1483, 2011. ISSN 00426989.

[25] Sutton RS, Barto AG. *Reinforcement Learning: An Introduction 2nd Ed*. 2018.

[26] Gustafson NJ, Daw ND. Grid cells, place cells, and geodesic generalization for spatial reinforcement learning. *PLoS Computational Biology*, 7(10), 2011. ISSN 1553734X.

[27] Stachenfeld KL, Botvinick MM, Gershman SJ. The hippocampus as a predictive map. *Nature Neuroscience*, 20(11):1643–1653, 2017. ISSN 15461726.

[28] Mattar MG, Daw ND. Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, 21(11):1609–1617, 2018. ISSN 15461726.

[29] Tomov MS, Yagati S, Kumar A, Yang W, Gershman SJ. *Discovery of hierarchical representations for efficient planning*, volume 16. 2020. ISBN 1111111111.

[30] Solway A, Diuk C, Córdova N, Yee D, Barto AG, Niv Y, Botvinick MM. Optimal Behavioral Hierarchy. *PLoS Computational Biology*, 10(8), 2014. ISSN 15537358.

[31] Liu Y, Dolan RJ, Kurth-Nelson Z, Behrens TE. Human Replay Spontaneously Reorganizes Experience. *Cell*, 178(3):640–652, 2019. ISSN 10974172.

[32] Ekman M, Kok P, De Lange FP. Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*, 8(May):1–9, 2017. ISSN 20411723.

[33] Kurth-Nelson Z, Economides M, Dolan RJ, Dayan P. Fast Sequences of Non-spatial State Representations in Humans. *Neuron*, 91(1):194–204, 2016. ISSN 10974199.

[34] Johnson A, Redish AD. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience*, 27(45):12176–12189, 2007. ISSN 02706474.

[35] Pfeiffer BE, Foster DJ. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447):74–79, 2013. ISSN 00280836.

[36] Dragoi G, Tonegawa S. Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature*, 469(7330):397–401, 2011. ISSN 00280836.

[37] Brown TI, Carr VA, LaRocque KF, Favila SE, Gordon AM, Bowles B, Bailenson JN, Wagner AD. Prospective representation of navigational goals in the human hippocampus. *Science*, 352(6291):1323–1326, 2016. ISSN 10959203.

[38] Lakshminarasimhan KJ, Avila E, Neyhart E, DeAngelis GC, Pitkow X, Angelaki DE. Tracking the Mind's Eye: Primate Gaze Behavior during Virtual Visuomotor Navigation Reflects Belief Dynamics. *Neuron*, 106(4):662–674, 2020. ISSN 10974199.

[39] Javadi AH, Emo B, Howard LR, Zisch FE, Yu Y, Knight R, Pinelo Silva J, Spiers HJ. Hippocampal and prefrontal processing of network topology to simulate the future. *Nature Communications*, 8, 2017.

[40] Hikosaka O, Nakamura K, Nakahara H. Basal ganglia orient eyes to reward. *Journal of Neurophysiology*, 95(2):567–584, 2006. ISSN 00223077.

[41] Koenig S, Kadel H, Uengoer M, Schubö A, Lachnit H. Reward draws the eye, uncertainty holds the eye: Associative learning modulates distractor interference in visual search. *Frontiers in Behavioral Neuroscience*, 11(July):1–15, 2017. ISSN 16625153.

[42] Postle BR, Sala SD, Baddeley AD. The selective disruption of spatial working memory by eye movements. *Q J Exp Psychol*, 59(1):100–120, 2006.

[43] Buhry L, Azizi AH, Cheng S. Reactivation, replay, and preplay: How it might all fit together. *Neural Plasticity*, 2011. ISSN 16875443.

[44] Caspi A, Beutter BR, Eckstein MP. The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 101(35):13086–13090, 2004. ISSN 00278424.

[45] Crowe DA, Averbeck BB, Chafee MV. Mental Maze Solving. *Journal of Cognitive Neuroscience*, 12(5):813–827, 2000.

[46] Redish AD. Vicarious trial and error, 2016.

[47] Gottlieb J, Oudeyer PY, Lopes M, Baranes A. Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*, 17(11):585–593, 2013. ISSN 13646613.

[48] Gottlieb J, Hayhoe M, Hikosaka O, Rangel A. Attention, reward, and information seeking. *Journal of Neuroscience*, 34(46):15497–15504, 2014. ISSN 15292401.

[49] Ahmad S, Yu AJ. Active sensing as bayes-optimal sequential decision-making. *Uncertainty in Artificial Intelligence - Proceedings of the 29th Conference, UAI 2013*, pages 12–21, 2013.

[50] Smittenaar P, FitzGerald TH, Romei V, Wright ND, Dolan RJ. Disruption of Dorsolateral Prefrontal Cortex Decreases Model-Based in Favor of Model-free Control in Humans. *Neuron*, 80(4):914–919, 2013. ISSN 08966273.

[51] Simon DA, Daw ND. Neural correlates of forward planning in a spatial decision task in humans. *Journal of Neuroscience*, 31(14):5526–5539, 2011. ISSN 02706474.

[52] De Cothi W, Nyberg N, Griesbauer EM, Ghanamé C, Zisch F, Lefort J, Fletcher L, Newton C, Renaudineau S, Bendor D, Grieves R, Duvelle Barry C, Spiers HJ. Predictive Maps in Rats and Humans for Spatial Navigation. *bioRxiv*, page 2020.09.26.314815, 2020.

[53] Ghamari H, Golshany N. Wandering Eyes: Using Gaze-Tracking Method to Capture Eye Fixations in Unfamiliar Healthcare Environments. *Health Environments Research and Design Journal*, 2021. ISSN 19375867.

[54] Pierrot-Deseilligny C, Rivaud S, Gaymard B, Muri R, Vermersch AI. Cortical control of saccades. *Ann Neurol*, 37:557–567, 1995.

[55] Johnston K, Everling S. Monkey dorsolateral prefrontal cortex sends task-selective signals directly to the superior colliculus. *Journal of Neuroscience*, 26(48):12471–12478, 2006. ISSN 02706474.

[56] Pierrot-Deseilligny C, Müri RM, Nyffeler T, Milea D. The role of the human dorsolateral prefrontal cortex in ocular motor behavior. *Annals of the New York Academy of Sciences*, 1039:239–251, 2005. ISSN 00778923.

[57] Tervo DGR, Kuleshova E, Manakov M, Proskurin M, Karlsson M, Lustig A, Behnam R, Karpova AY. The anterior cingulate cortex directs exploration of alternative strategies. *Neuron*, 109(11):1876–1887, 2021. ISSN 10974199.

[58] Gaymard B, Rivaud S, Cassarini JF, Dubard T, Rancurel G, Agid Y, Pierrot-Deseilligny C. Effects of anterior cingulate cortex lesions on ocular saccades in humans. *Experimental Brain Research*, 120(2):173–183, 1998. ISSN 00144819.

[59] Frey M, Nau M, Doeller CF. Magnetic resonance-based eye tracking using deep neural networks. *Nature Neuroscience*, 24(12):1772–1779, 2021. ISSN 15461726.

[60] Diba K, Buzsaki G. Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience*, 10(10):1241–1242, 2007.

[61] Zhou R, Hansen EA. Combining breadth-first and depth-first strategies in searching for treewidth. *International Symposium on Combinatorial Search, SoCS 2008*, pages 162–168, 2008.

[62] Sosa M, Giocomo LM. Navigating for reward, 2021.

[63] Gillespie AK, Maya DAA, Denovellis EL, Liu DF, Kastner DB, Coulter ME, Roumis DK, Eden UT, Frank LM. Hippocampal replay reflects specific past experiences rather than a plan for subsequent choice. *BioRXiv*, 2021.

[64] Liu Y, Mattar MG, Behrens TE, Daw ND, Dolan RJ. Experience replay is associated with efficient nonlocal learning. *Science*, 372(6544), 2021. ISSN 10959203.

[65] Eldar E, Lièvre G, Dayan P, Dolan RJ. The roles of online and offline replay in planning. *eLife*, 9:1–23, 2020. ISSN 2050084X.

[66] Larson AM, Loschky LC. The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, 9(10):1–16, 2009. ISSN 15347362.

[67] Wilming N, König P, König S, Buffalo EA. Entorhinal cortex receptive fields are modulated by spatial attention, even without movement. *bioRxiv*, pages 1–16, 2017.

[68] Hannula D, Ranganath C. The Eyes Have It: Hippocampal Activity Predicts Expression of Memory in Eye Movements. *Neuron*, 63(5):592–599, 2009.

[69] Crowe DA, Chafee MV, Averbeck BB, Georgopoulos AP. Neural Activity in Primate Parietal Area 7a Related to Spatial Analysis of Visual Mazes. *Cerebral Cortex*, 14(1):23–34, 2004. ISSN 10473211.

[70] Turk-Browne NB. The hippocampus as a visual area organized by space and time: A spatiotemporal similarity hypothesis. *Vision Research*, 165(October):123–130, 2019. ISSN 18785646.

[71] Liu ZX, Shen K, Olsen RK, Ryan JD. Visual sampling predicts hippocampal activity. *Journal of Neuroscience*, 37(3):599–609, 2017. ISSN 15292401.

[72] Monaco J, Rao G, Roth E, Knierim J. Attentive Scanning Behavior Drives One-Trial Potentiation of Hippocampal Place Fields. *Nature Neuroscience*, 17(5):725–731, 2014.

[73] Jun JJ, Longtin A, Maler L. Active sensing associated with spatial learning reveals memory-based attention in an electric fish. *Journal of Neurophysiology*, 115(5):2577–2592, 2016. ISSN 15221598.

[74] Fotowat H, Lee C, Jun JJ, Maler L. Neural activity in a hippocampus-like region of the teleost pallium are associated with navigation and active sensing. *eLife*, 8:e44119, 2019.

[75] Ringo JL, Sobotka S, Diltz MD, Bunce CM. Eye movements modulate activity in hippocampal, parahippocampal, and inferotemporal neurons. *Journal of Neurophysiology*, 71(3):1285–1288, 1994. ISSN 00223077.

[76] Bakkour A, Palombo DJ, Zylberberg A, Kang YH, Reid A, Verfaellie M, Shadlen MN, Shohamy D. The hippocampus supports deliberation during value-based decisions. *eLife*, 8(e46080), 2019.

[77] Tajima S, Drugowitsch J, Pouget A. Optimal policy for value-based decision-making. *Nature Communications*, 7, 2016. ISSN 20411723.

[78] Killian NJ, Jutras MJ, Buffalo EA. A map of visual space in the primate entorhinal cortex. *Nature*, 491(7426):761–764, 2012. ISSN 00280836.

[79] Meister ML, Buffalo EA. Neurons in primate entorhinal cortex represent gaze position in multiple spatial reference frames. *Journal of Neuroscience*, 38(10):2430–2441, 2018. ISSN 15292401.

[80] Killian NJ, Buffalo EA. Grid cells map the visual world. *Nature Neuroscience*, 21(2):161–162, 2018. ISSN 15461726.

[81] Eckstein MK, Collins AG. Computational evidence for hierarchically structured reinforcement learning in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 117(47):29381–29389, 2020. ISSN 10916490.

[82] Balaguer J, Spiers H, Hassabis D, Summerfield C. Neural Mechanisms of Hierarchical Planning in a Virtual Subway Network. *Neuron*, 90(4):893–903, 2016. ISSN 10974199.

[83] Rasmussen D, Voelker A, Eliasmith C. *A neural model of hierarchical reinforcement learning*, volume 12. 2017. ISBN 1111111111.

[84] Sipatchin A, Wahl S, Rifai K. Eye-tracking for low vision with virtual reality (VR): testing status quo usability of the HTC Vive Pro Eye 2. *bioRxiv*, (1898):2020.07.29.220889, 2020.

[85] Bermudez-Contreras E, Clark BJ, Wilber A. The Neuroscience of Spatial Navigation and the Relationship to Artificial Intelligence. *Frontiers in Computational Neuroscience*, 14(July):1–16, 2020. ISSN 16625188.

[86] Bellman R. Some Problems in the Theory of Dynamic Programming. *Econometrica*, 22(1):37, 1954. ISSN 00129682.