Eye Movements Reveal Spatiotemporal Dynamics of Active Sensing and Planning in Navigation

Seren Zhu[1,5], Kaushik Lakshminarasimhan[2,5], Nastaran Arfaei[3], and Dora 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

April 26, 2021

Abstract

Goal-oriented navigation is widely understood to depend critically upon internal maps. Although this may be the case in many settings, many animals, such as humans, tend to rely on vision when the environment is complex or unfamiliar. The nature of gaze during visually-guided navigation remains unknown. To study this, we tasked humans to navigate to transiently visible goals in virtual mazes of varying complexity, observing that they achieved near-optimal path lengths in all arenas. By analyzing subjects' eye movements, we gained insights into the principles of active sensing and prospection that would not have been gleaned from observing navigational behavior alone. The spatial distribution of fixations 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. Furthermore, the temporal evolution of gaze revealed rapid, sequential prospection of the future path, evocative of neural 'preplay'. These findings suggest that the spatiotemporal characteristics of eye movements during navigation are significantly shaped by the unique cognitive computations underlying real-world, sequential decision making.

Introduction

"The eye, the window of the soul, is the chief means whereby the [central sense] can most fully and abundantly appreciate the infinite works of Nature." – Leonardo da Vinci ca. 1500

By swiftly parsing a large, complex scene on a millisecond timescale, the eyes actively interrogate and efficiently gather information to facilitate complex computations [1, 2]. At the same time, eye movements reveal the contents of internal deliberation and the prioritization of goals in real-time [3–6]. Thus, eye tracking lends itself as a valuable tool for investigating both sensing and cognition [7–9]. However, gaze dynamics is seldom studied in the context of sequential decision making tasks such as goal-directed navigation.

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 [10–12]. During routine activities such as making tea, we tend to foveate specifically on objects relevant to the task being performed (e.g. boiling water), while ignoring salient distractors [11, 13]. The strategy of orienting the sensory apparatus to reduce uncertainty about task variables in the service of decision making is known as *active sensing* [3]. Computational modeling studies on active sensing have shown that human eye movements during visual search [14–16], categorization [17], pattern matching [18], and instrumental sampling [19–21] are near-optimal with respect to the task structure. In contrast to the search and binary decision making tasks listed above (e.g. deciding whether a pattern belongs to a cheetah or a zebra), navigation entails deciding upon and carrying out a *sequence* of actions to achieve a goal, thereby introducing unique cognitive demands that must be addressed during the study of gathering information. Specifically, knowledge about the structure of an environment —– whether transiently encoded in working memory [22], or solidified in the form of a cognitive map [23–25] —– is a crucial ingredient that allows navigators to efficiently plan favorable trajectories to their goals [26]. However, neural representations of the environment are noisy and rarely accurate. Furthermore, environmental volatility would render learned representations unreliable [27–29]. The precise advantage that active visual sampling confers upon navigational planning in structured environments, and the extent to which humans exploit it, are unclear.

A candidate framework to formalize this hypothesis and extend current insights on the role of active sensing to the more naturalistic domain of navigation is Reinforcement Learning (RL), whereby the goal of behavior is cast in terms of maximizing total long-term reward [30]. 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 [31, 32]. Incidentally, RL provides a formal interpretation of active sensing, which can be understood as exploratory actions taken with the purpose of improving knowledge about the environment, allowing for better planning and ultimately greater long-term reward [3]. Here, we address this gap in the role of active sensing in navigation by using the RL framework to mathematically predict the precise spatial locations toward which eye movements should be directed, given the objective of navigating from a given starting position to a goal. We hypothesized that eye movements during navigation would reduce uncertainty about the model of relationships between spatial positions in the world, allowing subjects to learn the transitions available to them. Our results illustrate that artificial agents who sample information in accordance with these predictions outperform agents who use undirected strategies or simple heuristics, successfully overcoming model uncertainty to achieve near-optimal path lengths. To test this observation experimentally, we designed a virtual reality navigation task in which human subjects navigated to transiently visible targets using a joystick in unfamiliar arenas of various degrees of complexity. We found that human subjects balanced foveating the hidden reward location with viewing highly task-consequential transitions both prior to and during active navigation.

In addition to ascertaining the model via sensing, planning — the process of simulating future steps prior to taking action [33] — is a crucial component of model-based sequential decision making tasks like goal-oriented navigation. In rodents, a neural signature of such an ordered list of steps during navigation occurs in the form of the sequential activation of hippocampal neurons along trajectories prior to movement, a phenomenon termed 'preplay' [34–38]. Inspired by this well-documented phenomenon, we sought to utilize gaze to better understand mechanisms of human navigational prospection, hypothesizing that subjects' eye

movements would faithfully manifest their trajectory plans. Our experiment revealed that subjects' gaze indeed swept along the trajectory which they subsequently embarked upon. Furthermore, subjects seemed to decompose convoluted trajectories by focusing on one turn at a time until they reached their goal. Taken together, our results suggest that during naturalistic navigation, the spatiotemporal dynamics of gaze are significantly shaped by active sensing and planning, thus reflecting the unique cognitive demands of real-world sequential decision making.

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 subjects 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. Subjects 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 — left). 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.

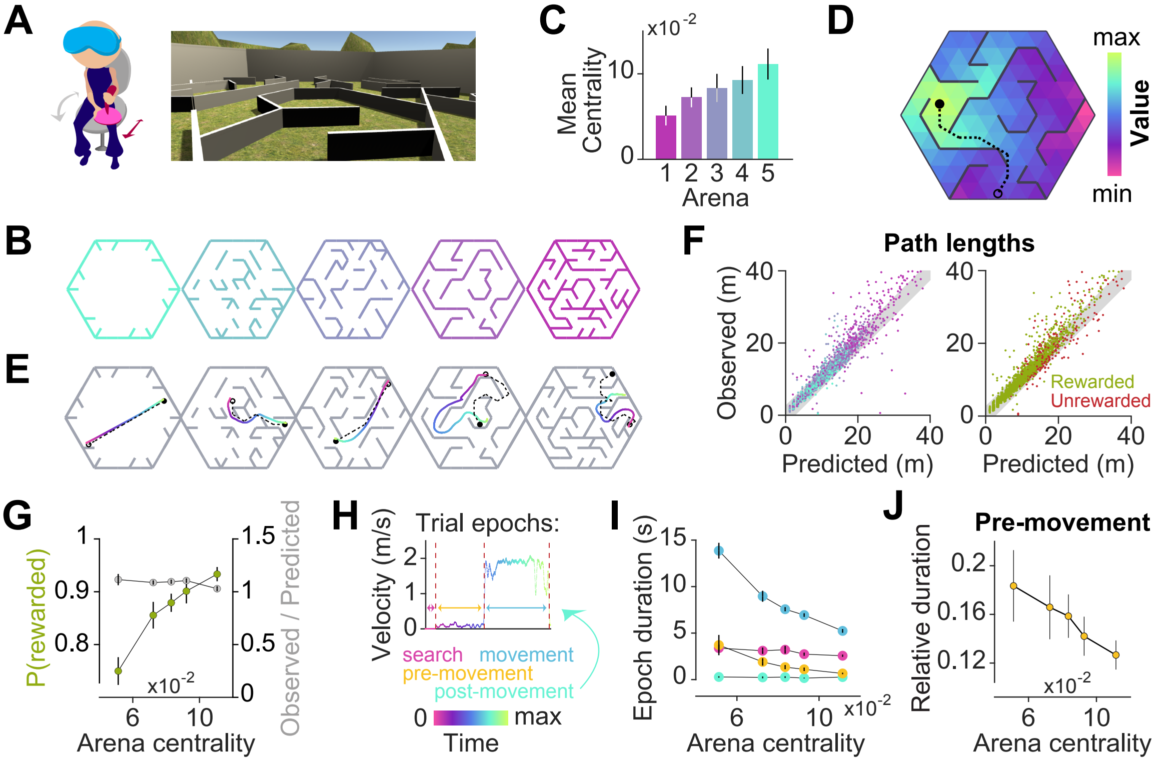

Figure 1: **Subjects exhibit near-optimal navigation performance across multiple environments. A.** Left: Human subjects wore a VR headset and executed turns by rotating in a swivel chair, while translating forwards or backwards using an analog joystick. Right: A screenshot of the first-person view of 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 mean state closeness centrality, with a lower centrality value corresponding to a more complex arena. Error bars denote ±1 SE across states. **D.** Heatmap showing the value function corresponding to an arbitrary goal state (closed circle) in one of the arenas. For each goal location, the value of each state directly corresponded 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 human subject. 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 (two states wide; see Figure S1a). Top: Data points are colored in accordance to the colors of each arena as depicted in **B**. Bottom: Unrewarded trials (red) vs. rewarded trials (green) had similar path lengths. For both plots, all trials for all subjects and all arenas are superimposed. **G.** The mean closeness centrality metric predicts the fraction of rewarded trials in each arena. However, 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 (banana) 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**) decreased with arena centrality. **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. Subjects could take ***actions*** to achieve ***transitions*** between adjacent states which were not separated by obstacles. As subjects were free to rotate and/or translate, the space of possible actions was continuous such that subjects did not report knowledge about the tessellation. Furthermore, subjects experienced a relatively high vantage point and were able to gaze over the tops of all of the obstacles (Figure 1a — right). On each trial, subjects 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 subjects had to locate and foveate in order to unlock the joystick. The banana disappeared 200 ms after foveation, and 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 subjects that they had received either two points for stopping within the goal state, one point for stopping in a state neighboring the goal state, or zero points for stopping in any other state. While subjects viewed the feedback, a new goal for the next trial was spawned without breaking the continuity of the task. In separate blocks, subjects 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; Figure 1c) [39]. Lower centralities correspond to more complex arenas. One of the blocks involved an open arena that contained only a few obstacles at the perimeter, such that on most trials, subjects 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).

To quantify behavioral performance, we first computed the optimal trajectory for each trial using dynamic programming (Figure S1b). This technique uses two pieces of information — the goal location (***reward function***) and the obstacle configuration (***transition structure***) — to calculate 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 (Figure 1d). The ***optimal policy*** requires that subjects 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 subjects' position in each trial. Subjects took near-optimal paths (i.e. optimal to within the width of the reward zone) on most trials (Figure 1f), scoring 74±5% of the points across all arenas and stopping within the reward zone on 86±4% of all trials (Figure S1c). However, participants occasionally took a suboptimal route (Figure 1e — second from right), or sometimes misremembered the goal location and took a near-optimal trajectory to a different location (Figure 1e — rightmost). The latter type constituted a majority (68±13%) of the unrewarded trials — most of the unrewarded trajectories would have been near-optimal if the subjects' stopping location was deemed to be the goal state (Figure S1d — left), which suggests that remembering the goal location was not straightforward. We quantified the degree of optimality by computing the ratio of observed vs. optimal path lengths to the subjects' stopping location. Across all trials (rewarded and unrewarded), this ratio was close to unity (median = 1.06, interquartile range = 0.14), suggesting that subjects were able to navigate efficiently in all arenas (Figure 1g, gray). This was the case even in unrewarded trials, where the observed path lengths were within 3% of the optimal path length to the subjects' stopping location (Figure S1d — right). Navigational performance was near-optimal from the beginning, such that there was no visible improvement with experience (Figure S1e). However, the fraction of rewarded trials decreased with increasing arena complexity (r = 0.70, p = 9.3 x10^{-8}), suggesting that the ability to remember the goal location is compromised in challenging environments (Figures 1g, light green).

In order to understand how subjects tackled the computational demands of the task, it is critical to break down each trial into three main epochs: ***search*** — when subjects sought to locate the goal, ***pre-movement*** — when subjects surveyed their route prior to utilizing the joystick, and ***movement*** — when subjects 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.

We compared the fraction of time spent in different epochs. Although subjects spent a major portion of each trial navigating to the target, the relative duration of other epochs was not negligible (mean fraction ± SD — search: 0.26±0.03, pre-movement: 0.11±0.04, movement: 0.63±0.05; Figure S1f). There was considerable variability across subjects in the fraction of time spent in the pre-movement phase (coefficient of variation (CV) — search: 0.11, pre-movement: 0.38, movement: 0.07), although this did not translate to a significant difference in navigational precision (Figure S1g). One possible explanation is that some participants were simply more efficient planners or were more skilled at planning on the move. Because we are interested in principles that are conserved across subjects, we pooled subjects for all subsequent analyses. 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. Notably, the pre-movement duration was also higher in more complex arenas, reflecting the subject's commitment to meet the increased planning demands in those arenas (Figures 1j, S1h). Nonetheless, the relative pre-movement duration did not correlate with the probability of reward, and was similar for rewarded and unrewarded trials (Figures S1h, S1h). This suggests that the participants' performance is limited by their success in remembering the reward location, rather than in meeting planning demands. 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 relatively complex, unfamiliar environments.

A computational analysis supports that human eye movements are task relevant

Aiming to gain insights from subjects' 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 ± SD — search: 6.3±2.4 m, pre-movement: 1.9±0.6 m, movement: 3.0±0.6 m; Figure 2b — left). This pattern was reversed when examining the spatial spread across trials (mean spread ± SD — search: 5.4±2.4 m, pre-movement: 6.7±1.2 m, movement: 6.5±0.6 m; Figure 2b — right). This suggests that subjects' eye movements during pre-movement and movement were chiefly dictated by trial-to-trial fluctuations in task demands.

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 [42, 43]. Consistent with this, subjects spent a large fraction of time looking at the reward zone (pre-movement: 67±3%, movement: 53±6%). However, this

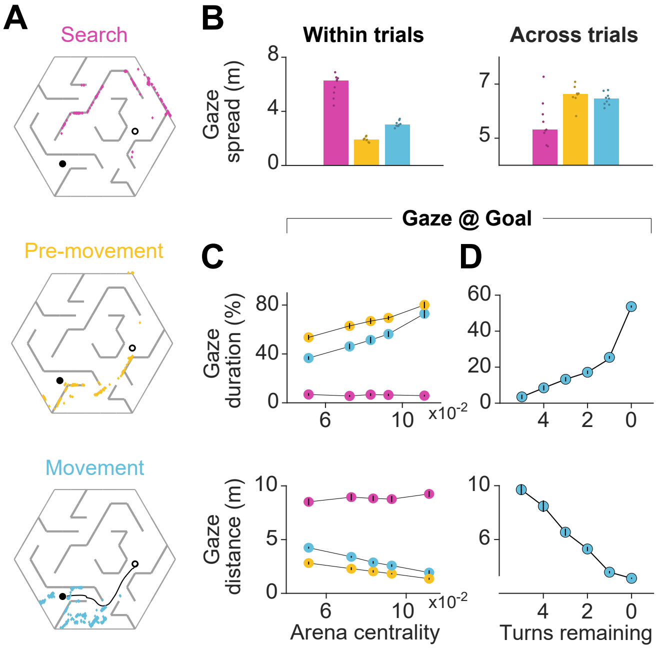

Figure 2: **Eye movements are modulated by goal location and environment complexity. A.** Eye positions on a representative trial for one subject during the three main trial epochs. Each datapoint corresponds to one frame. An open black circle denotes the start location, while a closed black circle denotes the goal. **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 subject data are overlaid on top of the bars. **C.** Top: The fraction of time for which subjects' gaze was within 2 m of the center of the goal state for each arena. Bottom: Average distance between the gaze position and the goal state for each arena. Epochs are colored in accordance to the color scheme in **A**. **D.** Top: Across all subjects and all trials, the greater the number of turns remaining in the subjects' trajectory, the lower the fraction of time that subjects looked within 2 m of the goal location. Bottom: The greater the number of turns remaining, the greater was the average distance of the point of gaze from the goal location. All error bars denote ±1 SEM.

fraction decreased with arena complexity (Figures 2c — top) resulting in a larger mean distance between the gaze and the goal in more complex arenas (Figure 2c — bottom). This effect could not be attributed to subjects forgetting the goal location in more complex arenas, as we found a similar trend when analyzing gaze in relation to the eventual stopping location (believed goal location; Figure S2). A more plausible explanation is that constantly looking at the goal may not enable subjects to efficiently learn the task-relevant transition structure of the environment, as the transition 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 subjects approach the target. Indeed, subjects 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 (Figure 2d). 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 goal allows for the study of the extent to which eye movements are dedicated to encoding the reward function, how may we assess the effectiveness with which subjects interrogate the transition structure of the environment to solve the task of navigating from point A to point B? This is the paramount challenge that we address in the remainder of this section by constructing a novel theoretical measure to guide data analysis.

In the case that a subject has a precise model of the transition structure of the environment, they would be theoretically capable of planning trajectories to the remembered goal location without vision. However, in this experiment, the arena configurations were unfamiliar to the subjects, such that they would be quite uncertain about the transition structure. The finding that subjects achieved near-optimal performance on even the first few trials in each arena (Figure S1e) indicates that humans are capable of using vision to rapidly reduce their uncertainty about the aspects of the model needed to solve the task. Motivating the quantitative characterization of the task relevance of visual samples, we show an illustration of the consequence of mistaken beliefs about the passability of specific transitions on the subjective value function. Transition ***toggling*** can be defined as the act of removing an obstacle between two states if an obstacle was previously present, or adding an obstacle at that location if one was previously absent. In alignment with intuition, some transitions are more important to veridically represent (Figure 3a — middle), as toggling them results in a dramatic change to the value function (which is essential to computing the optimal set of actions to reach the goal), while toggling some other transitions causes a relatively minimal change (Figure 3a — right).

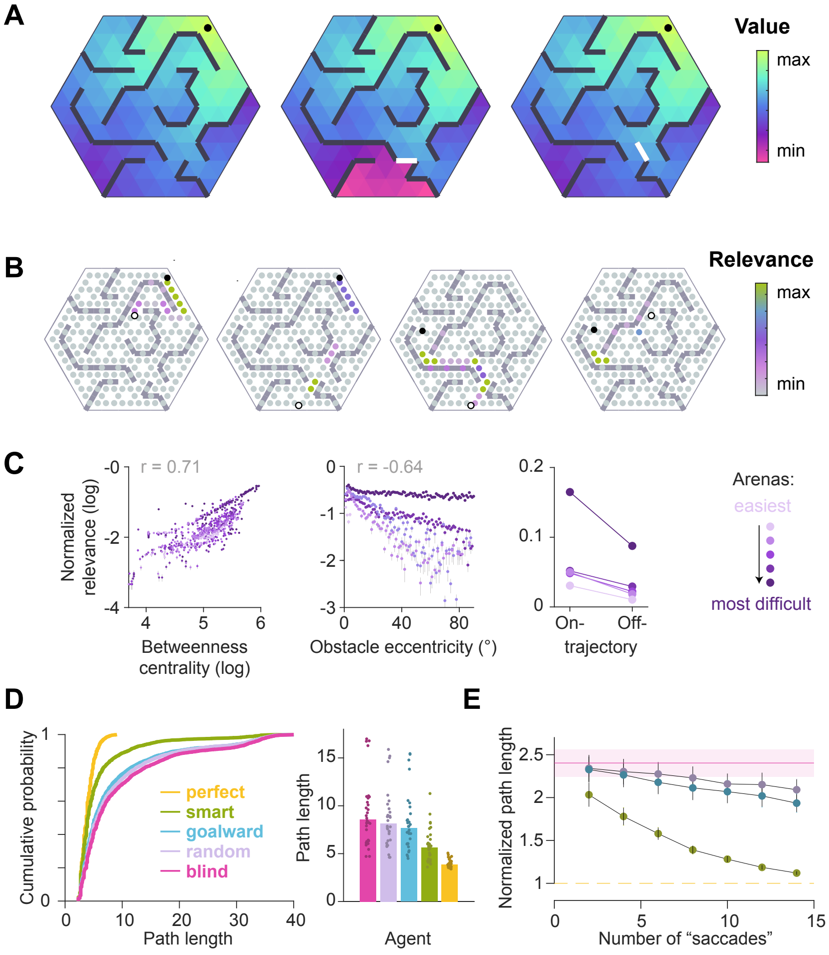

Figure 3: **Simulations validate the utility of precisely knowing the status of theoretically important transitions. A.** Value functions corresponding to an arbitrary goal location (closed circle) in an example arena (left) and in the arenas resulting from blocking either a bottleneck transition (center) or a transition that was not a bottleneck (right). **B.** Theoretical relevance of all transitions (circles) for the example arena for four different pairs of start (open black circle) and goal (closed black circle) states. **C.** The mean normalized relevance of non-obstacle transitions was positively correlated with betweenness centrality values (taken to be the average of the two states on either side of each transition). For this analysis, transitions within 1 m from the goal state were excluded due to their chance of having spuriously high relevance values. Middle: The mean normalized relevance of obstacles was negatively correlated with the eccentricity of each obstacle from the straight line connecting the current state to the goal state. Right: Transitions that fell on the optimal trajectory had greater relevance than those that fell outside of it. **D.** Simulation results of an agent instantiated with a perfect transition model (orange), and four agents with imperfect transition models, three of whom were endowed with the ability to correct their model according to different rules (see text). Those three agents were allowed to make eight 'saccades', each of which could update one transition. Left: Cumulative distributions (CDFs) of the path lengths of various agents (100 trials each from 25 different arenas; see Methods). Right: Median of trial-averaged path lengths across all simulated arenas; data points denote trial-averaged path lengths in individual arenas. **E.** Results of simulations similar to **D** but with a variable budget of 'saccades'. Each line denotes the average path length (across arenas) of one agent as a function of the number of 'saccades'. For each trial, path lengths of different agents were normalized by the optimal path length before trial-averaging. Error bars denote ±1 SEM.

We leveraged this insight and defined a novel metric to quantify the task-***relevance*** of each transition by computing the magnitude of the change in value of the subject's current state, for a given goal location, if the status of the transition was toggled:

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

where $\Omega_k(s_0, s_G)$ denotes the relevance of the k^{th} 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 k^{th} transition, provided that the transitions are stationary and the subject'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 utility of staying put. In the supplementary notes, we derive 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 subject is more certain about some transitions than others).

We found that the transitions with the highest relevance on each trial strikingly correspond to bottleneck transitions that bridge clusters of interconnected states (Figures 3b, 3c — left). Betweenness centrality is a metric describing the degree to which states are connected to other states (see Methods). The mean relevance of a transition strongly correlates with the betweenness centrality averaged across two states on either side of the transition. At the same time, the relevance was also high for obstacles that precluded a straight path to the goal (Figure 3c — center), as well as for transitions along the optimal trajectory (3c — right). By defining relevance of transitions according to Equation 1, we can thus capture multiple task-relevant attributes in a succinct manner. Theoretically investigating whether looking at task-relevant transitions improves navigational efficiency, we simulated artificial agents performing the same task that we imposed upon our human subjects. One agent ("perfect") had a veridical subjective model of the environment, and thus was capable of computing the optimal trajectory (Figure S4). Its antithesis ("blind") had an incorrect subjective model where half of the obstacle positions were 'misremembered' (toggled), but this agent was incapable of using vision to correct their model prior to taking actions according to their subjectively computed value functions. Performance at these two extremes was compared against the performance of three agents that were allocated a fixed budget of 'saccades' to rectify their incorrect models. These agents either randomly interrogated transitions ("random"), preferentially sampled transitions along the direction connecting the agent's starting location to the goal location ("goalward"), or chose the most task-relevant transitions as defined by the relevance metric in Eq 1 ("smart").

While all three agents showed an improvement over the "blind" agent, the agent with knowledge about the most task-relevant transitions resulted in much shorter average path lengths than agents looking at transitions along the general direction of the goal or looking at random transitions (mean path length ± SE

— perfect: 3.9±0.1, blind: 9.6±0.7, random: 8.9±0.6, goalward: 8.6±0.7, smart: 5.8±0.3; Figure 3d). Moreover, the performance of the smart sampling agent quickly approached optimality as the number of sampled transitions increased (Figure 3e). The rate of performance improvement was substantially slower for the goalward and random samplers (linear rather than exponential). These results were robust to the precise algorithm used to compute the value function in Eq 1. In particular, the successor representation (SR) has been proposed as a computationally efficient, biologically plausible alternative to pure model-based algorithms like value iteration for responding to changing goal locations [27, 32]. We found that estimating the task-relevance of transitions using values implied by SR resulted in a similar performance improvement (Figure S5). Nevertheless, we emphasize that our objective was to use the relevance metric simply as a means to probe whether humans preferentially looked at task-relevant transitions. Understanding how the brain might compute such metrics is outside the scope of this study.

To apply the above theoretical insights to the assessment of subjects' eye movements, we quantified the usefulness of subjects' 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 for each trial. This resulted in frame-by-frame gaze relevance values that ranged between zero (least relevant) and one (most relevant). Then, we constructed a distribution of shuffled relevances by analyzing gaze with respect to a random goal location. Figure 4a shows the resulting cumulative distributions across trials for the average subject during the three epochs in an example arena. As expected, the relevance of subjects' gaze was not significantly different from chance during the search epoch, as the subject had not yet determined the goal location. However, relevance values were significantly greater than chance both during pre-movement and movement (median relevance {and interquartile range} for the most complex arena, pre-movement — true: 0.21{0.06}, shuffled: 0.06{0.04}; movement — true: 0.21{0.07}, shuffled: 0.11{0.05}). Results were qualitatively similar in other arenas (Figure S6, Table S2).

To concisely describe subjects' 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 4a — 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 (Figures 4b; mean AUC ± SD — search: 0.52±0.03, pre-movement: 0.76±0.04, movement: 0.71±0.06). This suggests that subjects 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 3b — left), it is natural to wonder whether our evaluation of gaze relevance was confounded by the observation that subjects 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 4a-b) by computing the area under the ROC curves (AUC) constructed using the true vs. shuffled distributions of the fraction of the duration spent foveating the goal in each epoch (Figure 4c-d). Across all arenas, AUCs were high during the pre-movement and movement epochs, confirming that there was a strong tendency for subjects to look at the goal location (Figure 4e). When we excluded gaze positions that fell within the reward zone while computing relevance, we found that the degree to which subjects 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 4e; Pearson's r — pre-movement: -0.95, movement: -0.87). In contrast, the tendency to look at the goal location followed the opposite trend: it was greater in easier arenas (Figure 4f; Pearson's r — pre-movement: 0.95, movement: 0.88). 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 this task.

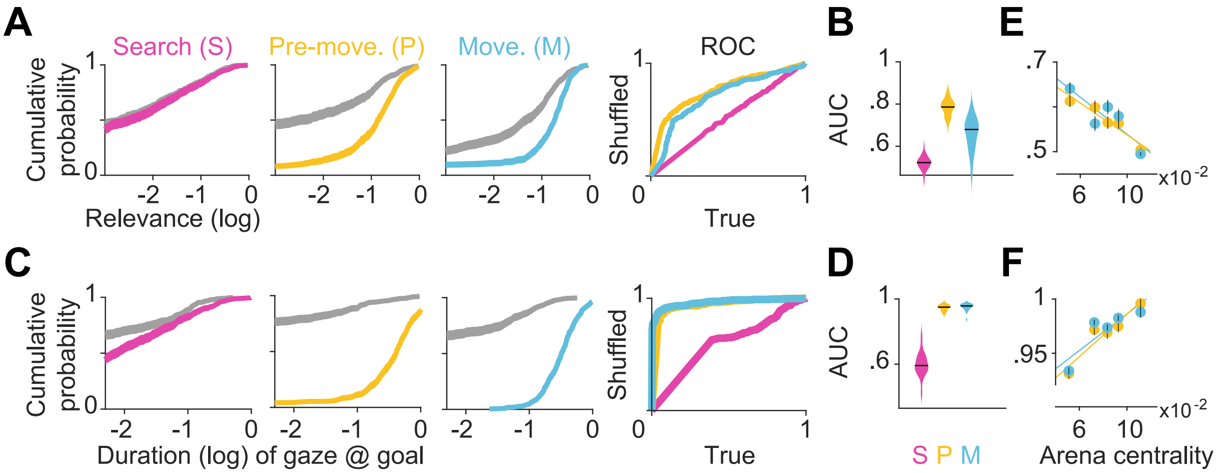

Figure 4: **Eye movements reveal a cognitive trade-off between reward and transition encoding.** **A.** Left: Cumulative distribution (computed by pooling trials from all subjects) of average log (normalized) relevance values (pink) and the corresponding shuffled distribution (gray) during search (left), 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 within two meters of the eventual stopping position (which was assumed to be the subjects' 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. **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 5a (top) 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. 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, subjects seemed to rapidly navigate their future paths with their eyes, and all subjects exhibited sweeping eye movements without being explicitly instructed to plan their trajectories prior to navigating. The fraction of time that subjects looked near the trajectories which they subsequently embarked upon increased with arena difficulty (Figure S7a). 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 geodesic 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 automatically 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 5a — bottom; see Methods).

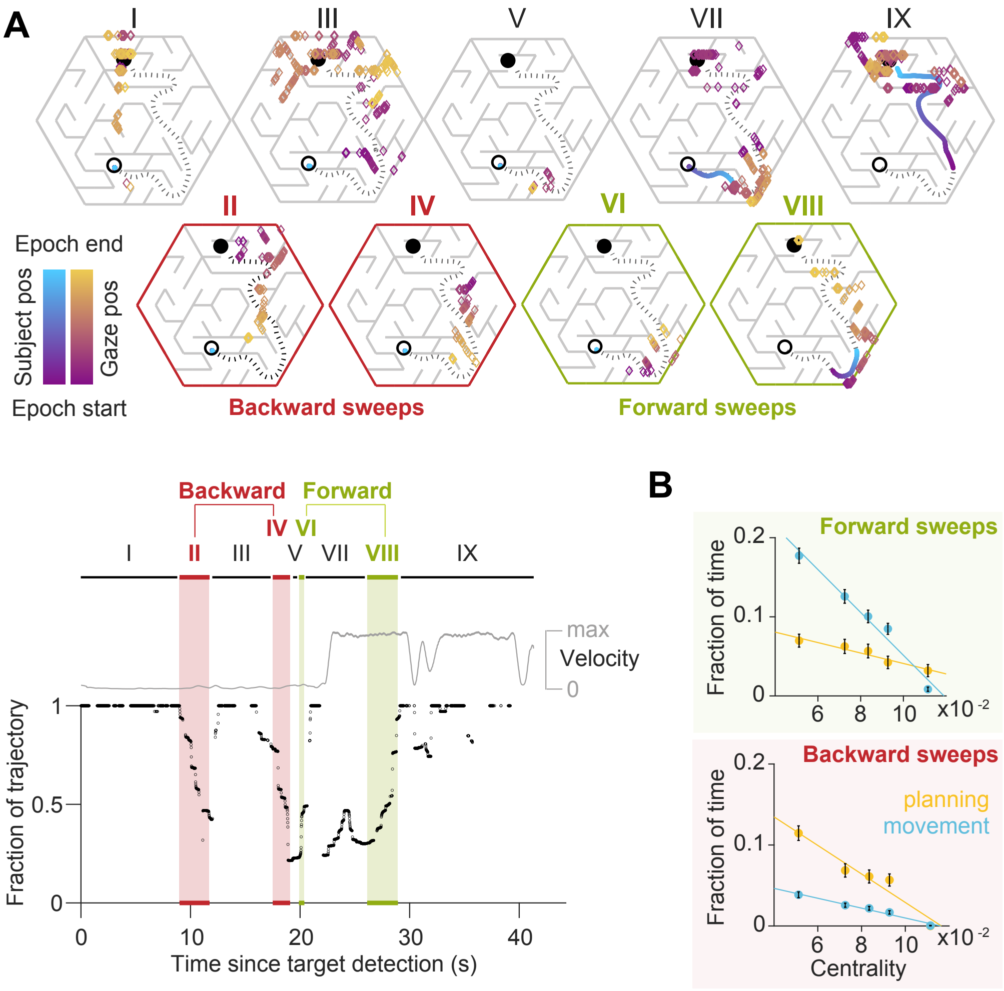

Figure 5: **Gaze traveled forwards and backwards along the intended trajectory. A.** Top: Spatial locations of gaze positions (the arrow of relative time within each window increases from violet to orange) and subject positions (violet to blue) during individual time windows demarcated in the bottom panel. Panels in the bottom row correspond to time periods corresponding to sweeps. The subject's trajectory from the starting location (open black circle) to the goal (closed black circle) is denoted by a black dashed line. Bottom: Time-series of the points on the trajectory that were closest to the subject's gaze on each frame, expressed in percentiles (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 subject 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. **B.** Across all subjects, 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). Error bars denote ±1 SEM.

The complexity of the transition structure exerted a strong influence on the probability of sweeping: the fraction of trials in which this phenomenon occurred was significantly correlated with arena centrality scores (Pearson's r = -0.95, p = 0.01; Figures S7b). This suggests that sweeping eye movements could be integral to trajectory planning. It turns out that aside from gazing upon the target or the trajectory on each trial, roughly 20% of eye movements were made to other locations in space (Figure S7c), and this percentage did not vary across arenas nor epochs. Furthermore, 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: 6.0±1.2%, movement: 2.0±0.2%; forward sweeps

— pre-movement: 5.5±0.6%, movement: 10.2±1.5%; Figure 5b). This suggests that the initial planning is primarily carried out by sweeping backwards from the goal. The mean speed of backward sweeps was greater than the speed of forward sweeps across all arenas (backward sweeps: 11.3±0.9 m/s, forward sweeps: 6.7±0.2 m/s; Figure S7d). Notably, sweep velocities were 3-4 times greater than the maximum movement velocity (2.26 m/s) and more than fivefold greater than the average subject velocity during the movement epoch (1.47±0.08). This is reminiscent of the hippocampal replay or preplay 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) [44, 45]. Both sweep speeds and durations slightly increased with arena complexity (Figures S7d). Furthermore, sweeps were comprised of increasingly more saccades as arenas became complex, and saccade rates were higher during sweeps than at other times after goal detection (Figure S7e). An explanation for this finding requires first recognizing that peripheral vision processing must lead the control of central vision to allow for sequential eye movements to trace a viable path [46, 47]. 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. However, due to the lengthier trajectories in those arenas, the gaze must cover greater distances, resulting in sweeps which last longer.

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 (Figure S7f). The latency between goal detection and the first sweep increased with arena difficulty (Figure S7g), suggesting that sweep initiation is preceded by brief processing of the arena. While the sequential nature of eye movements could constitute a swift and efficient way to perform instrumental sampling, we emphasize that task-relevant eye movements were not necessarily sequential. When we reanalyzed the spatial distribution of gaze positions by removing periods of gazing upon the trajectory on each trial, the resulting relevance values remained significantly greater than chance (Figure S7h).

What task conditions promote sequential eye movements? To find out, we computed the probability that the subjects engaged in sweeping behavior as a function of time and position, during the pre-movement and movement epochs respectively, and focusing on the dominant type of sweep during those periods (backward and forward sweeps respectively; Figure 5b). 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 subjects may be gathering some preliminary information about the environment (Figure 6a — left). During movement, on the other hand, it turns out that the probability of sweeping is heavily influenced by whether subjects 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 broken down 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 all trials with respect to subgoals. The likelihood of sweeping the trajectory in the forward direction tended to spike precisely when subjects reached a subgoal (Figure 6a — 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 6b — right). In contrast to backward sweeps, which were made predominantly to the most proximal subgoal prior to navigating (Figure 6c — left), forward sweeps that occurred during movement were not regularly directed toward one particular location. Instead, in a strikingly precise and stereotyped manner, subjects appeared to lock their gaze upon the *upcoming* subgoal when rounding each bend in the trajectory (Figure 6c — right). This suggests that subjects 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.

To summarize, we found subjects 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 typically traced the trajectory backwards from the goal to the first subgoal (Figure 6d — orange). While moving through the arena, they tended to lock their gaze upon the upcoming subgoal until they reached it, thereafter sweeping their gaze forward to the next subgoal (Figure

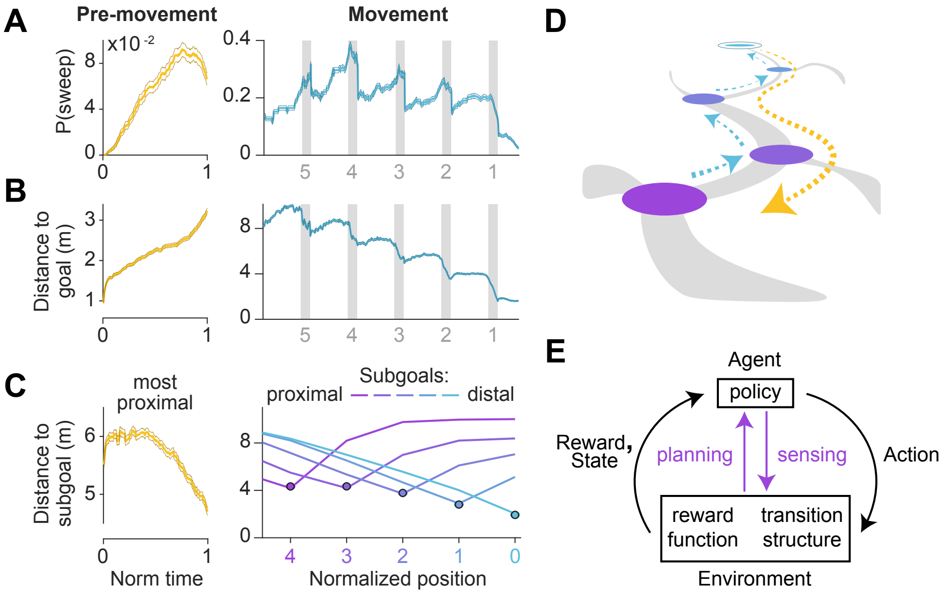

Figure 6: **Timing of sweeps reveals task decomposition.** Trials across all arenas and all subjects were aligned and scaled for the purpose of trial-averaging. This process was carried out separately for pre-movement and movement epochs. **A.** Left: Prior to movement, the probability of (backward) sweeps increased with time. Right: During movement, the probability of (forward) sweeps transiently increased at the precise moments when subjects reached each subgoal. Subject 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 target 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 subjects approached that subgoal. **D.** A graphical summary of the spatiotemporal dynamics of eye movements in this task. Subgoals are depicted in the same color scheme used in **C**. **E.** 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.

6d — blue). Via eye movements, navigators could perform active sensing and construct better-informed plans to refine their policies such that the sequence of actions that they choose would most efficiently lead to rewards (Figure 6e).

Discussion

In this study, we highlight the critical role of eye movements for model-based computations in spatial sequential-decision making tasks such as navigation in fully-observable environments. We found that humans took near-optimal trajectories when navigating to the goal, regardless of the environmental complexity. Participants divided their gaze between task-relevant transitions and the hidden reward location, with the tipping point of this balance depending upon the complexity of the environment. This compromise allowed subjects 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. In the temporal domain, subjects often traced trajectories to and from the goal (sweeping), and subsequently concentrated on one subgoal at a time until the goal was reached. All the while, the gaze was attracted to the hidden goal location either extensively or transiently depending on the relative environmental complexity. Therefore, 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.

Active sensing in navigation. Eye movements provide a natural means for researchers to understand active sensing strategies [48, 49]. Common paradigms for goal-oriented navigation block large portions of the environment from view, precluding the rigorous study of active sensing, usually in the interest of distinguishing between different navigational strategies [28, 50]. By removing such constraints, we extend previous results on active sensing from simple decision making tasks to the domain of sequential decision-making (specifically navigation) with one key adaptation: rather than testing whether eye movements reduce uncertainty about the *state* of the environment (such as whether a change in an image has occurred) [17, 18, 51], we tested whether they reduce uncertainty about the *model* of the environment. In particular, we find that the gaze is distributed between the two components of the model required to plan a path —- the transition function and the reward function —- with the distribution skewed in favor of the former in more complex environments. In general, whether active sensing reduces uncertainty about state or model would depend on whether it is engaged in the service of inference or learning. When the world state is only partially observable (e.g. due to visual occlusions), active sensing might be used primarily to reduce state uncertainty. However, when the environment is complex and unfamiliar, our results show that humans use active sensing to mitigate model uncertainty.

Navigational planning. An active topic of research in navigation is the computational algorithms underlying action selection. One set of studies investigated saccade scheduling while multitasking — e.g. keeping on the sidewalk, avoiding obstacles, and picking up litter [52–55]. More relevant to goal-oriented navigation are the studies on habitual choices (model-free decision making) [56] vs. using explicit representations of the reward function and transition structure (model-based decision making). Two example VR studies on human navigation in partially observable state spaces found evidence against model-free decision making [28, 50]. The first study involved subjects navigating between rooms arranged in a grid [28], while the second tasked subjects to navigate in simple mazes where they could only view options for the immediate next step [50]. In both studies, subjects had to take physical actions to learn the model from experience. In contrast, participants in our task could interrogate the transition structure with their eyes, resulting in the semblance of a model-based strategy. The first piece of evidence in support of a role for eye movements in model-based computation is that subjects spent more time prospecting in more complex arenas. Furthermore, shortly after fixating on the goal, subjects' gaze often swept backwards along their future trajectory, evocative of a depth-first tree search, a model-based algorithm for path discovery [57].

Perceptual grouping and task decomposition. In Crowe et. al. (2000), humans solved 2D visual mazes and were found to smoothly and accurately trace paths without having studied the maze layout [47, 58], leading the authors to conclude that peripheral vision processing leads the generation of the next saccade during path tracing. We suspect that a similar mechanism could underlie the sweeping eye movements in our task. A combination of bottom-up and top-down processes may contribute to peripheral vision processing: top-down inputs may convey the general direction in which to move the eyes (from the goal to the subject and vice versa), and bottom-up processes may convey information such as the presence and orientation of borders [59, 60]. Perceptual grouping operations may rapidly decompose abstract feature maps into computationally tractable chunks [61–63], and the oculomotor system may operate upon these chunks to sequentially discover navigable trajectories. Thus, when the environment is fully observable, cognitive computations for navigation may depend on cortical algorithms acting on sensory, rather than cognitive maps. The tendency of humans to break larger problems into smaller, more tractable subtasks has been previously established in domains outside of navigation [64–68]. When it comes to space, theories suggest that a non-flat representation would reduce memory demands, and the RL framework is frequently invoked to explain why and how [69, 70]. However, these theoretical insights have not been empirically validated in the context of navigation, primarily due to the difficulty in distinguishing between flat and hierarchical representations from behavior alone. Although our paradigm was not explicitly designed to study hierarchical path representations, the observation that subjects' locus of concentration was often locally oriented towards navigating to the end of each path segment suggests that subjects viewed turns as subgoals of the overall plan.

Neural mechanisms. Eye movements are driven by many diverse neural systems. Hippocampal projections to higher oculomotor controllers (e.g. supplemental eye fields through the orbitofrontal cortex) may guide the embodiment of simulation and memory through eye movements [71–73]. Mentally testing sequences of actions is precisely a theorized function of hippocampal preplay, which is the ordered activation of potential future state representations with or without any prior experience of traversing the space represented [34–38, 74]. Modulation also occurs in the opposite direction — the contents of gaze influences activity in the hippocampus [75–77, 77–81] and entorhinal cortex [64–66]. Although the hippocampal contribution to the task used in this study is uncertain, especially given the relatively brief exposure subjects had to each arena, it would be enlightening to examine the flow of information between brain regions with strong spatial representations and the oculomotor circuitry during fully-observable navigation tasks throughout model learning and consolidation.

Conclusion. We hope that the study of active sensing and 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 subject'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 Subject Details. Nine human subjects (all >18 years old, six males) participated in the experiments. All but two subjects (S6 and S9) were unaware of the purpose of the study. Four of the subjects, including S6 and S9, were exposed to the study earlier than the rest of the subjects, and part of the official dataset for two of these subjects (S4 and S8) was collected two months prior to the rest of data collection as a safety precaution during the COVID-19 pandemic. Six additional human subject recruits (all >18 years old, two males) were disqualified due to experiencing motion sickness while in the VR environment. All experimental procedures were approved by the Institutional Review Board at New York University and all subjects signed an approved consent form.

Stimulus. Subjects 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 subjects' eyes with respect to the ground plane) was 1.72 meters. The subject 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 of 2.26 m/s. Subjects executed angular rotations inside the arena by turning their head, while the joystick input enabled translation in the direction in which the subject'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 a greater path length between two given states, making them more complex to navigate. The order of arenas presented to each subject was randomly permuted but not 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, subjects calibrated the VIVE Pro eye tracker using inbuilt Tobii software which prompted subjects to foveate several points tiling a 2D plane in the VR environment. Both eyes were tracked, and the subject'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) [82]. This means that when the subject 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 tiny. Furthermore, Sipatchin et. al.

reported that the system latency was 58.1 ms. While there is reason to suspect that the subject's position was recorded with a similar latency of around five frames, even if the gaze data lagged the position data, the subject would only have moved 13 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 subject foveated the target, but the subject 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 subjects were tasked with navigating to the remembered target location without time constraints. Subjects were not given instructions on what strategy to use to complete the task. After reaching the target, subjects 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 subjects did not remember seeing the target before it disappeared, and these trials were recorded and excluded from the analyses (< 1%).

Reward. If subjects 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 subject'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), subjects completed one block of fifty trials in each of five arenas (Figure 1). 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 subject's speed and the length of the breaks that they needed from the virtual environment, ranging from 1.5-2 hours. Subjects 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 $27.72).

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 [83]. 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 $s \rightarrow s'$ 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 $s \rightarrow s'$ 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'. 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 $s \rightarrow s_G$ 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 methods such as value iteration, an algorithm that involves iteratively unrolling the recursion in this equation [84]. 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 1e).

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

We incorporated 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 S1b, 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 value iteration to converge.

Relevance computation. To compute the relevance $\Omega_k(s_0, s_G)$ of the k^{th} 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 more elaborate 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 [32]. 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.

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

Simulations. Behavior of three qualitatively different artificial agents with 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. 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 subject was deemed to be moving, and when the input fell below this threshold for the last time on each trial, the subject was deemed to have stopped moving. Subjects' 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, S1d, S1f, and 1g, the first trial of each run was removed from the analyses due to an occasional rapid teleportation of the subject 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 subjects 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.

Relevance estimation. Prior to estimating the task-relevance of the subjects' gaze positions at each time point, the closest transition k to the subject'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 subject'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 randomly selecting 20 trials and recomputing f based on the subject'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 (278 ms) 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 subjects subsequently took.

Saccade detection. Saccade times were classified to be eye movement speeds v which crossed the threshold of 50°/s from below, where speeds were computed using Eq 7, 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.

$$\alpha(t) = tan^{-1}(\frac{x(t)}{\sqrt{y^2(t) + z^2(t)}}),\ \beta(t) = tan^{-1}(\frac{z(t)}{\sqrt{y^2(t) + x^2(t)}}),\ v(t) = \sqrt{\frac{d}{dt}\alpha^2(t) + \frac{d}{dt}\beta^2(t)} \qquad (7)$$

Subgoal analysis. Turns in the subjects' trajectories 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 6, all trials (for all subjects in all arenas) were stop-aligned and periods of turns vs. periods of navigating straight segments were independently interpolated to fit an arbitrarily defined common timeline of 25 time points per turn and 100 time points per straight segment. Note that 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 6 increases from left to right.

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

[2] 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.

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

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

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

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

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

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

[15] 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.

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

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

[18] 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.

[19] Foley NC, Kelly SP, Mhatre H, Lopes M, Gottlieb J. Parietal neurons encode expected gains in instrumental information. *Proceedings of the National Academy of Sciences of the United States of America*, 114(16):E3315–E3323, 2017. ISSN 10916490.

[20] Gottlieb J. Understanding active sampling strategies: Empirical approaches and implications for attention and decision research. *Cortex*, 102:150–160, 2018.

[21] Daddaoua N, Lopes M, Gottlieb J. Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates. *Scientific Reports*, 6(February):1–15, 2016. ISSN 20452322.

[22] Blacker KJ, Weisberg SM, Newcombe NS, Courtney SM. Keeping Track of Where We Are: Spatial Working Memory in Navigation. *Vis Cogn*, 25(7-8):691–702, 2017.

[23] Tolman EC. Cognitive maps in rats and men. *Psychological Review*, 55(4):189–208, 1948. ISSN 0033295X.

[24] Epstein RA, Patai EZ, Julian JB, Spiers HJ. The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, 20(11):1504–1513, 2017. ISSN 15461726.

[25] Behrens TE, Muller TH, Whittington JC, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson Z. What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, 100(2):490–509, 2018. ISSN 10974199.

[26] Maguire EA, Frackowiak RSJ, Frith CD. Recalling Routes around London: Activation of the Right Hippocampus in Taxi Drivers. *Journal of Neuroscience*, 17(18):7103–7110, 1997.

[27] 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.

[28] 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.

[29] Paulus MP, Potterat EG, Taylor MK, Van Orden KF, Bauman J, Momen N, Padilla GA, Swain JL. A Neuroscience Approach to Optimizing Brain Resources for Human Performance in Extreme Environments. *Neurosci Biobehav Rev*, 22(7):1080–1088, 2009.

[30] Sutton RS, Barto AG. *Reinforcement Learning: An Introduction 2nd Ed*. 2013.

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

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

[33] Craik KJW. Hypothesis on the nature of thought. In *The Nature of Explanation*, chapter 5, pages 51–61. Cambridge University Press, 1943.

[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] 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.

[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] 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.

[43] 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.

[44] 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.

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

[46] 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.

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

[48] 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.

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

[50] 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.

[51] 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.

[52] Sprague N, Ballard D. Eye movements for reward maximization. *Advances in Neural Information Processing Systems*, 2004. ISSN 10495258.

[53] Rothkopf CA, Ballard DH, Hayhoe MM. Task and context determine where you look. *Journal of Vision*, 7(14):1–20, 2007. ISSN 15347362.

[54] Shinoda H, Hayhoe MM, Shrivastava A. What controls attention in natural environments? *Vision Research*, 41(25-26):3535–3545, 2001. ISSN 00426989.

[55] Sullivan BT, Johnson L, Rothkopf CA, Ballard D, Hayhoe M. The role of uncertainty and reward on eye movements in a virtual driving task. *Journal of Vision*, 12(13):1–17, 2012. ISSN 15347362.

[56] 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.

[57] 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.

[58] 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.

[59] McMains S, Kastner S. Interactions of top-down and bottom-up mechanisms in human visual cortex. *Journal of Neuroscience*, 31(2):587–597, 2011. ISSN 02706474.

[60] Julian J, Keinath A, Frazzetta G, Epstein R. Human entorhinal cortex represents visual space using a boundary-anchored grid. *Nature Neuroscience*, 21(2):191–194, 2018.

[61] Xu Y, Chun MM. Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 104(47):18766–18771, 2007. ISSN 00278424.

[62] Roelfsema PR. Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, 29:203–227, 2006. ISSN 0147006X.

[63] Barsalou LW. Grounded cognition. *Annual Review of Psychology*, 59:617–645, 2008. ISSN 00664308.

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

[65] 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.

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

[67] 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.

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

[69] 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.

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

[71] 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.

[72] 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.

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

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

[75] Meister ML, Buffalo EA. Getting directions from the hippocampus: The neural connection between looking and memory. *Neurobiology of Learning and Memory*, 134:135–144, 2016. ISSN 10959564.

[76] 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.

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

[78] 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.

[79] 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.

[80] 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.

[81] 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.

[82] 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.

[83] 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.

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

Supplemental materials

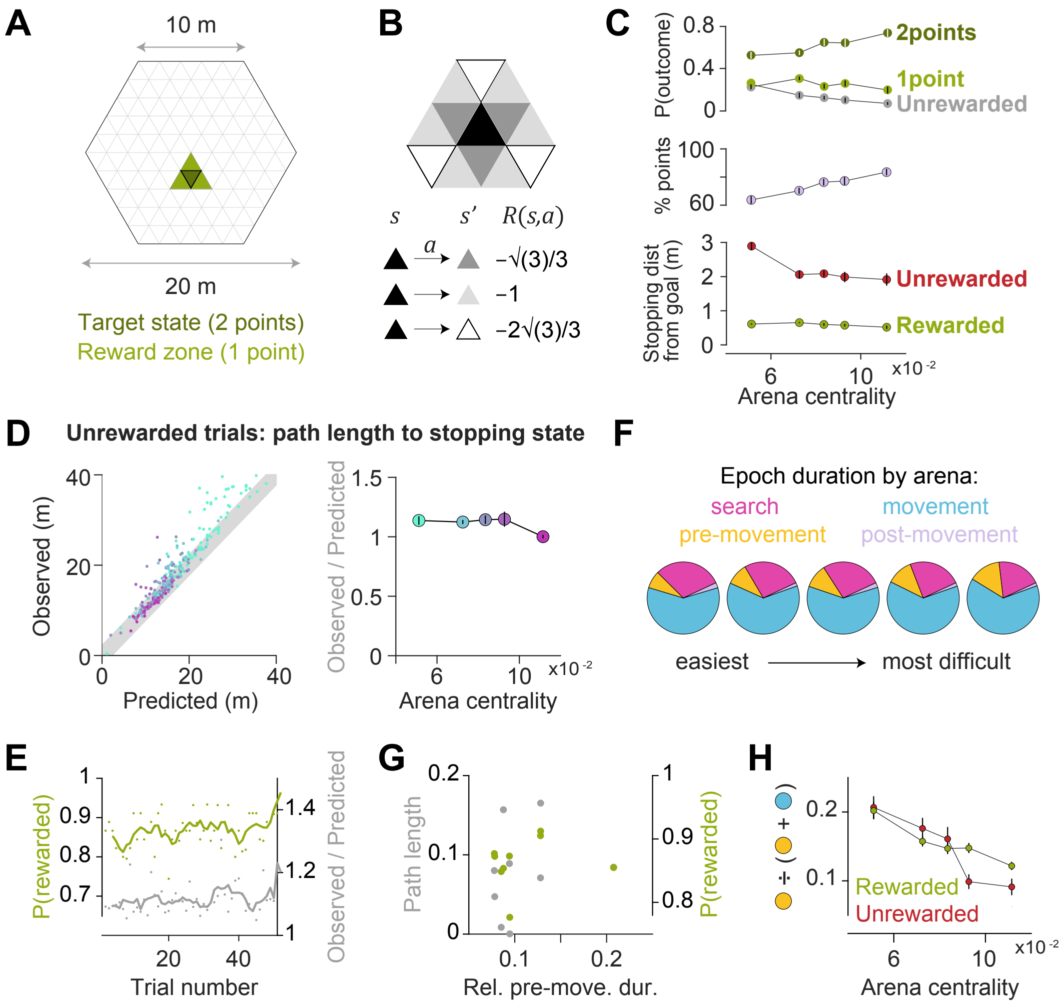

Figure S1: **A.** Arenas were regular hexagons with a side length of 10 m, and a triangular tessellation with a unit length of 2 m. Two points were rewarded if participants reached the goal state (green), and one point was rewarded if participants reached a state neighboring the goal state (light green). **B.** To incorporate twelve degrees of freedom in translation, value functions were computed by dynamic programming whereby the cost of actions scaled in accordance with the center-to-center distance between states s and s' (pertaining to the transition which results from taking action a). **C.** Top: Across all subjects and all trials, the probability of being awarded two points (green) increased with arena centrality, while the probability of being awarded one point (light green) is relatively constant across all arenas. Gray denotes the probability of not being rewarded. Middle: The total fraction of points earned increased as a function of arena centrality. Bottom: Distance between the stopping location and goal in rewarded (green) and unrewarded (red) trials. Error bars denote ±1 SEM. **D.** Left: Across all arenas (colored according to the arena coloring scheme introduced in Figure 1b), the path lengths observed in unrewarded trials were close to the optimal trajectory lengths between the starting state and the state at which subjects stopped on these trials, suggesting that unrewarded trials were predominantly caused by subjects forgetting the precise location of the target. Right: The ratio of observed to optimal path lengths (to the subjects' stopping location on unrewarded trials) was close to unity in all arenas. **E.** Performance was stable across each block, as measured by the average probability of being rewarded on each trial (green), as well as the average ratio between the empirical and optimal path lengths (gray). **F.** Distribution of epoch durations across all subjects and all trials. Pre-movement occupied a greater fraction of the total trial time for more complex arenas. **G.** Some subjects spent lesser time deliberating before movement, but this did not impact task performance. Relative pre-movement duration was defined as the average ratio of the duration of the pre-movement epoch to the duration of the entire trial after goal detection. The average proportion of time that subjects spent making prospective eye movements prior to using the joystick did not correlate with their average path lengths across all arenas (gray), nor with the overall probability of them being rewarded (green). **H.** The relative pre-movement duration did not differ between rewarded and unrewarded trials (after matching the mean trial duration of the two groups, separately for each arena), except for the two easiest arenas, where planning demands are low. This suggests that failure to obtain rewards is not mere due to poor planning. Error bars denote ±1 SEM.

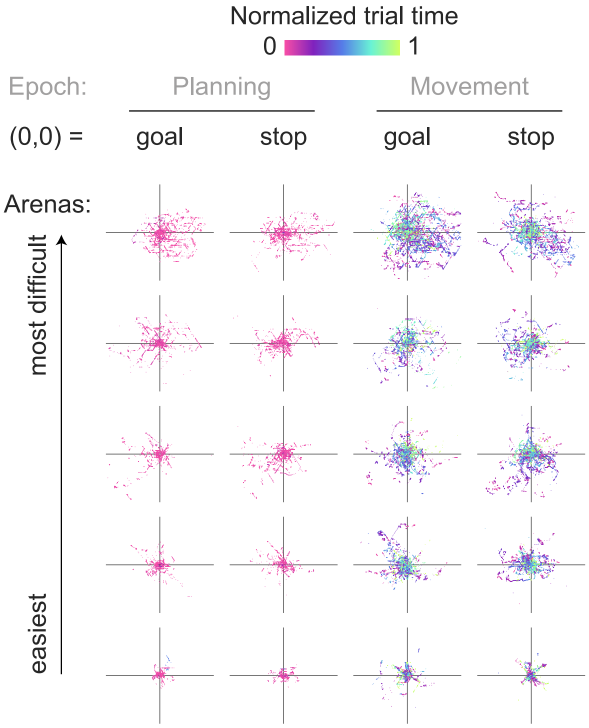

Figure S2: Gaze is increasingly concentrated around the goal/stopping location for easier arenas. The believed goal location was assumed to be the subjects' stopping position, and the point of gaze was visibly more concentrated around the stop location than around the true goal location (especially in the most complex arena). The effects of working memory on gaze were more apparent for easier, more open arenas. Each panel depicts eye movements on a random subset of trials (3 trials x 9 subjects) in each arena. The origin (0,0) denotes the goal location or the stopping location. Raw gaze positions relative to these points are depicted during the pre-movement and movement epochs. Axis limits are ±15 m for all panels.

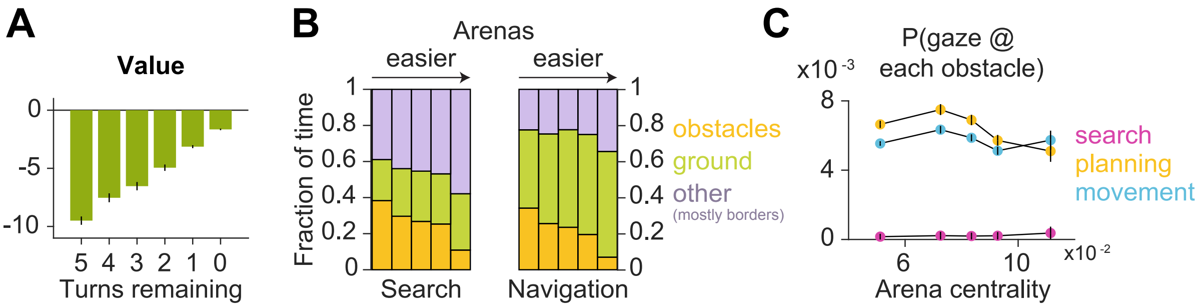

Figure S3: **A.** The average value (calculated using dynamic programming) of the states upon which subjects gaze increased as the subject approached the remembered goal location. Specifically, the greater the number of turns remaining in the trajectory (where turns were classified using a threshold applied on angular velocity), the lower the value of the states upon which the subjects looked, partly due to a lower probability of gazing upon or near the goal state (see Figure 6). Values shown here are negative due to the formulation of value as a function of the path length from a state to the target state (the value of the target state is zero). **B.** During search, subjects spend a greater fraction of time foveating the arena borders (purple) than during the active navigation phase (which consists of both the pre-movement and movement epochs). During navigation (both before and during movement), subjects spend more time foveating the ground (green). While there appears to be a trend in the fraction of time foveating obstacles (orange) vs. arena centrality, this is explained by a higher obstacle density in the more complex arenas. **C** Across all subjects and all trials, the probability of gazing upon each obstacle remains relatively constant across all arenas during each epoch. All error bars denote ±1 SEM.

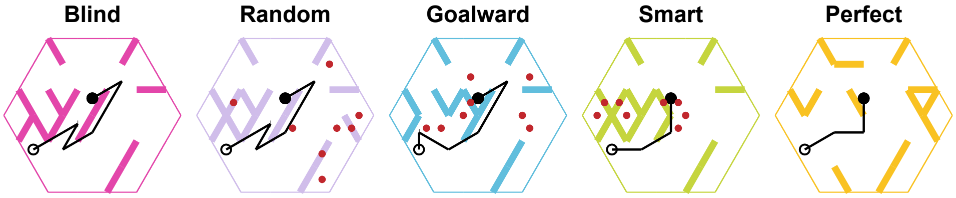

Figure S4: Example simulated trajectories, as well as the gaze samples (red dots, if applicable), taken by each agent. The configuration of the arena reflects the agent's subjective model at the *end* of all eye movements. Note that the subjective model of the "smart" agent was still quite mismatched with the true world model after eight eye movements, but the visual samples allowed for the correction of the model at crucial locations such that the trajectory of the "smart" agent was closer to optimal than that of the other agents.

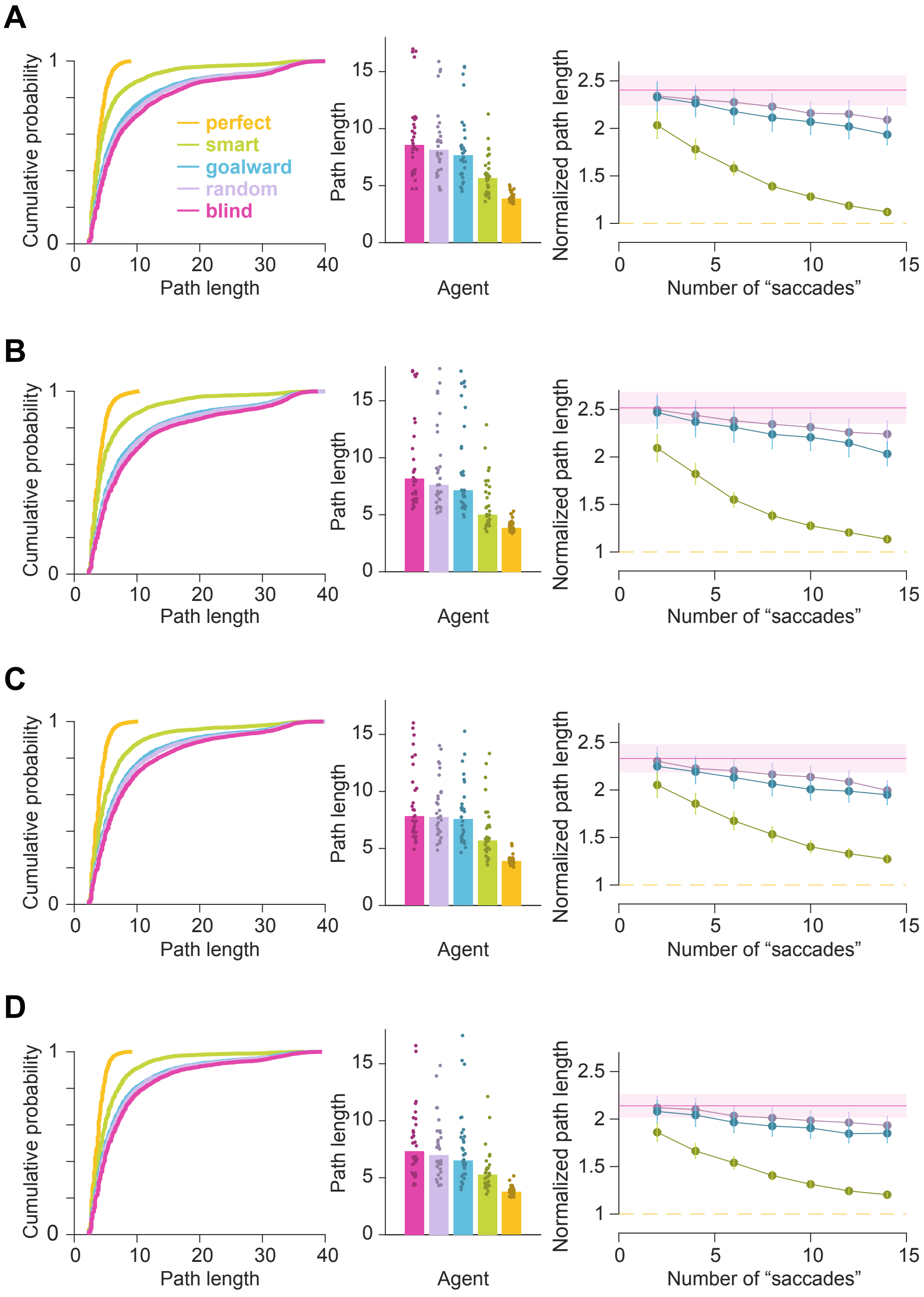

Figure S5: Simulations reveal that foveating 'relevant transitions' reduces path length. Results were robust to the precise algorithm (value iteration vs. successor representation) as well as the degree of temporal abstraction (current state vs. optimal trajectory) used to estimate the relevance of transitions. Plots similar to Figure 3c and 3d are shown for relevance values calculated with **A** value iteration, current state, **B** value iteration, entire trajectory, **C** successor representation, current state, and **D** successor representation, entire trajectory.

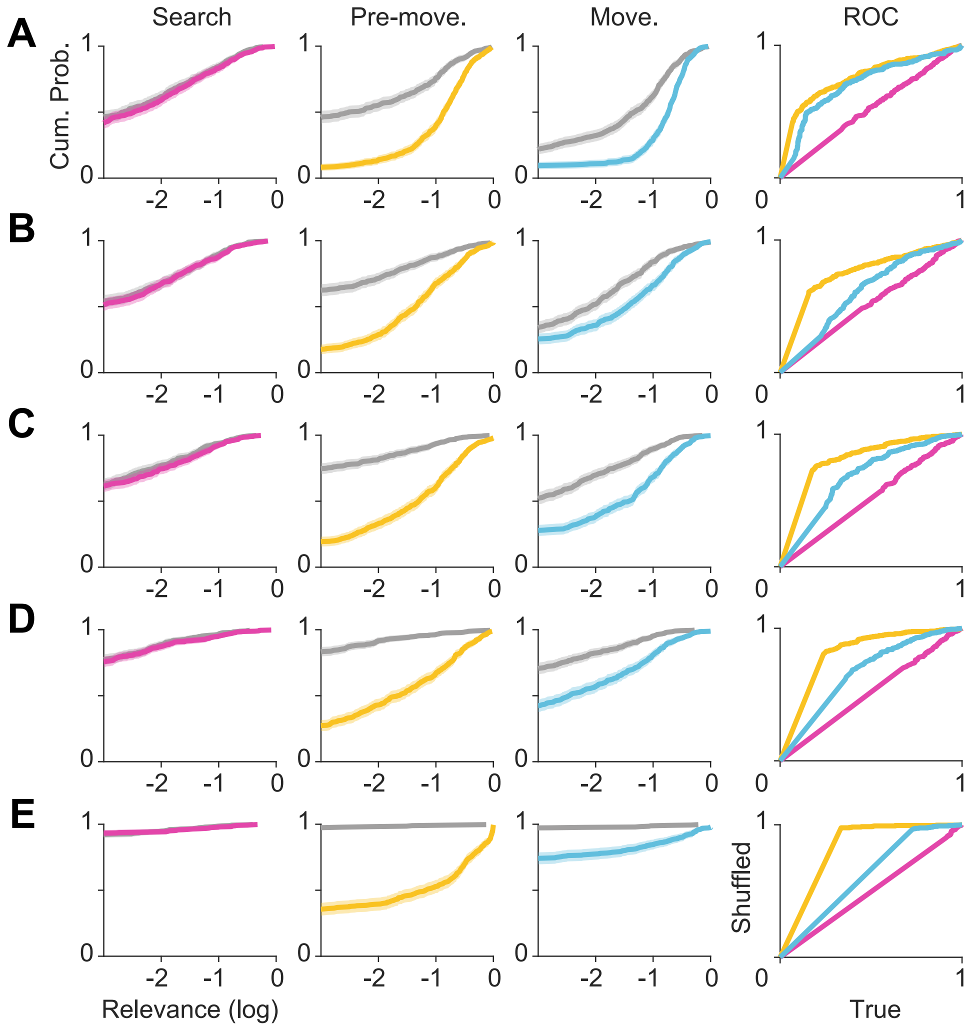

Figure S6: **A-E:** Breakdown of Figure 4a for arenas 1 (most complex) through 5 (least complex).

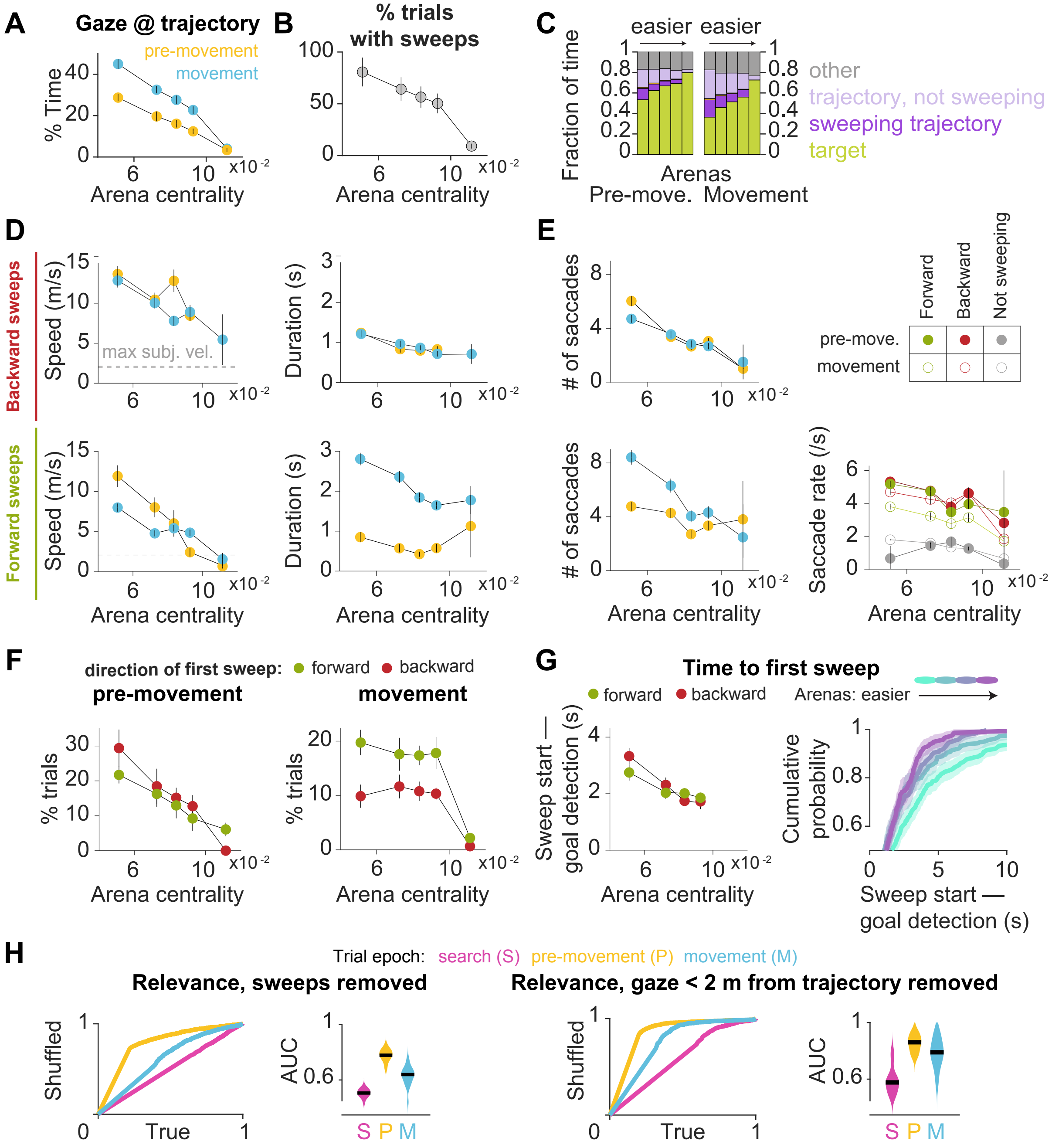
A
Gaze @ trajectory
pre-movement
movement
% Time
0
20
40
6
10
x10⁻²
Arena centrality
B
% trials with sweeps
0
50
100
6
10
x10⁻²
Arena centrality
C
easier easier
Fraction of time
0
0.2
0.4
0.6
0.8
1
Arenas
Pre-move. Movement
other
trajectory, not sweeping
sweeping trajectory
target
D
Backward sweeps
Speed (m/s)
0
5
10
15
max subj. vel.
6
10
x10⁻²
Duration (s)
0
1
2
3
Forward sweeps
Speed (m/s)
Duration (s)
Arena centrality
Arena centrality
E
# of saccades
0
4
8
Forward
Backward
Not sweeping
pre-move.
movement
# of saccades
Saccade rate (/s)
0
2
4
6
Arena centrality
Arena centrality
F
direction of first sweep: ● forward ● backward
pre-movement
movement
% trials
0
10
20
30
% trials
0
10
20
Arena centrality
Arena centrality
G
Time to first sweep
● forward ● backward
Arenas: easier
Sweep start — goal detection (s)
0
2
4
6
10
x10⁻²
Arena centrality
Cumulative probability
0.6
0.8
1
0
5
10
Sweep start — goal detection (s)
H
Trial epoch: search (S) pre-movement (P) movement (M)
Relevance, sweeps removed
Relevance, gaze < 2 m from trajectory removed
Shuffled
0
True
1
AUC
0.6
1
S P M

Figure S7: **A.** The fraction of time that subjects spent gazing within 2 m from the trajectory that they took on each trial (excluding points of gaze within 2 m from the goal location) decreased with arena centrality. **B.** The fraction of trials with sweeps was lower for less complex arenas. Error bars for **A** and **B** denote ±1 SEM across subjects. **C.** Eye movements on each trial was decomposed into fixations within 2 m of the target (green), fixations within 2 m of the subject's trajectory (excluding the target) during sweeps vs. outside of sweeps, and fixations outside of sweeps that were neither made to the target nor trajectory. Subjects viewed the hidden target location more in easier arenas, and gazed upon the rest of the trajectory more in difficult arenas. The fraction of time which was spent looking elsewhere was relatively constant across arenas and epochs. **D.** Left: Across all subjects and all trials, the speed of backward sweeps (bottom plot) was greater than the speed of forward sweeps (top plot) for all arenas. Before movement (orange), the speed of forward sweeps was faster than that during movement (blue) in the two most complex arenas. Right: Forward sweeps before movement, as well as backwards sweeps before and during movement, were relatively constant in duration across different arenas. However, forward sweeps during movement were longer in the two most complex arenas. **E.** Left: The average number of saccades per sweep was lower for easier arenas, and was highest during forward sweeps while the subject was actively moving. Right: The saccade rate was higher during sweeps than outside of sweeps. **F.** The direction of the first sweep was more likely to be backwards if it occurred prior to movement (left), and forwards if it occurred during movement (right). While the fraction of trials with a sweep occurring prior to movement increased as a function of arena difficulty, the fraction of trials for which the first sweep occurred during movement was rather constant for all but the easiest arena. **G.** Left: The average delay between goal detection and the first sweep increased with arena difficulty. Right: The cumulative distribution of the delays across all subjects and all trials. **H.** Left: ROC curves constructed as described in Figure 4a (rightmost) for the distributions of true vs. shuffled average relevance values for each trial (pooled across all arenas, all subjects, and all trials), with periods of sweeping eye movements removed, reveals that during the pre-movement (orange) and movement (blue) epochs, non-sequential eye movements are still directed towards task-relevant locations. AUC plots were constructed with sweeps removed, as described in Figure 4b. The AUC values remain well above chance during the pre-movement (orange) and movement (blue) epochs. Right: The same analysis was performed with gaze positions falling within 2 m of the subject's trajectory on each trial removed, revealing that the remaining visual samples were still made to relevant locations in space.

Subject ID	Block 1	Block 2	Block 3	Block 4	Block 5
1	3	2	1	4	5
2	5	4	1	2	3
3	5	3	1	2	4
4	4	5	2	3	1
5	3	4	2	5	1
6	5	2	1	3	4
7	3	2	5	1	4
8	4	5	3	1	2
9	2	1	4	5	3

Table S1: The order of arena presentation was randomized across subjects.

Epoch	Arena 1	Arena 2	Arena 3	Arena 4	Arena 5
search	3.3 {6}	5.2 {5}	0 {3}	0 {4}	0 {2}
pre-movement	135 {10}	45 {11}	51 {11}	22 {11}	67 {17}
movement	184 {8}	36 {10}	38 {9}	3.5 {10}	0 {10}

Table S2: Median true relevance values {with interquartile range (IQR)} for each arena (x10^{-3}).

Relevance derivation: In this section, we derive a general measure to quantify the relevance of transitions with respect to the task of navigating between two given states. The following derivation focuses on the general setting when external noise (stochastic transitions) is present and internal noise (model uncertainty) is inhomogeneous. As we will show, the measure used to quantify transition relevance in the main text (Equation 1) corresponds to the special case where transitions are deterministic and uncertainty is homogeneous. Let T_k denote the status of the k^{th} stochastic transition (1 or 0) and p_k be the parameter of the true probability distribution (p.d.) of that transition such that $P(T_k = 1) = p_k$ and $P(T_k = 0) = 1 - p_k$. Let $\hat{p}_k$ be the parameter of the subjective probability distribution of the transition. I.e. The agent thinks that $P(T_k = 1) = \hat{p}_k$ and $P(T_k = 0) = 1 - \hat{p}_k$. Given a particular goal state, let V_s^k denote the value of the agent's current state s evaluated using the true transition status T_k such that $V_s^k = V_s(T_k = 1)$ if $T_k = 1$ and $V_s^k = V_s(T_k = 0)$ if $T_k = 0$. Let $\hat{V}_s^k$ denote the expectation of the value of state s evaluated using the subjective transition p.d. of the k^{th} transition such that $\hat{V}_s^k = \hat{p}_k\, V_s(T_k = 1) + (1 - \hat{p}_k)\, V_s(T_k = 0)$. Since looking at a transition will dramatically reduce the uncertainty about the status of that transition, this can impact the subjective value of the current state, provided that transition is critical to the task at hand. For instance, discovering a subway line linking your neighborhood and downtown will increase the value of your neighborhood if your workplace is located downtown, but will have no impact if your workplace is located crosstown. Therefore, we define relevance (Ω_k) of the k^{th} transition as the expectation of the (log) change in subjective value about the current state s induced by looking at that transition. Then, we have:

$$
\begin{aligned}
\Omega_k = \mathbb{E}[\,\log(|V_s^k - \hat{V}_s^k|)\,]_{p_k} \;&\text{where } \mathbb{E}[\,.\,]_{p_k} \text{ denotes expectation taken w.r.t the true p.d.}\\
&= \mathbb{E}[\,\log(|V_s^k - \hat{p}_k\, V_s(T_k = 1) - (1 - \hat{p}_k)\, V_s(T_k = 0)|)\,]\\
&= p_k \log(|V_s(T_k = 1) - \hat{p}_k\, V_s(T_k = 1) - (1 - \hat{p}_k)\, V_s(T_k = 0)|) +\\
&\qquad (1 - p_k) \log(|V_s(T_k = 0) - \hat{p}_k\, V_s(T_k = 1) - (1 - \hat{p}_k)\, V_s(T_k = 0)|)\\
&= p_k \log((1 - \hat{p}_k)|\Delta V|) \;+\; (1 - p_k) \log(\hat{p}_k|\Delta V|) \text{ where } \Delta V = V_s(T_k = 1) - V_s(T_k = 0)\\
&= p_k \log(1 - \hat{p}_k) \;+\; (1 - p_k) \log(\hat{p}_k) \;+\; \log(|\Delta V|)\\
&= (p_k - 1 + 1) \log(1 - \hat{p}_k) \;-\; p_k \log(\hat{p}_k) \;+ \log(\hat{p}_k) \;+\; \log(|\Delta V|)\\
&= -(1 - p_k) \log(1 - \hat{p}_k) \;-\; p_k \log(\hat{p}_k) \;+ \log(\hat{p}_k) \;+ \log(1 - \hat{p}_k) \;+\; \log(|\Delta V|)\\
&= \mathsf{H}(p_k, \hat{p}_k) \;+\; \log(\hat{p}_k\,(1 - \hat{p}_k)) \;+\; \log(|\Delta V|)\\
&\qquad \text{where } \mathsf{H}(X, Y) \text{ denotes the cross entropy between } X \text{ and } Y\\
&= \mathsf{H}(p_k) \;+\; D_{\mathsf{KL}}(p_k||\hat{p}_k) \;+\; \log(\hat{p}_k\,(1 - \hat{p}_k)) \;+\; \log(|\Delta V|)\\
&\qquad \text{where } \mathsf{H}(X) \text{ denotes the entropy of } X\\
&= \mathsf{H}(p_k) \;+\; D_{\mathsf{KL}}(p_k||\hat{p}_k) \;+\; \log(\mathsf{Var}[\hat{T}_k]) \;+\; \tfrac{1}{2}\log(|\Delta V|^2)\\
&\qquad \text{where } \hat{T}_k \text{ denotes the subjective knowledge about the status of the } k^{th} \text{ transition}
\end{aligned}
$$

Observe that Ω_k is comprised of four factors: (**I**) $\mathsf{H}(p_k)$, the entropy of p_k, which captures transition volatility, (**II**) $D_{\mathsf{KL}}(p_k||\hat{p}_k)$, the Kullback-Leibler divergence between true and subjective p.d., which captures the degree of mismatch between the subjective and true transition models, (**III**) $\log(\mathsf{Var}[\hat{T}_k])$, the log variance of the subjective status of the transition, which captures the agent's uncertainty, and (**IV**) $\log(|\Delta V|^2) = \log([V_s(T_k = 1) - V_s(T_k = 0)]^2)$, the log change in the value of the current state induced by changing the transition status, which captures the sensitivity of the value function to the transition. The first and third terms suggest that an agent should prioritize looking at transitions with high volatility and high subjective uncertainty. The second term suggests that it is best to look at transitions whose subjective status is known to be wrong. Although this is mathematically correct, agents would not know the true model to begin with and therefore cannot direct their attention at such transitions. Therefore, if external and internal noise are homogeneous, the best strategy would be to look at transitions which the value function is highly sensitive to, as postulated by Equation 1 in the main text. Note that in deriving Ω_k, we have neglected the contribution of model mismatches that may exist at other transitions. This approximation will be valid if the model mismatch is small, and the solution works well in practice as demonstrated by the simulations (Figure 3).