Appendix A. Appendix

A.1. Experimental setup

Thirteen human participants participated in the experiments. All experimental procedures were approved by the Institutional Review Board at the university and all participants signed an informed consent form (IRB-FY2019-2599). Participants were seated on a swivel chair with 360° of freedom in physical rotation and navigated in a full-immersion hexagonal virtual maze with several obstacles. The stimulus was rendered at a frame rate of 90 Hz using the Unity game engine and was viewed through an HTC VIVE Pro virtual reality headset. The subjective vantage point was 1.72 m above ground level and the field of view of 110.1° of visual angle. Forward and backward translation was enabled via a continuous control CTI Electronics M20U9T-N82 joystick. Participants executed angular rotations inside the maze by turning their head, while the joystick input enabled translation in the direction in which the participant's head was facing. Obstacles and maze boundaries appeared as gray, rectangular slabs of concrete.

A.2. Maze structure

The maze was a regular hexagon enclosing an area of approximately $260m^2$ of navigable space. For ease of simulation and data analyses, the maze 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. The maze contained several obstacles in the form of obstacles (0.4 m high) located along the edges between certain triangles (states). Outer boundary walls of height 2.5 m enclosed the maze. We chose five mazes (including an open maze) spanning a large range in average state closeness centrality, where closeness centrality is defined as the inverse average path length from state to every other state. On average, mazes with lower centrality will impose greater path lengths between two given states, making them more complex to navigate. The order of mazes presented to each participant was randomly permuted.

A.3. Trial structure

At the beginning of each trial, a target in the form of a realistic banana appeared hovering 0.4 m over a state randomly drawn from a uniform distribution over all possible states. The joystick input was disabled until the participant foveated the target, but the participant was free to scan the environment by rotating in the swivel chair during the visual search period. About 200 ms after target foveation, the banana disappeared and participants were tasked with navigating to the remembered target location without time constraints. Participants were not given instructions on what strategy to use to complete the task. After reaching the target, participants pressed a button on the joystick to indicate that they have completed the trial. Feedback was displayed immediately after the button press.

A.4. Reward structure

If participants stopped within the triangular state which contained the target, they were rewarded with two points. If they stopped in a state sharing a border with the target

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

A.5. Sweep detection

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

A.6. Quantification of sequence similarity

We quantified the similarity between the sequence of states *viewed* during planning and the sequence of states *visited* while subsequently navigating towards the goal using *ordering-based sequence similarity* (OSS) (Gomez-Alonso & Valls, 2008). Briefly, let $i = (x_{i,1}, ..., x_{i,card(i)})$ and $j = (x_{j,1}, ..., x_{j,card(j)})$ be two sequences of different lengths and let $S = (s_1, ..., s_n)$ be the set of all possible states comprising the elements of those sequences. The the (dis)similarity between two sequences i and j is quantified as:

$$d_{\mathrm{OSS}}(i,j) = \frac{f(i,j) + g(i,j)}{\mathrm{card}(i) + \mathrm{card}(j)} \tag{6}$$

where

$$g(i,j) = \mathrm{card}(\{x_{ik} | x_{ik} \notin j\}) + \mathrm{card}(\{x_{jk} | x_{jk} \notin i\}) \tag{7}$$

measures the number of uncommon elements in the sequences and

$$f(i,j) = \frac{\sum_{k=1}^{n} | \sum_{p=1}^{\Delta} i_{(l_k)}(p) - j_{(l_k)}(p)|}{\max\{\mathrm{card}(i), \mathrm{card}(j)\}} \tag{8}$$

measures the relative distance between the position of common elements across the two sequences. d_{OSS} ranges between 0 (identical) and 1 (non-overlapping).

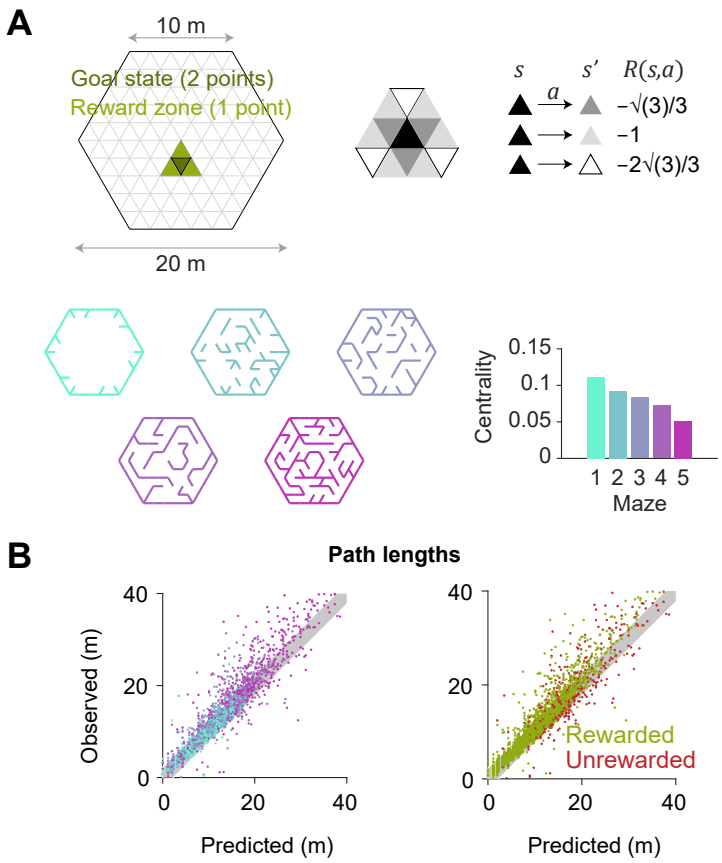

Figure 6: Participants exhibit near-optimal navigation performance across multiple environments. **A**. Top left: Mazes were regular hexagons of side length 10 m with a triangular tessellation of side length 2 m. Two points were rewarded if participants reached the goal state (green), and one point was rewarded if participants reached a state neighboring the goal state (light green). Top right: To incorporate twelve degrees of freedom in translation, value functions were computed using 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). Bottom left: Aerial view showing the layout of the mazes. Bottom right: Mazes ranged in mean state closeness centrality (mazes with higher complexity have lower centrality). **B**. Comparison of the empirical path length against the path length predicted by the optimal policy calculated using dynamic programming. The gray shaded region denotes the width of the outer reward zone (see panel A). Left: Data points are colored in accordance to the colors of each maze as depicted in A. Right: Unrewarded trials (red) vs rewarded trials (green) had similar path lengths. For both plots, all trials for all participants and all mazes are superimposed. Trials with observed path length lower than optimal correspond to paths that do not terminate exactly on the goal state, but some of these may terminate within the reward zone.

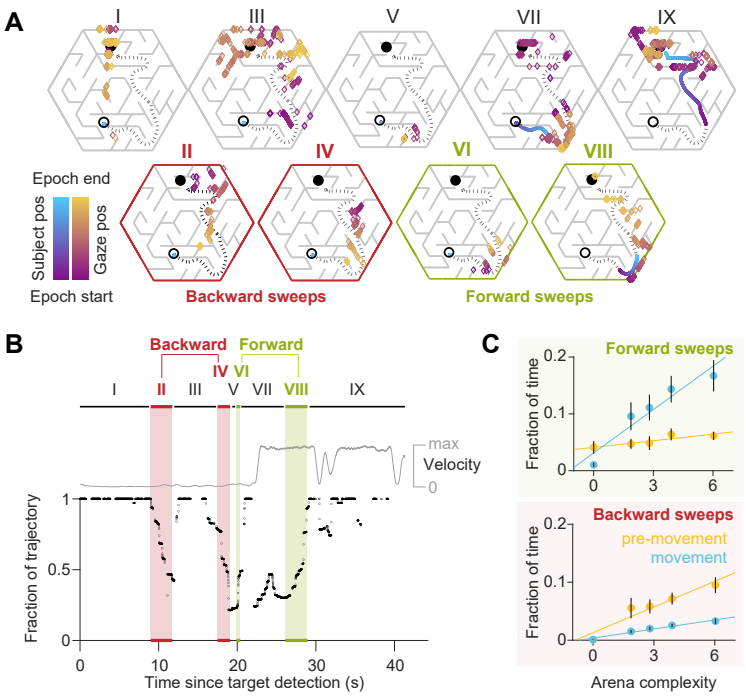

Figure 7: Gaze traveled forwards and backwards along the intended trajectory. **A**. Spatial locations of gaze positions (the arrow of relative time within each window increases from violet to orange) and participant positions (violet to blue) during individual time windows demarcated in panel B. Panels in the bottom row correspond to time periods corresponding to sweeps. The participant's trajectory from the starting location (open black circle) to the goal (closed black circle) is denoted by a black dashed line. **B** Time-series of the points on the trajectory that were closest to the participant's gaze on each frame, expressed as a fraction (0: start of trajectory, 1: end of trajectory) during one example trial. Only frames during which the gaze position fell within 2 m of the trajectory are plotted. The gray trace shows the movement velocity of the participant during this trial. Red and green shaded regions highlight time windows during which the sweep classification algorithm detected backward and forward sweeps, respectively. In this trial, there were two backward sweeps before movement, and one forward sweep each before and during movement. **C**. Across participants, the fraction of time spent sweeping in the forward and backward directions increased with the maze complexity. Error bars denote ±1 SEM. Figure 2F of the manuscript does not distinguish between forward and backward sweeps.

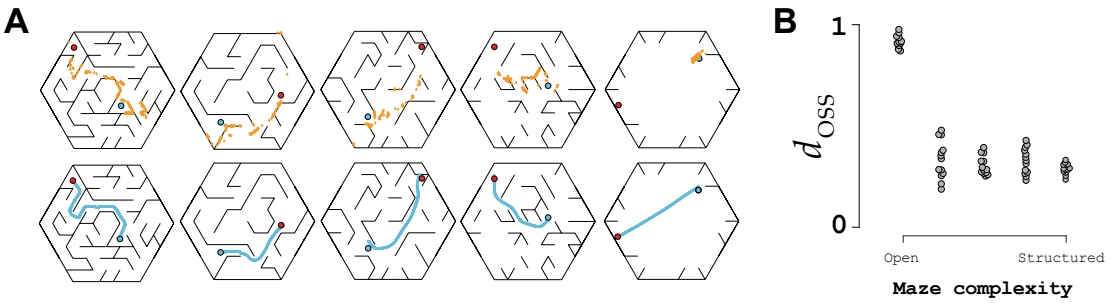

Figure 8: Similarity between active sensing strategy during planning and navigation. **A**. Example trials showing the spatial pattern of gaze during planning (orange) and path subsequently taken during navigation (blue) in different mazes. **B**. The ordering-based sequence similarity, d_{OSS} (similar sequences have lower values) between the states viewed during planning and states visited during navigation for individual participants ($n = 13$).