eNeuro

Research Article: New Research

Cognition and Behavior

Learning-Induced Shifts in Mice Navigational Strategies Are Unveiled by a Minimal Behavioral Model of Spatial Exploration

Christina-Anna Vallianatou,* Alejandra Alonso,* Adrian Zapata Aleman, Lisa Genzel,* and Federico Stella*

https://doi.org/10.1523/ENEURO.0553-20.2021

Donders Institute for Behaviour and Cognition, Radboud University, Nijmegen 6500GL, The Netherlands

Abstract

Shifts in spatial patterns produced during the execution of a navigational task can be used to track the effects of the accumulation of knowledge and the acquisition of structured information about the environment. Here, we provide a quantitative analysis of mice behavior while performing a novel goal localization task in a large, modular arena, the HexMaze. To demonstrate the effects of different forms of previous knowledge we first obtain a precise statistical characterization of animals' paths with sub-trial resolution and over different phases of learning. The emergence of a flexible representation of the task is accompanied by a progressive improvement of performance, mediated by multiple, multiplexed time scales. We then use a generative mathematical model of the animal behavior to isolate the specific contributions to the final navigational strategy. We find that animal behavior can be accurately reproduced by the combined effect of a goal-oriented component, becoming stronger with the progression of learning, and of a random walk component, producing choices unrelated to the task and only partially weakened in time.

Key words: schema; cognitive map; spatial navigation; behavioral model

Significance Statement

This work presents a novel statistical analysis we applied to describe mice behavior during a goal-reaching task in a large, modular environment (HexMaze). By combining sub-trial quantification of animal navigation with mathematical modeling of the task, we aim at developing analysis tools that can match the demands of rich, articulated experimental paradigms. We show how mice progressively incorporate information about the task and the maze structure and how such knowledge accumulation unfolds over multiple time scales. We also demonstrate how mice never completely converge to optimal behavior: even in late phases of learning a substantial part of their behavior can be ascribed to purely random choices with no relationship with the location of the goal.

Introduction

The problem of learning, and especially of the integration of new information into an already existing knowledge structure, is at the center of the effort to understand brain functioning (Alonso et al., 2020b). When using rodent animal models, such problem has often been addressed in the context of spatial navigation and map learning (Wang and Morris, 2010; Richards et al., 2014). Animals' knowledge about the environment, and the degree to which they can acquire new information, can be linked to their ability to easily navigate to specific locations and flexibly adapt to changes in the environment (Behrens et al., 2018). Nevertheless, the characterization of the effects of

Received December 22, 2021; accepted June 9, 2021; First published July 30, 2021.

The authors declare no competing financial interests.

Author contributions: A.A., L.G., and F.S. designed research; C.-A.V., A.A., and A.Z.A. performed research; F.S. contributed unpublished reagents/analytic tools; C.-A.V., A.Z.A., and F.S. analyzed data; L.G. and F.S. wrote the paper.

learning has been mostly restricted to simple tasks, with limited spatial and temporal complexity, focusing on isolating specific components of the learning process with highly controlled paradigms. Indeed, the difficulties in precisely monitoring the animal behavior pose one of the major limiting factors in the development of more comprehensive experimental paradigms (Fonio et al., 2009). Here, we aim at filling this gap by providing a quantitative framework for the description of navigational strategies, expressed by mice while completing a spatial task.

Assessing the effects of the accumulation of learning on the performance in a spatial orientation task requires the combination of two elements. On the one hand, the complexity of the task should be high enough to allow for the expression of rich behavioral patterns and of different grades of information acquisition (Benjamini et al., 2011). Disentangling the different components informing animal choices requires providing animals with multiple options over a sizable spatial and temporal interval. Such availability is also a requirement in the interest of understanding animal behavior in its naturalistic setting (Tchernichovski and Benjamini, 1998). Wild rodents experience will include an articulate system of burrows together with the surrounding layout, a situation that can only be captured in the laboratory by studying spatial learning in larger, more complex environments (Wood et al., 2018).

As a consequence of the richer behavioral repertoire accessible to the animal, successfully tracking the evolution of task-related abilities requires the deployment of specific quantification tools, aimed not only at measuring task performance but also the specifics of animal behavior that accompany it (Dvorkin et al., 2008). Such tools should also provide a link between observable changes in the animal choice patterns and shifting navigational strategies underlying such choices (Ruediger et al., 2012; Gehring et al., 2015).

In this study, we use a novel spatial task, featuring a goal-localization paradigm with extended spatial and temporal dimensions, the HexMaze (Fig. 1, top). Mice learn to locate a reward location in a larger, modularly structured maze, providing precise control over animals' paths. Testing animals over a long temporal period, and after a modification of the environment as either we introduce a novel reward location or alternatively, we place a set of barriers to block some paths (Fig. 1), we look at the effects of previous knowledge on their performance and on their ability to flexibly incorporate novel information. We are thus able to track different contributions to the observed animal behavior, including those linked with information encoding, memory consolidation and cognitive map formation (for a complete description of learning dynamics on the maze see companion paper; Alonso et al., 2020a). To test the effects of such components, we apply statistical analysis to obtain a fully characterized picture describing the evolution of animal choices with sub-trial resolution. Importantly, such detailed phenomenological description of trial-by-trial behavioral profiles is then complemented with a generative model of task trajectories based on a mathematical description of task completion process (Fig. 1).

This modeling approach unveils a limited set of principles guiding animal choices. Our results show how animal behavior can be faithfully reproduced by a minimal mixture of random walking and goal-directed runs over limited distances. The relative importance of these two components over the different phases of learning (initial goal-location acquisition, consolidation over multiple sessions, cognitive map update in coincidence with environmental modifications) not only provides a concise characterization of animal approach to the task, but mirrors the emergence of task-specific memory constructs, offering a direct quantification of learning induced patterns of behavior. We find that, although we can observe an increasing amount of knowledge about the task and the maze structure being incorporated by animals, their performance never really converges to be completely goal-directed. Instead, we can measure the influence of a consistent random component, interfering with optimal task performance, and being only partially reduced with increasing familiarity to the task. Persistence of such task-independent activity could be a product of exposing animals to an expanded task complexity, effectively giving them increased freedom to divert from task completion, but it might also reflect mice-specific idiosyncratic behavior, possibly triggered by their propensity to hyperactivity (Jones et al., 2017). In both cases, further application of our modeling approach is likely to provide further insight on the diversity of behavioral approaches linked to different cognitive demands and different species.

Materials and Methods

Subjects

Five cohorts of four male C57BL/6J mice each (Charles River Laboratories) aged two months at arrival, were group-housed in the Translational Neuroscience Unit of the Centraal Dierenlaboratorium (CDL) at Radboud University Nijmegen, Netherlands. They were kept at a 12/12 h light/dark cycle and were before training food deprived overnight during the behavioral testing period. Weight was targeted to be at 90–85% of the animals' estimated free-feeding weight. All animal protocols were approved by the Centrale Commissie Dierproeven (CCD; protocol number 2016-014-018). The first cohort (coh 1) was used to establish general maze and task parameters and are not included in this dataset.

F.S. is supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement 840704 (BrownianReactivation), A. A. by the European Union's Horizon 2020 Research and Innovation Programme under the Marie-Sklodowska-Curie Grant M-Gate 765549.

*C.-A.V., A.A., L.G., and F.S. contributed equally to this work.

Correspondence should be addressed to Federico Stella at f.stella@donders.ru.nl.

https://doi.org/10.1523/ENEURO.0553-20.2021

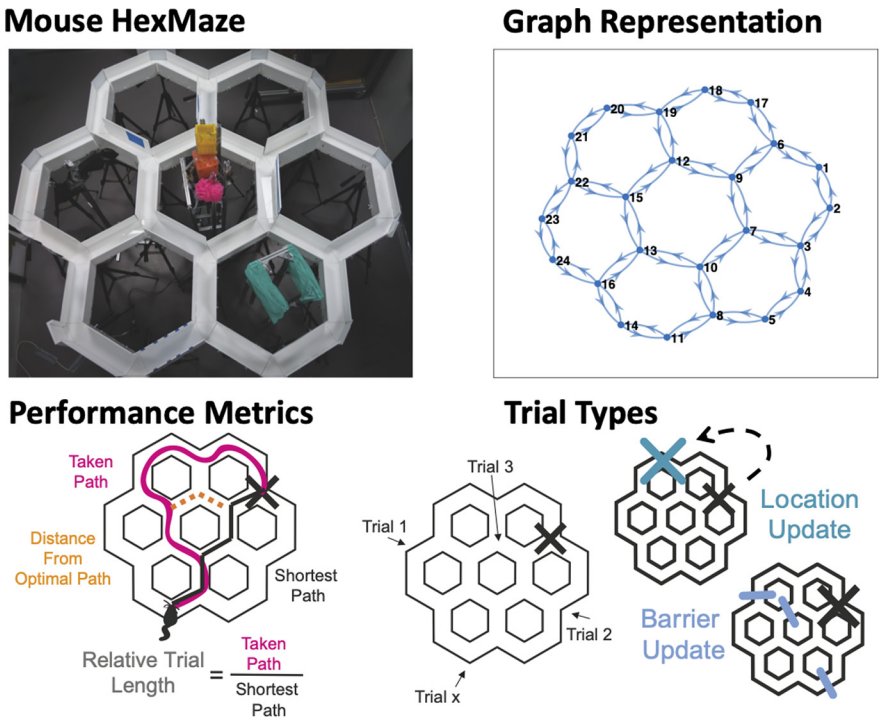


Figure 1. HexMaze structure and experimental paradigm. Top, View of the maze (left) and its graph representation used in the analysis (right). Bottom left, Two main performance metrics are used. (1) RTL is the length of the paths taken by the animal divided by the shortest possible path to the GL (indicated by the big X). (2) DFOP is the distance of the animal position at any time from the closest point of the shortest path. Bottom right, During training, animals started each trial from a different location and had to navigate to a fixed GL. After the animals had acquired the general maze knowledge during the build-up, updates were performed with inclusion of new barriers (barrier update) or new GLs (location update).

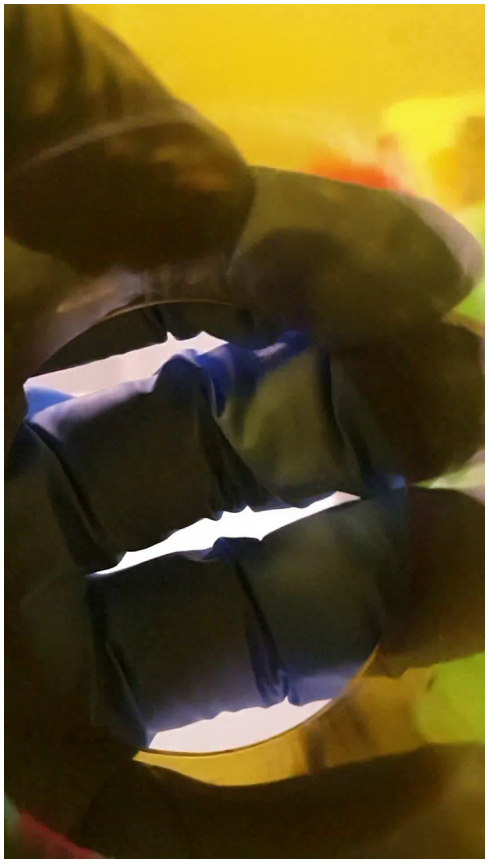
HexMaze

The HexMaze was assembled from 30 10-cm-wide opaque white acrylic gangways connected by 24 equilateral triangular intersection segments, resulting in 36.3-cm distance center-to-center between intersections (Fig. 1*A*). Gangways were enclosed by either 7.5- or 15-cm-tall white acrylic walls. Both local and global cues were applied to provide visual landmarks for navigation. Barriers consisted of transparent acrylic inserts tightly closing the space between walls and maze floor as well as clamped plates to prevent subjects bypassing barriers by climbing over the walls. The maze was held 70 cm above the floor to allow easy access by the experimenters.

Behavioral training

After arrival and before training initiation, mice were handled in the housing room daily for one week (until animals freely climbed on the experimenter) and then habituated to the maze in two 1-h sessions (all four cage mates together) with intermittent handling for maze pick-ups (tubing; Gouveia and Hurst, 2017). Mice were trained either on Mondays, Wednesdays, and Fridays (coh 1–3) or Tuesday and Thursday (coh 4 + 5). Per training day (session) each mouse underwent 30 min of training in the maze, resulting in up to 30 trials per session (Table 1). The maze was cleaned with 70% ethanol between animals (later clean wipes without alcohol to avoid damaging the acrylic), and to encourage returning in the next trial, a heap of food crumbles (Coco Pops, Kellogg's) was placed at a previously determined goal location (GL), which varied for each animal. GLs were counterbalanced across animals, as well as within animals across GL switches. e.g., one out of four animals, and one out of four GL per animal would be located on the inner ring of the maze while the others were on the outer ring (to shape animal behavior against circling behavior). Start locations for each day were generated based on their relation to the GL and previous start locations (locations did not repeat in subsequent trials, at least 60% of the trials had only one shortest path possible, first trial was different to the last and first trial of the previous session and locations had at least two choice points distance to each other as well as the GL). On average 30 start locations were needed per day per mouse, which were generated the day before training. After the mouse reached the food and ate a reward, the animal would be manually picked up with a tube, carried around the maze to disorient the mouse, and placed at the new start location. All pick-ups in the maze were done by tubing (Gouveia and Hurst, 2017). After placing the animal at the start location, the experimenter quickly but calmly moved behind a black curtain next to the maze to not be visible to the animal during training trials. An example of the view of the animal within the maze can be seen in Movie 1.

Training consisted of two blocks: build-up and updates. During probe sessions (each second session of a GL switch and additionally in build-up GL1: S6, GL2: S5,



Movie 1. Mimicking animal view in the maze. [View online]

GL3–GL5: S4), there was no food in the maze for the first and ninth trial of the day and each time for the first 60 s of the trial to ensure that olfactory cues did not facilitate navigation to the GL. After 60 s, food was placed in the GL, while the animal was in a different part of the maze (to avoid the animal seeing the placement). All other trials of the day were run with food at the GL. Probe trials and GLs switches were initially minimized, to help shape the animal behavior. In the first trial of the day, animals would not find food at the last presented location for both the first session of a new GL as well as probe trial days (e.g., always the second session of a new GL); thus, these sessions were interleaved with normal training sessions with food present at the last known location in the first trial of the day to avoid the animals learning the rule that food is initially not provided.

To measure the animals' performance, the actual path a mouse took was divided by the shortest possible path between a given start location and the GL, resulting in the log of normalized path length (Fig. 1*B*) and functioning as a score value. Given a sufficient food motivation and an established knowledge-network of the maze a mouse should navigate the maze efficiently. A score of 0 indicated that the mouse chose the shortest path and navigated directly to the goal. On average, animals would improve from a 3- to 1.5–2 times longer path length than the shortest path, corresponding to 0.4 and 0.2–3 log values. Random walks through the maze are estimated with a model to result in a four times longer path (0.6 in log). The normalized path length of any first trial of a session was used to measure long-term memory since training sessions were 2–3 d apart.

First trial of the second sessions (probe trials) of each GL in build-up and update phase were watched to score the number of times that animals crossed their current and previous GL as well as the amount of time they dwelled there. As a control, same method was applied to two other nodes, one on the inner ring and the other on the outer ring of the maze. These nodes were selected in such a way that they were not close to each other and to the GLs, with at least three gangways between them. Further, to control a false positive result, nodes that were in the way between GLs were not chosen as a control

Food motivation was ensured by restricting access to food for 12 to 24 h before training and confirmed by both the number of trials ran each day as well as the count of trials during which the animal ate food at the first encounter with the food in each trial. If animals were not sufficiently motivated, the count of both would decrease. Additionally, animals were weighted three times a week and the average weekly weight was ensured to not fall below estimated 85% free-feeding weight, which was adapted for the normal growth of each animal across time.

Behavior analysis

The structure of the HexMaze experimental setup was reproduced as a directed graph with node numbering corresponding to the experimental one. Animal trajectories were thus described as sequences of visited nodes on this graph (Fig. 1, top).

When measuring the distance of the animal location from one of the nodes in the optimal path, one has to consider the possible presence of multiple shortest paths of equal length connecting the start location with the goal. To take into account this source of ambiguity, for each trial, we computed the optimal path between the two locations using the HexMaze graph with weighted edges. The shortest path was computed multiple times each time on a different graph, first initialized with uniform weights and then adding small random noise to the value of every edge. In this way, in the presence of alternative and equivalent paths, the noisy weights would lead to the selection of either of the existing ones on a random basis. By collecting all the nodes happening to be described as belonging to a shortest path we thus obtain a list of all the nodes to be considered when computing the distance of the animal from the optimal path.

The experimental trials are divided as following: each condition (build-up, location update, or barrier update) comprises three sessions. For each session we analyze separately the first trial and then the following trials in groups of 10 until trial number 31.

Distance from optimal path (DFOP) curves fitting

We used the following function to parametrize the animal performance.

$$F = A * ((N(0, L1) - N(0, L2))/Z).$$

The difference of Gaussian functions (N) is normalized (Z) so that its maximum value is equal to. Therefore, A then controls the peak value of fit, while $L1$ and $L2$ its descending and ascending length, respectively. The fit is performed by optimizing the values of A, $L1$, and $L2$.

Mean and variance of the different measures for each condition were evaluated using 50-fold bootstrapping.

GL shuffling analysis

The specificity of the results for goal-directed behavior was tested by randomly assigning each trial with a random GL drawn from any of those used in the experiment. We generated surrogate data by randomly shuffling GLs across all trials from a particular trial block used in the analysis. Behavioral performance analyses were then repeated using the newly assigned GL as the target location for the optimal path and for the evaluation of the relative trial length (RTL). In cases in which the mice trajectory for a particular trial did not include the surrogate GL, we assigned by default an RTL of 10.

Simulations

The simulations are performed as following: we create a virtual HexMaze as a directed graph having the same structure of the real one. At every time step the virtual mouse moves from one node to an adjacent one. We do not allow trajectory reversals, so the node visited at the previous time step is not taken into account as a target. The start and GLs are the same as those used in the experiment. Each run consists then in a sequence of nodes visited by the mouse and the run eventually ends when the animal reaches the goal. We augment the size of the modelled data by simulating multiple independent runs ($n = 50$) for each experimental trial.

The movements of the virtual animal are generated according to an algorithm with two components: random search and the foresight. The random search part consists in a procedure to select which node the animal is going to visit next and is meant to approximate an optimal search strategy. While performing random search the animal randomly picks the next node among the available ones. On top of this, we introduce the possibility for the animal to take long diagonal runs that take it to another section of the maze. These diagonal runs are initiated with a probability η at any time step. If a diagonal run is initiated, then a node is randomly picked among those in the outer ring and at a distance of at least three steps from the current position of the animal. The mouse then uses the following time steps to reach this target along the shortest available path. Once the target is reached, the random node selection is resumed. We use different simulations to vary the value of η. Decreasing the value of this probability makes the search strategy approximate more and more a purely random walk through the environment. Higher values introduce a larger amount of "optimality" as they allow the animal to more quickly leave an already explored area.

The foresight component, on the other hand, represents the ability of the animal to anticipate the location of the goal when getting within a certain distance from it. It is, therefore, aimed at representing the effect of experience and an increasing knowledge of the environment and of visual cues. At every step, we draw a random number from an exponential distribution with mean F: $p(x) = \frac{1}{F} e^{-\frac{x}{F}}$. If the shortest path from the animal location to the goal node is smaller of this number, then the animal takes a direct path to the goal and the trial is over. Also, in this case, we run different sets of simulations varying the value of F. $F = 0$ corresponds to an animal with no ability to remember the position of the goal from its current location, unless by running directly over it. As F increases the chances for the simulated mouse to detect the goal from some distance increase. Eventually a very large F would reproduce a goal-directed behavior.

For each set of parameters, we measure how well the statistics of the simulated runs reproduce those obtained from real animal behavior. To do so we use a combination of measures: (1) RTL (as the ratio between the actual length of the trajectory and the length of the shortest path between start and GL); (2) maximal distance from the optimal path reached during the trial; (3) amount of time spent in the external ring of the maze versus the internal ring. For each of these quantities we compare the distribution obtained from the experiment to the one generated simulating the trajectory of the animal according to a specific set of parameters. We measure the distance between the two distributions with Kolmogorov–Smirnov (KS) statistics. Therefore, for each set of experimental trials we obtain get how well the statistical properties of the animal behavior can be reproduced by a certain choice of the model parameters.

Th3e experimental trials are divided as following: for each condition (build-up, location update, or barrier update) we used the first three session. During build-up, more sessions were run for each GL, however here we focus on the first three to be able to compare it with the update phase. For each session we analyze separately the first trial and then the following trials in groups of 10 until trial number 31.

Statistical analysis

Sample size of the data available and used for each behavioral condition is reported in Table 1. Simulated data samples correspond to the same numbers multiplied by 50.

Throughout the paper, bootstrap estimates of mean and SD of different measured quantities are obtained from $n = 50$ resampling with replacement.

Whenever mentioned, KS test is meant to be two-samples. Reported non-significant differences were all associated with a $p > 0.1$ and a size effect < 0.08. Minimal $n \times m/(n + m)$ ratio was equal to 50.

Code accessibility

The code/software described in the paper is freely available online at https://github.com/fstella/HexMaze_BehaviorAnalysis and also as the Extended Data 1. All analysis

Table 1: Number of trials used in the analysis for each condition

	Session 1				Session 2				Session 3			
	Trial 1	Trials 2–11	Trials 12–21	Trials 22–31	Trial 1	Trials 2–11	Trials 12–21	Trials 22–31	Trial 1	Trials 2–11	Trials 12–21	Trials 22–31
Build-up	75	749	571	260	65	640	555	325	50	494	426	219
Location update	76	761	633	231	64	638	564	302	52	517	400	176
Barrier update	76	760	601	245	64	640	581	355	52	519	449	229

Number of simulated runs correspond to the same amounts multiplied by $n = 50$.

and simulations were performed using custom MATLAB code.

Results

The HexMaze experiment

The HexMaze is arranged as six regular densely packed hexagons, forming twelve two-way and twelve three-way choice points (nodes) 36.3 cm apart, in total spanning 2 × 1.9 m (Fig. 1, top). Gangways between nodes were 10 cm wide and flanked by either 7.5- or 15-cm-tall walls. Maze floor and walls were white and opaque, with local and global cues applied in and around the maze to enable easy spatial differentiation and good spatial orientation; overall leading to a complex, integrated maze. During training food was placed in one of the nodes and the animal had to learn to navigate efficiently from different start locations to the Goal Location (GL).

Animals went through two phases of training: build-up and updates. In the build-up the animals should create a cognitive map of the maze environment; in contrast, during updates, stable performance is achieved and they should be simply updating the cognitive map. These two phases also differed in the frequency of GL switches: during build-up, the GL remained stable for five or more sessions, while during updates a change occurred every three sessions (see also below). For more detailed analysis of the different phases of learning in this task, please see companion paper (Alonso et al., 2020a). Different update types were performed: including barriers in the environment (barrier update) and changing the GL (location update; Fig. 1, bottom). Number of trials used in the following analysis are reported in Table 1.

Characterization of animal behavior in the HexMaze

We first set out to quantify the time-evolution of animal behavior. The structure of the HexMaze allows for an efficient tracking of the animal choices, as its behavior is easily described by the sequence of visited maze nodes. Therefore, in the following we will focus on this descriptor to measure different aspects of the animal performance during the experiment. Clearly, the main element to be taken into account when analyzing behavior is the ability of the animal to efficiently localize the reward and reach it through a path as short as possible (Fig. 1).

A perfectly optimal, goal-oriented behavior, would imply that after a necessary learning transient the trajectories selected by the animal would progressively converge toward the shortest available given the start location and the reward one. By measuring the ratio between the length of the actual path (measured in terms of number of nodes visited before reaching the goal) and the length of the optimal path (Relative Trial Length, RTL), over a certain amount of trials, one would then expect to observe this distribution to be more and more skewed toward the value of 1, corresponding to the animal actually following the optimal path (Fig. 2). Indeed, what we find is a progressive increase in the percentage of trials with a low RTL score (Fig. 2, last column), within each session and across sessions. Multiple effects point to an actual presence of learning and to a growing awareness of the maze structure and GL in the animals. First, during build-up not only the score improves within each session (Welch's t test first session $p = 10e\text{-}90$, second $p = 10e\text{-}76$, third $p = 10e\text{-}68$), but animals consistently do better in the first trial of a new session compared with the previous one (Welch's t test first trial session 1 vs 2 $p = 10e\text{-}21$, session 2 vs 3 $p = 10e\text{-}12$). Then, while the score goes back to pre-learning values at the beginning of the location update, when a new unknown goal is introduced, it reaches its asymptotic value faster compared with build-up (Welch's t test first trial second session build-up vs location update, $p = 0.03$). And finally, the insertion of barriers has only a very limited effect on the animal performance. Furthermore, we tested the actual presence of goal-directed behavior and the specificity of the results for the GL by randomly reassigning the target node across all trials from the same trial block and recomputing the RTL (Fig. 2, bottom raw). Results obtained from the surrogate data show no effect of learning and no difference from un-directed behavior thus indicating the presence of genuine task-related learning. At the same time, the hypothesis of over-wise, totally committed mice is challenged by the fact that although increasing in time, the probability of a perfect (or almost perfect) run remains substantially below 1 over the entire arc of the experiment, even after the animals have been repeatedly exposed to the maze and to a specific reward location. The trial relative length distribution shows a long tail of values larger than 1 (Fig. 2, columns 1–3), indicating that the animal choices are far from corresponding to a purely optimal, goal-oriented strategy.

One way to further characterize the degree to which the mouse behavior is goal-directed is to measure how far during the trial, its trajectory would steer away from the optimal path joining the starting location to the reward. We thus look to expand the trial level RTL score into a description of the within-trial structure of behavior, looking at the specific sequence of choices made by the animal. For each trial and for each node visited by the animal during the trial, we then compute the Distance from optimal

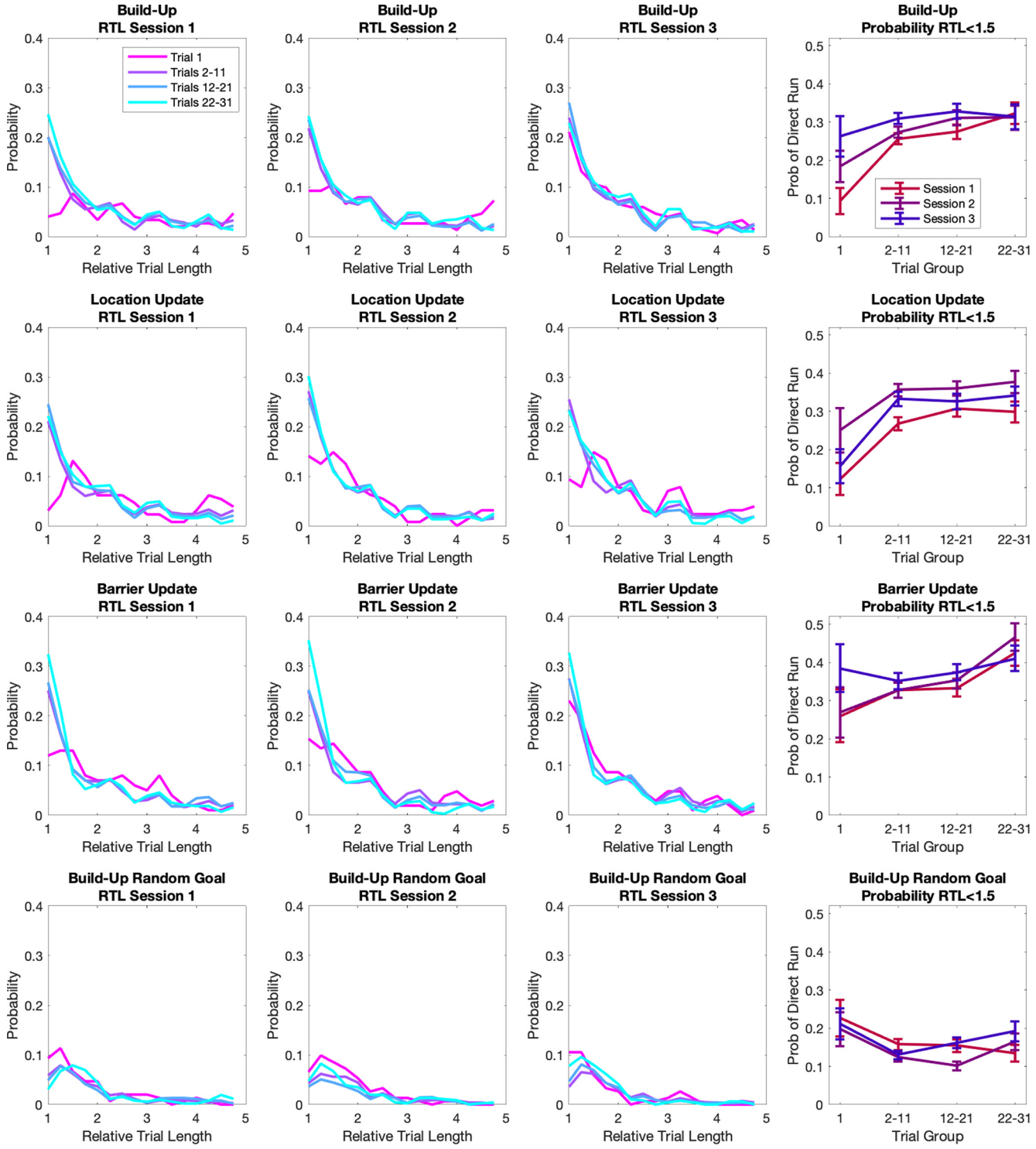


Figure 2. Animals are progressively more likely to take shorter paths to the goal. Distribution of Relative trial length (RTL) for different trial groups and sessions. For each session, trials from different animals were grouped in four categories: first trial only, from second to 11th trial, from 12th to 21st trial, and from 22nd to 31st trial. RTL = 1 corresponds to perfect trial. Last column, Probability of RTL < 1.5 (optimal trial) over time. Last row, RTL computed after randomly shuffling the Goal Locations across trials. The absence of learning-induced changes indicates their specificity for goal-directed behavior. Error bars show STD computed by 50 bootstraps.

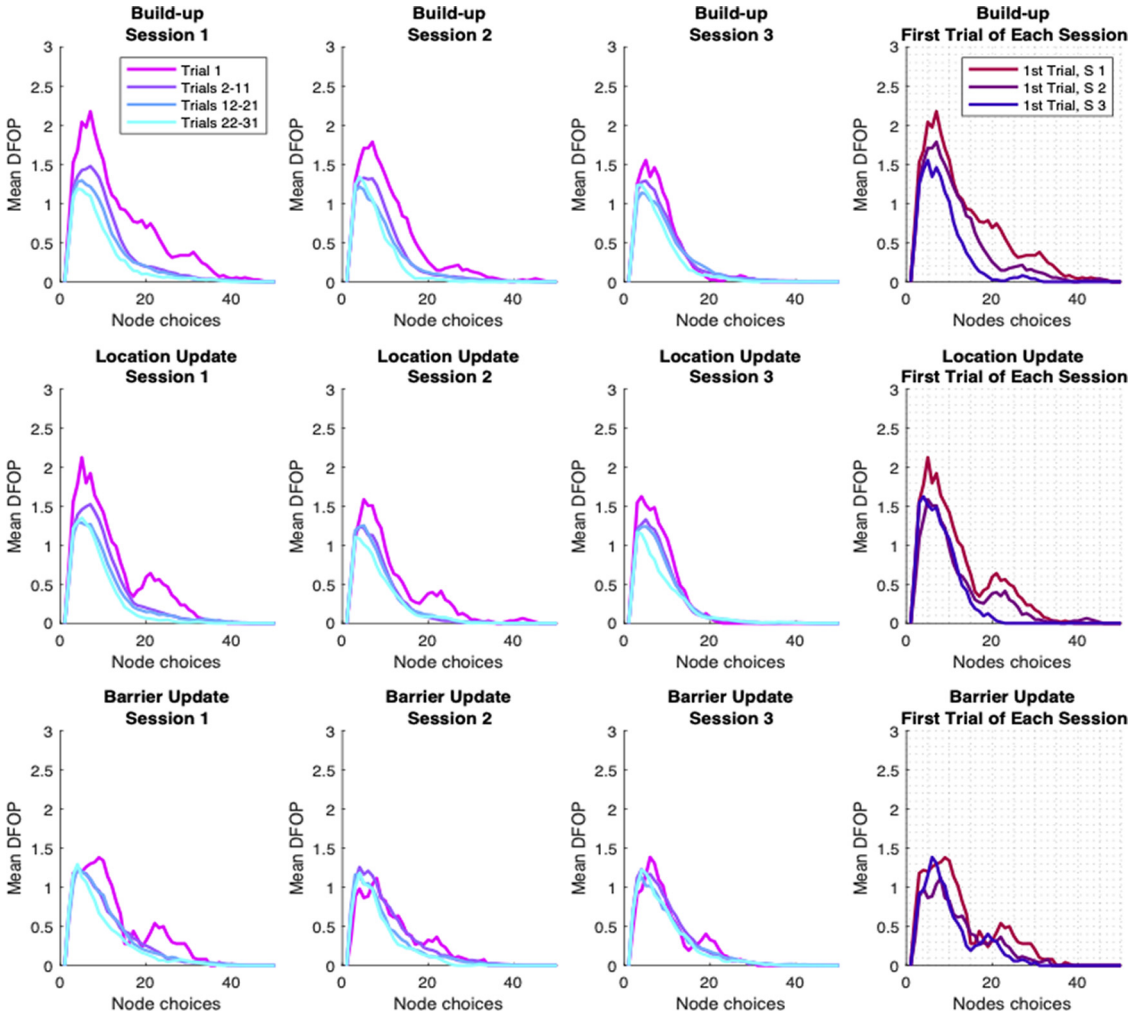


Figure 3. Quantification of animal trajectories departure from optimal path. Distance from optimal path (DFOP) over time for all trial groups and sessions. Last column panels, Comparison of the first trial for the three sessions. The effects of learning can be seen in the progressive reduction of the distance within each session. Additionally, DFOP decreases on the first trial of every successive session. Once a new Goal Location is introduced, convergence to asymptotic performance is faster than during initial learning. Finally, the insertion of barriers has only limited effects on behavior.

path (DFOP), that is, the distance between the visited node and any of the nodes comprising the optimal path. For each trial we then obtain a measure of "stray" over time, providing a profile of the animal approach to the goal (Fig. 3). When averaged over different set of trials, this profile shows a bump shape, quantifying the amount of deviation from optimal behavior, showing an increase at the beginning of the trial and eventually converging again toward the correct path. These curves provide us with different information about the animal trajectories over the course of the experiment: (1) the amount of "stray," that is the average maximum distance from the optimal path; (2) the average length of the trials; and (3) how fast the animal will go back to the correct path after straying away in the beginning. We can extract this information from the data by fitting a parametrized function to the DFOP average profiles. We use a normalized difference of Gaussians (see Materials and Methods) that provides us an excellent approximation of the experimental curve shapes (Fig. 4, top row). This fit depends on 3 parameters: (1) the maximum height; (2) the amount of steps to reach the maximum; and (3) the amount of steps to go back to the optimal path. Separately plotting the value of these three parameters, obtained by fitting the data from different stages of learning, we can identify the different components contributing to the overall change in performance (Fig. 4). In fact, we observe how the maximum height and the descending scale show a gradual decrease with time, consistently with the improving performance of the animal (Fig. 4, first and second column; build-up within session decrease maximum height: Welch's t test first session $p = 10e\text{-}51$, second $p = 10e\text{-}29$, third $p = 10e\text{-}28$, descending scale: Welch's t test first session $p = 10e\text{-}40$, second $p = 10e43$, third n.s. Build up first trial decrease maximum height: Welch's t test first vs second session $p = 0.025$, second vs third $p = 10e\text{-}8$, descending scale: Welch's t test first vs second session $p = 10e\text{-}20$, second vs third $p = 10e\text{-}31$. Location update second session first

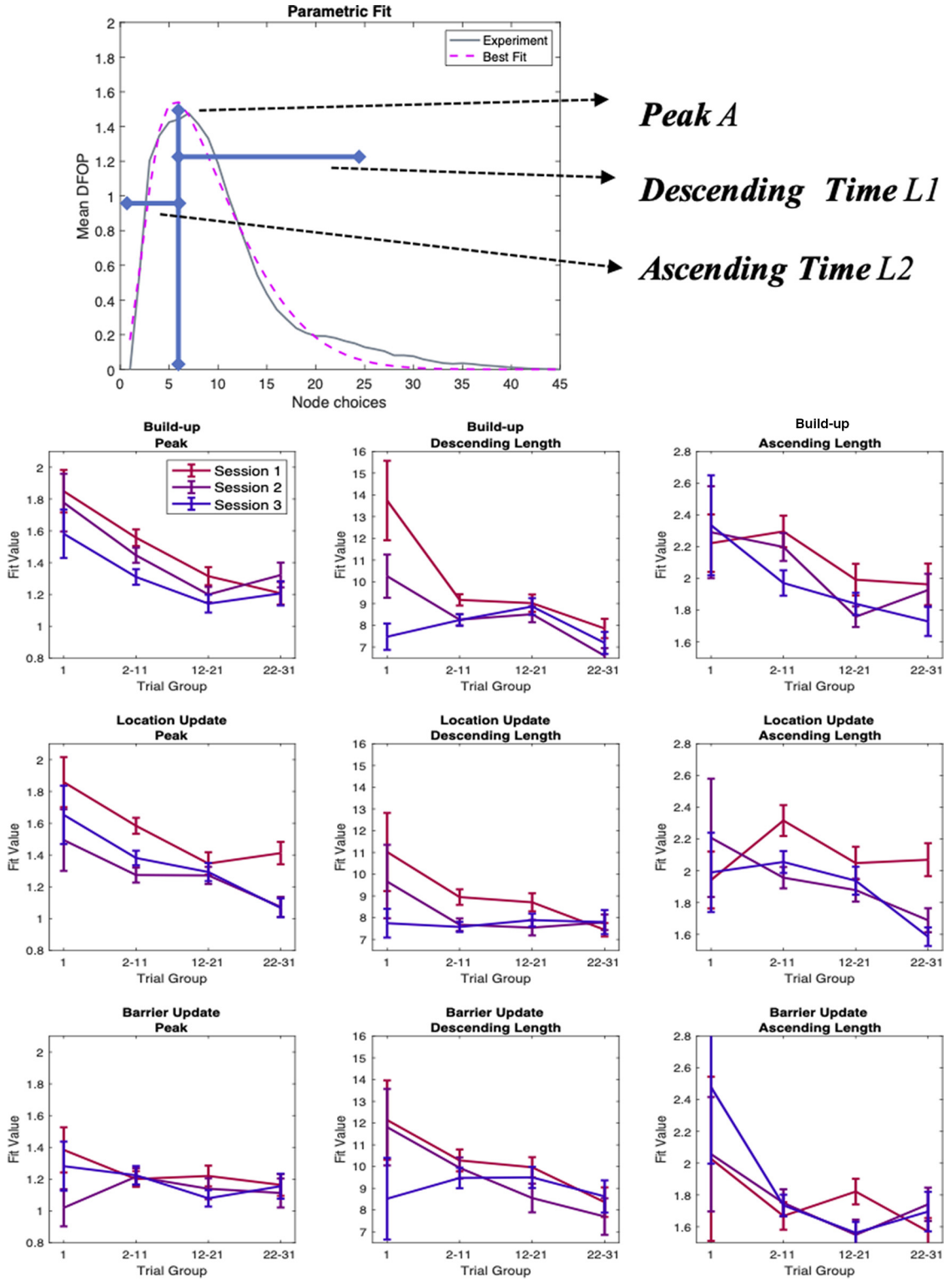


Figure 4. Learning sharpens animal performance by progressively reducing trajectory Distance from optimal path (DFOP). Results of the parametric fit of the curves in Figure 3. Top row, Example of the obtained match between experimental curves and the parametric fit. Bottom, value of the fit parameters over time for all conditions and sessions. Error bars STD from bootstrapping. This fit allows us to quantify: (1) the amount of "stray," that is, the average maximum distance from the optimal path; (2) the average length of the trials; and (3) how fast the animal will go back to the correct path after straying away in the beginning. Maximum distance and descending length show a decreasing modulation over time: within one session, across sessions, and during Goal Location shift, consistently with learning effects. Descending length shows instead no significant improvement, in line with a persistent influence of a random component on behavior.

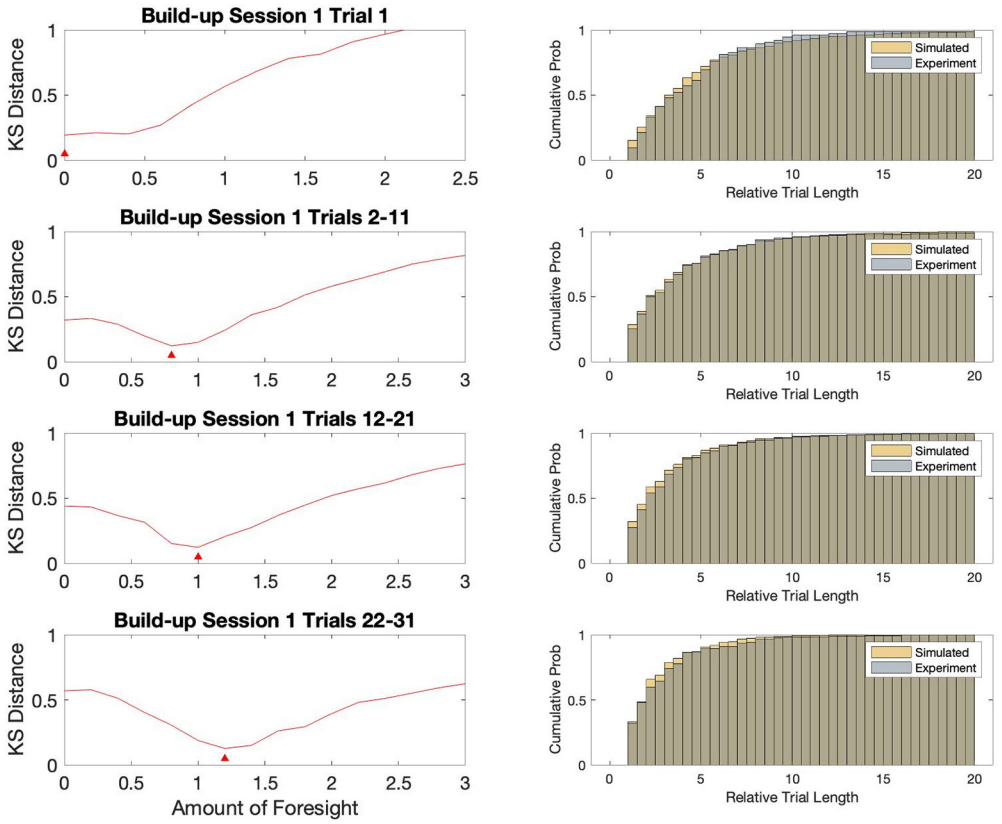


Figure 5. A minimal mathematical model reproduces the main properties of animal behavior. Model fitting to experimental data. Left column, Simulated-experimental Kolmogorov-Smirnov (KS) distance for a range of tested foresight values. Red triangles indicate location of best-fit *F* value for different sets of trials. Right column, Comparison of cumulative distributions for different trial groups and corresponding simulation results with best *F* value. All shown data are from build-up phase.

trial build-up vs location update, maximum height: Welch's *t* test $p = 0.033$, descending scale: Welch's *t* test $p = 10e\text{-}11$). Again, also these two parameters show an overall trend across all the phases of the experiment although their evolution is not monotonous, but rather has a seesaw shape because of the partial rollbacks happening between the last trials of one session and the first trial of the following one. At the same time, the ascending phase is not significantly affected by learning, showing that while the animals progressively strayed less and eventually took more direct runs to the goal, they nevertheless maintained a comparable amount of undirected behavior in the first part of the trial. Together with the previous analysis, our quantification of the animal behavior, shows how the navigation of mice in the HexMaze, can be described as a combination of learning-based choices (evident in the progressive improvement in all goal-related metrics) and of a persistent non-optimal component, keeping the overall behavior away from perfect performance even for late sessions and trials.

Minimal mathematical model describing animal choices

We then asked whether such results could be reproduced by a simplified model of the animal behavior. We simulate the trajectories produced by a virtual agent navigating the same HexMaze used in the experiment as it searches for the reward location. In these simulations the mouse moves in the environment selecting the next node to visit according to a set of predefined rules. We define these rules as a combination of a random walk through the environment and direct goal-runs based on the knowledge of the reward location. Crucially, the model depends on only two parameters, η, the probability of taking a long diagonal run while randomly moving through the environment; and F, determining the probability that the animal will at any time start to run directly toward the reward location (a quantity that for this reason, we named foresight). We thus simulate the animal behavior using different combinations of these two parameters and compare the obtained statistics with those collected during the real experiment. The comparison is based on quantifying the distance between the distribution of RTLs in real and virtual trajectories.

We find that our simple behavioral models very accurately approximate the animal strategy in every part of learning. For each set of trials, we find a combination of η and *F* that makes the distributions not significantly different (all $p > 0.1$, all effect sizes < 0.08, as evaluated using two samples KS statistics; Fig. 5).

We thus consider the (η, *F*) pair that minimizes the distance between experimental and simulation statistics for a specific set of trials (Fig. 5). This pair of values is taken

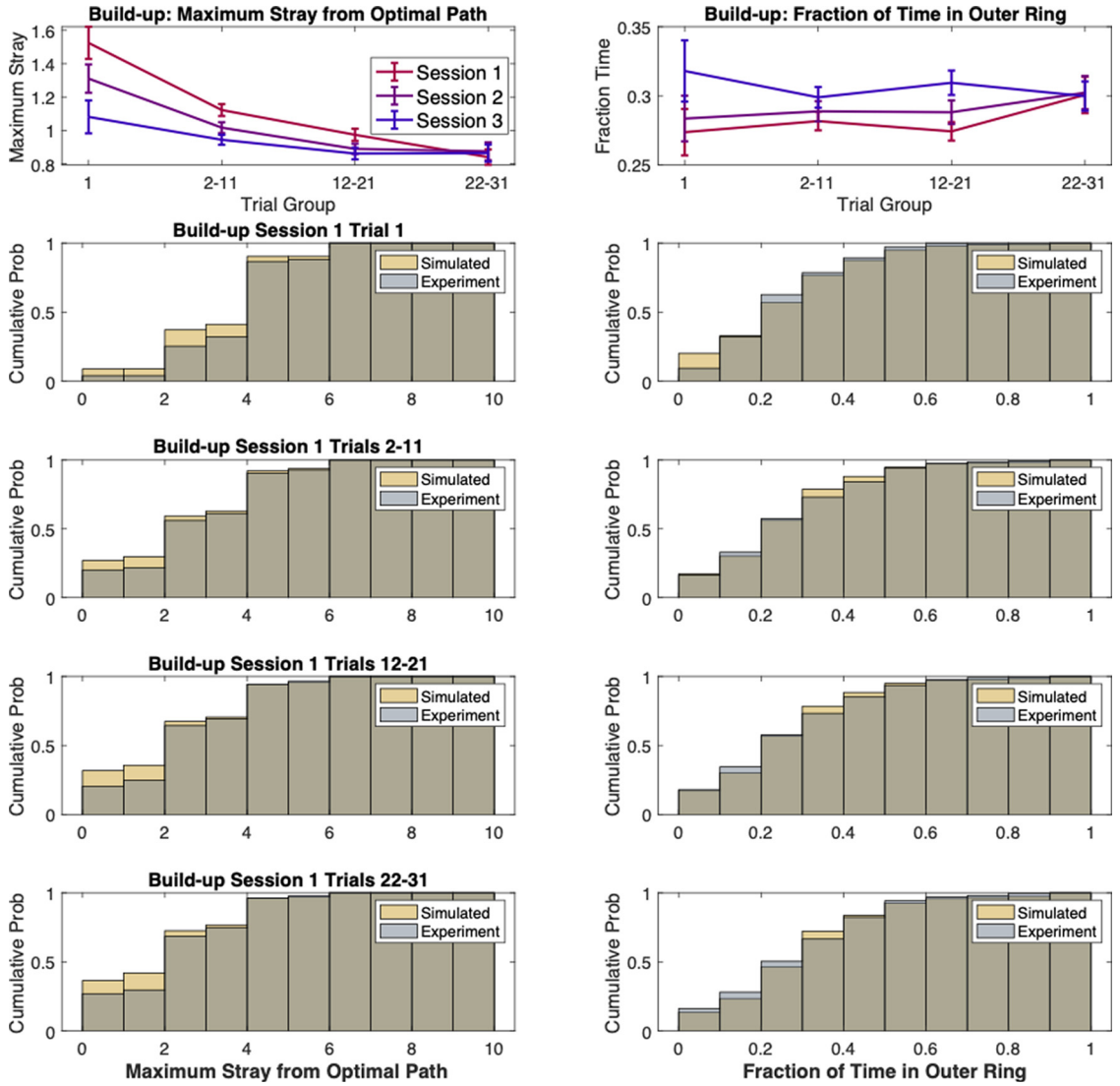


Figure 6. Further behavioral features matched by model fits. Top row, Evolution of the maximal DFOP and the fraction of time spent in the maze outer ring for the build-up phase. Rows 2–5, Same data as Figure 5. Using the same parameters obtained from fitting RTL distributions, the model also reproduces other aspects of animal behavior such as the distribution of maximal distance from the optimal path (left) and the distribution of relative time spent in the outer or inner ring of the maze on each trial (right). Real and model-based distributions are non-significantly different for every experimental phase and trial group (KS, all $p > 0.1$).

as best describing the behavioral characteristics of the animal navigation through the maze. The evolution of these values in time provides us with a measure of the effects of learning on animal's behavior. Moreover, the model relevance is corroborated by finding that the simulated trajectories obtained with parameters optimized to fit the trial length distribution also reproduced other statistical features of the animal behavior. In fact, both the distribution of maximal distance from the optimal path, and that of average time spent on the inner versus outer ring of the maze were captured by our simulations (comparison between behavior and model: two samples KS all $p > 0.1$, all effect sizes < 0.07; Fig. 6).

Studying the time evolution of the model parameter we first determine that the best value of η is not significantly affected by the progression of learning, and that it remains confined to 0 across the entire experiment. Therefore, learning does not change the properties of random movement across the maze, and indeed, this movement pattern appears to be largely unstructured, being captured by a simple sequence of random turns. On the other hand, the effects of learning are instead reflected in the evolution of the best value for foresight (Fig. 7). As shown in the figure, its value starts at 0 for the first trial of the build-up phase, compatibly with an animal with no knowledge of the reward location and only randomly moving across the maze. F then progressively increases with the accumulation of trials, indicating a growing awareness for the location of the reward, its relationship to visual cues and possibly for the geometrical structure of the maze itself. Interestingly, foresight increase is significant

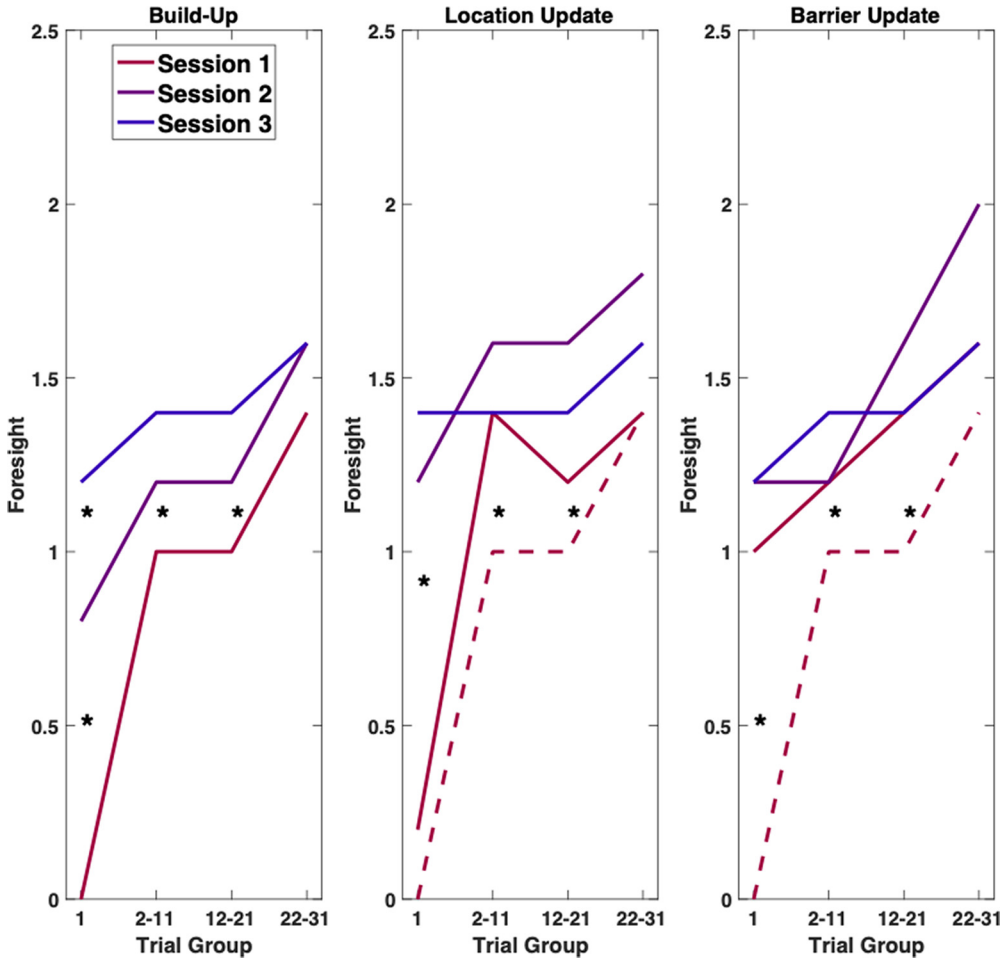


Figure 7. Modelling of animal behavior shows the accumulation of spatial information over the course of the experiment. Foresight evolution: best-fit values of foresight for different experimental phases, sessions and trial groups. Our model reproduces the different phases of learning identified from behavioral analysis. The foresight quantity appears to increase over the course of a session and with the accumulation of sessions. GL change is followed by a return to prelearning values, but successive increase in faster than during initial task learning (here shown as for reference with a dashed line). Also, in terms of inferred navigational strategy, the insertion of barriers in the maze has only limited effects.

(KS $p < 0.05$ comparing model distributions) both across trials within one session (single lines in the plot) and across the first trial for each session, indicating a non-monotonic increase in performance (as we already found while analyzing the animal trajectories), reflecting a drop in performance between the end of one session and the start of the next one. What sort of conclusions can be drawn from the model results about the animal behavior in the maze? Even at late stages of the build-up phase, the foresight value remains relatively low, never raising above a value of 2. This limit points to a significant presence of random walking even for mice that have completed a substantial number of trials. They appear to initiate goal-directed runs only when in close proximity to the reward and only rarely from the very beginning of the trial.

Should the persistence of random behavior be taken as proof of a failure of the animal to build a complete "cognitive-map" of the maze? We can partially address this issue by looking at the performance of the model in different experimental conditions. Taking the location update phase, we see how foresight is again close to 0 for the very first trial, consistently with the presence of a novel reward location. Nevertheless, in the following trials, the value of F increases at a significant faster rate compared with the build-up phase (KS $p < 0.05$). The mice are thus able to quickly integrate the new information into the knowledge they accumulated in the previous build-up sessions. This effect is not limited to the first update session but can be seen in a rapid saturation of F in the following session, although again its value does not grow beyond 2.

Similarly, using barrier update sessions, when barriers are added to the maze, while maintaining the reward locations stable, the behavior of the animal appears to be at the same level of the build-up trials from the very beginning of this phase. Improvement in the goal-directed behavior can be still seen within one session, but no significant difference can be observed between sessions. The relatively low impact of barrier

introduction is consistent with animal strategy being dominated by a random walk and only affected by the presence of the goal when in the proximity of it.

Discussion

The HexMaze provides the ideal setting to characterize animal spatial cognition, as it combines the availability of options to express complex behavior, with the possibility to precisely monitor and quantify animal navigational choices (Alonso et al., 2020a). In particular, in this set of experiments we leverage on its structure to study the development of goal-directed behavior in mice learning to localize a reward location. We first provide a statistical characterization of the animal navigational patterns as they traverse the maze, by comparing them to the path expected from optimal goal-oriented behavior. Our results show a clear effect of learning in mice, as they progressively tune their trajectories to reach the reward in a shorter time and visiting less nodes on the maze. We find that their trajectories are less likely to stray away from the optimal path, and that eventual detours are shorter lasting. Indeed, we can show that all of the trajectory quantifiers evolve according to a superposition of different time courses: (1) they steadily improve on the first trial of each session; (2) their improvement over the course of a session becomes faster in later sessions; (3) the rate of improvement is enhanced after a novel reward location is introduced during the update phase (GL switches after 12 weeks of learning) in contrast to the build-up phase (initial GL switches); (4) introducing path-blocking barriers in the maze has only a very limited effect, when the reward location is already familiar to the animal. Such sharpening of trajectories over the course of the experimental paradigm is in direct agreement with the presence of different forms of previous knowledge, all contributing to enhance the animal performance in the task (Gire et al., 2016). The emergence of an allocentric representation of the maze, the linking of specific cues to the proximity of the GL, the strengthening of memory encoding and consolidation, can be all considered to affect the measured properties of navigational patterns. For a more detailed discussion on different potential previous knowledge effects in the task, see companion paper (Alonso et al., 2020a).

At the same time, our measures also bring to the foreground how mice behavior remains significantly distant from a purely optimal one. Mice never develop a completely goal-oriented pattern of movements, as a consistent part of their choices on the maze appears to be independent of the GL. To identify the nature of this layer of non-optimal behavior, we develop a computational model of HexMaze navigation producing virtual animal trajectories based on specific generative principles. This minimal mathematical model indicates how two components are sufficient to reproduce, with little parameter tuning, most of the statistical properties of mice real behavior. As expected from the previous results, we find that the first of these components is an increasing propensity to directly run toward the GL, whose effects are felt further and further away from the goal with the accumulation of learning. Goal-directed runs stemming from this component are however combined with a basis of purely random choices that remains present throughout the experimental paradigm (Thompson et al., 2018).

One could find rather surprising this persistent neglect of the task requirements, leading to mice spending a considerable amount of time exploring portions of the maze distant from the goal, even when other behavioral measures indicate their awareness of its actual location. Indeed, the likelihood of direct runs does not seem to be completely determined by the amount of knowledge about the reward location, which appears to be available to the animal long before the end of the build-up phase. Instead, the steady and progressive increase of goal-directed behavior can be attributed to a shift in the animal approach to the task, giving more and more saliency to the known reward location. This shift does not happen abruptly, neatly separating a random walk phase from a goal-oriented one, but instead is embedded in a mixture of the two behaviors, appearing together on trial-by-trial basis. It is possible that this tendency to persist in random-like behavior, could have been hidden in other experimental paradigms by the lack of options available to the animals, as they were given very little possibilities to show behavior not related to the task. In the context of a larger spatial arena the balance between "exploratory" and "exploitative" behavior (Jackson et al., 2020; Wilson et al., 2021) might shift sensibly toward the former, leading to an increase in undirected behavior. We are also aware that this pattern of behavior could be specific to mice. In fact, random exploration might be a consequence of mice hyperactivity (Jones et al., 2017) and their reluctance to consistently focus on the accomplishment of a specific task. A tendency that in the present case could be further exacerbated by the absence of sheltered locations in the maze, an element that has been shown to be of great relevance for these animals' sense of security. With this in mind, it is not too far out to expect rats to perform very differently in this same experimental setting, which is currently under investigation. In a previous study, Jones et al. (2017) could show that rats and mice differed in levels of baseline activity measured as shuttle rate during inter-trial intervals; mice shuttled two to three times as frequently as rats. Species differences in behavioral ecology may underlie this difference, as for example mice needing to move rapidly when outside burrows to minimize predation risk. They tend to use bursts of speed to run from a more sheltered position to the next. Applying our modeling approach to another experimental condition will allow us to extend it to include novel behavioral components, and might open the way to turn the current descriptive approach into a predictive one, producing different strategy combinations depending on the external context.

Regardless of the final explanation of this finding, we acknowledge how these results have significant consequences for the interpretation of the neural correlates of goal-directed navigation. Place cell activity in the rodent hippocampus has been shown to organize in sequential "sweeps" linking the current location of the animal to that of one or more GLs, when the animal is asked to take a decision (Wikenheiser and Redish, 2015). Similarly, it has

been proposed that sequences of activated place cells during a sharp wave ripple bear information about future navigational paths (Pfeiffer and Foster, 2013). Our results, demonstrating the co-existence of goal-oriented behavior with a substantial amount of random choices when testing mice in a more naturalistic experimental setting, suggests that such neural episodes might be circumscribed both in time and space, and might play a more limited role when considering animals with a richer set of behavioral options.

References

Alonso A, Bokeria L, van der Meij J, Samanta A, Eichler R, Spooner P, Lobato IN, Genzel L (2020a) The HexMaze: a previous knowledge and schema task for mice. bioRxiv 441048.

Alonso A, van der Meij J, Tse D, Genzel L (2020b) Naïve to expert: considering the role of previous knowledge in memory. Brain Neurosci Adv 4:2398212820948686.

Behrens T, Muller TH, Whittington J, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson Z (2018) What is a cognitive map? Organizing knowledge for flexible behavior. Neuron 100:490–509.

Benjamini Y, Fonio E, Galili T, Havkin GZ, Golani I (2011) Quantifying the buildup in extent and complexity of free exploration in mice. Proc Natl Acad Sci USA 108 [Suppl 3]:15580–15587.

Dvorkin A, Benjamini Y, Golani I (2008) Mouse cognition-related behavior in the open-field: emergence of places of attraction. Plos Comput Biol 4:e1000027.

Fonio E, Benjamini Y, Golani I (2009) Freedom of movement and the stability of its unfolding in free exploration of mice. Proc Natl Acad Sci USA 106:21335–21340.

Gehring TV, Luksys G, Sandi C, Vasilaki E (2015) Detailed classification of swimming paths in the Morris water maze: multiple strategies within one trial. Sci Rep 5:14562.

Gire DH, Kapoor V, Arrighi-Allisan A, Seminara A, Murthy VN (2016) Mice develop efficient strategies for foraging and navigation using complex natural stimuli. Curr Biol 26:1261–1273.

Gouveia K, Hurst JL (2017) Optimising reliability of mouse performance in behavioural testing: the major role of non-aversive handling. Sci Rep-uk 7:44999.

Jackson BJ, Fatima GL, Oh S, Gire DH (2020) Many paths to the same goal: balancing exploration and exploitation during probabilistic route planning. eNeuro 7:ENEURO.0536-19.2020.

Jones S, Paul ES, Dayan P, Robinson ESJ, Mendl M (2017) Pavlovian influences on learning differ between rats and mice in a counter-balanced Go/NoGo judgement bias task. Behav Brain Res 331:214–224.

Pfeiffer BE, Foster DJ (2013) Hippocampal place-cell sequences depict future paths to remembered goals. Nature 497:74–79.

Richards BA, Xia F, Santoro A, Husse J, Woodin MA, Josselyn SA, Frankland PW (2014) Patterns across multiple memories are identified over time. Nat Neurosci 17:981–986.

Ruediger S, Spirig D, Donato F, Caroni P (2012) Goal-oriented searching mediated by ventral hippocampus early in trial-and-error learning. Nat Neurosci 15:1563–1571.

Tchernichovski O, Benjamini Y (1998) The dynamics of long term exploration in the rat. Biol Cybern 78:433–440.

Thompson SM, Berkowitz LE, Clark BJ (2018) Behavioral and neural subsystems of rodent exploration. Learn Motiv 61:3–15.

Wang SH, Morris R (2010) Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. Annu Rev Psychol 61:49–79.

Wikenheiser AM, Redish AD (2015) Hippocampal theta sequences reflect current goals. Nat Neurosci 18:289–294.

Wilson RC, Bonawitz E, Costa VD, Ebitz RB (2021) Balancing exploration and exploitation with information and randomization. Curr Opin Behav Sci 38:49–56.

Wood RA, Bauza M, Krupic J, Burton S, Delekate A, Chan D, O'Keefe J (2018) The honeycomb maze provides a novel test to study hippocampal-dependent spatial navigation. Nature 554:102–105.