

# Mice in a labyrinth: Rapid learning, sudden insight, and efficient exploration

Matthew Rosenberg<sup>1†</sup>, Tony Zhang<sup>1†</sup>, Pietro Perona<sup>2</sup>, Markus Meister<sup>1\*</sup>

\*For correspondence:

[mhrosenberg@caltech.edu](mailto:mhrosenberg@caltech.edu) (MR); [tonyzhang@caltech.edu](mailto:tonyzhang@caltech.edu) (TZ); [perona@caltech.edu](mailto:perona@caltech.edu) (PP); [meister@caltech.edu](mailto:meister@caltech.edu) (MM)

<sup>1</sup>Division of Biology and Biological Engineering, Caltech, USA; <sup>2</sup>Division of Engineering and Applied Science, Caltech, USA

†These authors contributed equally to this work

**Abstract** Animals learn certain complex tasks remarkably fast, sometimes after a single experience. What behavioral algorithms support this efficiency? Many contemporary studies based on two-alternative-forced-choice (2AFC) tasks observe only slow or incomplete learning. As an alternative, we study the unconstrained behavior of mice in a complex labyrinth and measure the dynamics of learning and the behaviors that enable it. A mouse in the labyrinth makes ~2000 navigation decisions per hour. The animal quickly discovers the location of a reward in the maze and executes correct 10-bit choices after only 10 reward experiences – a learning rate 1000-fold higher than in 2AFC experiments. Many mice improve discontinuously from one minute to the next, suggesting moments of sudden insight about the structure of the labyrinth. The underlying search algorithm does not require a global memory of places visited and is largely explained by purely local turning rules.

## Introduction

How can animals or machines acquire the ability for complex behaviors from one or a few experiences? Canonical examples include language learning in children, where new words are learned after just a few instances of their use, or learning to balance a bicycle, where humans progress from complete incompetence to near perfection after crashing once or a few times. Clearly such rapid acquisition of new associations or of new motor skills can confer enormous survival advantages.

In laboratory studies, one prominent instance of one-shot learning is the Bruce effect (*Bruce, 1959*). Here the female mouse forms an olfactory memory of her mating partner that allows her to terminate the pregnancy if she encounters another male that threatens infanticide. Another form of rapid learning accessible to laboratory experiments is fear conditioning, where a formerly innocuous stimulus gets associated with a painful experience, leading to subsequent avoidance of the stimulus (*Fanselow and Bolles, 1979; Bourchuladze et al., 1994*). These learning systems appear designed for special purposes, they perform very specific associations, and govern binary behavioral decisions. They are likely implemented by specialized brain circuits, and indeed great progress has been made in localizing these operations to the accessory olfactory bulb (*Brennan and Keverne, 1997*) and the cortical amygdala (*LeDoux, 2000*).

In the attempt to identify more generalizable mechanisms of learning and decision making, one route has been to train laboratory animals on abstract tasks with tightly specified sensory inputs that are linked to motor outputs via arbitrary contingency rules. Canonical examples are a monkey reporting motion in a visual stimulus by saccading its eyes (*Newsome and Pare, 1988*), and a mouse in a box classifying stimuli by moving its forelimbs or the tongue (*Burgess et al., 2017; Guo et al., 2014*). The tasks are of low complexity, typically a 1 bit decision based

42 on 1 or 2 bits of input. Remarkably they are learned exceedingly slowly: A mouse typically  
43 requires many weeks of shaping and thousands of trials to reach asymptotic performance; a  
44 monkey may require many months.

45 What is needed therefore is a rodent behavior that has the hallmark of complex decision  
46 making, with many input variables and many possible choices. Ideally the animals would  
47 perform this task without excessive intervention by human shaping, so we may be confident that  
48 they employ innate brain mechanisms rather than circuits created by the training. Obviously  
49 the behavior should be compatible with laboratory experiments. Finally, it would be satisfying  
50 if this behavior showed a glimpse of rapid learning.

51 Navigation through space is a complex behavior displayed by many animals. It typically  
52 involves integrating multiple cues to decide among many possible actions. It relies intimately  
53 on rapid learning. For example a pigeon or desert ant leaving its shelter acquires the information  
54 needed for the homing path in a single episode. Major questions remain about how the brain  
55 stores this information and converts it to a policy for decisions during the homing path. One  
56 way to formalize the act of decision-making in the laboratory is to introduce structure in  
57 the environment in the form of a maze that defines straight paths and decision points. For a  
58 burrowing rodent a maze of tunnels is in fact a natural environment. Early studies of rodent  
59 behavior did place the animals into true labyrinths (*Small, 1901*), but their use gradually  
60 declined in favor of linear tracks or boxes with a single choice point.

61 We report here on the behavior of laboratory mice in a complex labyrinth of tunnels. The  
62 mouse is placed in a home cage from which it has free access to the maze for one night. No  
63 handling, shaping, or training by the investigators is involved. By continuous video-recording  
64 and automated tracking we observe the animal's entire life experience with the labyrinth. Some  
65 of the mice are water-deprived and a single location deep inside the maze offers water. We find  
66 that these animals learn to navigate to the water port after just a few reward experiences. In  
67 many cases one can identify unique moments of "insight" when the animal's behavior changes  
68 discontinuously. This all happens within ~1 hour. Underlying the rapid learning is an efficient  
69 mode of exploration driven by some simple navigation rules. Mice that do not lack water  
70 show the same patterns of exploration. This laboratory-based navigation behavior may form a  
71 suitable substrate for studying the neural mechanisms that implement few-shot learning.

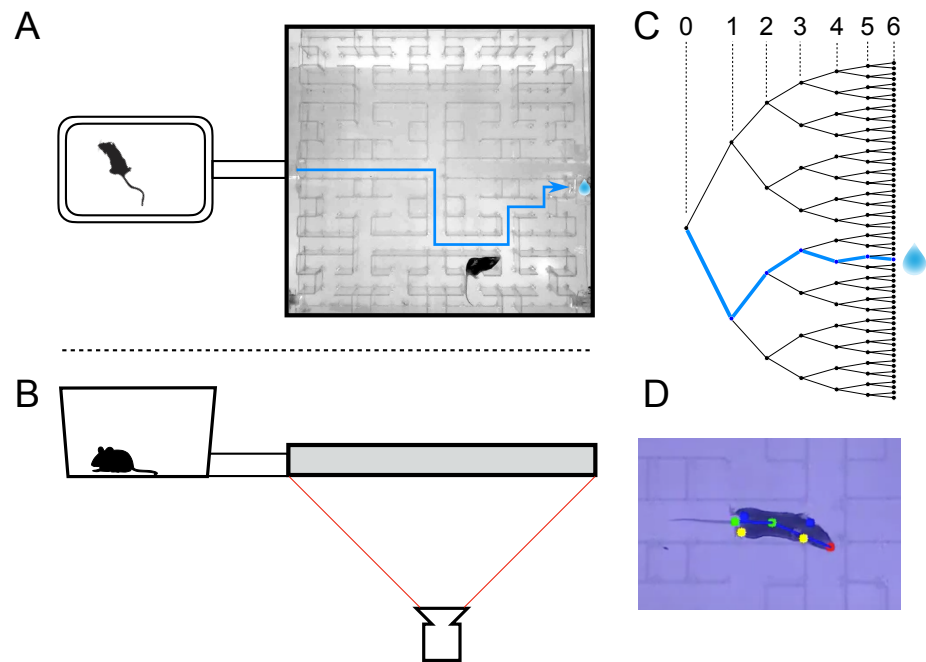
## 72 Results

### 73 Adaptation to the maze

74 At the start of the experiment a single mouse was placed in a conventional mouse cage with  
75 bedding and food. A short tunnel offered free access to a maze consisting of a warren of  
76 corridors (*Figure 1A-B*). The bottom and walls of the maze were constructed of black plastic  
77 that is transparent in the infrared. A video camera placed below the maze captured the animal's  
78 actions continuously using infrared illumination (*Figure 1B*). The recordings were analyzed  
79 offline to track the movements of the mouse, with keypoints on the nose, mid-body, tail base,  
80 and the four feet (*Figure 1D*). All observations were made in darkness during the animal's  
81 subjective night.

82 The logical structure of the maze is a binary tree, with 6 levels of branches, leading from the  
83 single entrance to 64 endpoints (*Figure 1C*). A total of 63 T-junctions are connected by straight  
84 corridors in a design with maximal symmetry (*Figure 1A, Figure 3-figure supplement 1*),  
85 such that all the nodes at a given level of the tree have the same local geometry. One of the 64  
86 endpoints of the maze is outfitted with a water port. After activation by a brief nose poke, the  
87 port delivers a small drop of water, followed by a 90-s time-out period.

88 We report observations from 20 animals using a frozen protocol developed from an initial

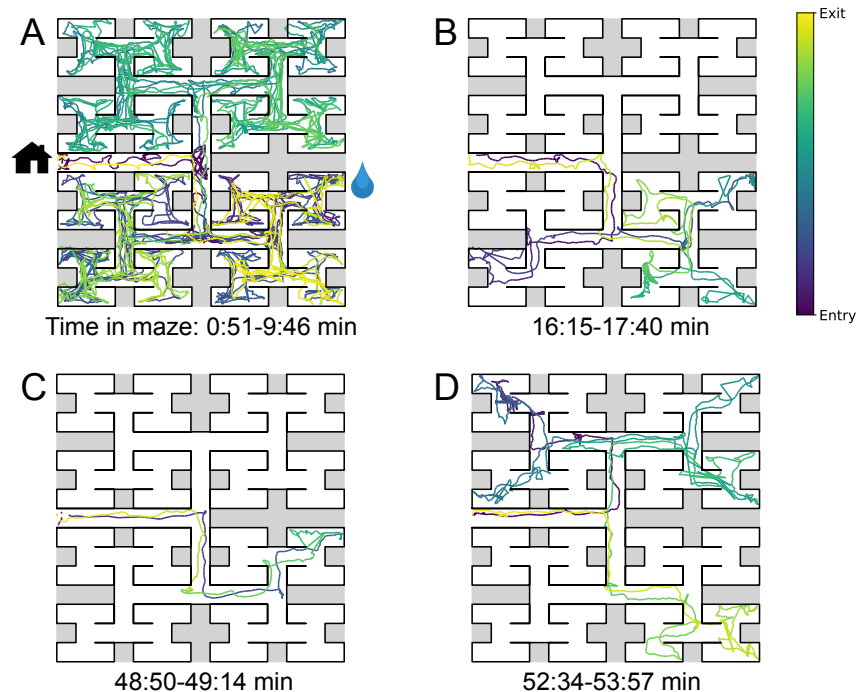


**Figure 1. The maze environment.** Top (A) and side (B) views of a home cage, connected via an entry tunnel to an enclosed labyrinth. The animal's actions in the maze are recorded via video from below using infrared illumination. (C) The maze is structured as a binary tree with 63 branch points (in levels numbered 0,...,5) and 64 end nodes. One end node has a water port that dispenses a drop when it gets poked. Blue line in A and C: path from maze entry to water port. (D) A mouse considering the options at the maze's central intersection. Colored keypoints are tracked by DeepLabCut: nose, mid body, tail base, 4 feet.

**Figure 1–figure supplement 1.** Occupancy of the maze.

**Figure 1–figure supplement 2.** Fraction of time in maze by group.

**Figure 1–figure supplement 3.** Transitions between cage and maze.



**Figure 2. Sample trajectories during adaptation to the maze.** Four sample bouts from one mouse (B3) into the maze at various times during the experiment (time markings at bottom). Position of the animal's nose as a function of time during the bout, which is encoded by the color of the trace. In panel A the entrance from the home cage and the water port are indicated with respective symbols.

**Figure 2–figure supplement 1. Speed of locomotion.**

89 period of exploratory experiments. Ten of these animals had been mildly water-deprived for  
90 24 hours; they received food in the home cage and water only from the port hidden in the maze.  
91 The other ten animals were sated and had free access to food and water in the cage. Each  
92 animal's behavior in the maze was recorded continuously for 7 h during the first night of its  
93 experience with the maze, starting the moment the connection tunnel was opened (watch a  
94 sample video [here](#)). The investigator played no role during this period, and the animal was free  
95 to act as it wished including travel between the cage and the maze.

96 All of the mice except one passed between the cage and the maze readily and frequently  
97 (*Figure 1–figure supplement 1*). The single outlier animal barely entered the maze and never  
98 progressed past the first junction; we excluded this mouse from subsequent analysis. On average  
99 over the entire period of study the animals spent 46% of the time in the maze (*Figure 1–figure  
100 supplement 2*). This fraction was similar whether or not the animal was motivated by water  
101 rewards (47% for rewarded vs 44% for unrewarded animals). Over time the animals appeared  
102 increasingly comfortable in the maze, taking breaks for grooming and the occasional nap.  
103 When the investigator lifted the cage lid at the end of the night some animals were seen to  
104 escape into the safety of the maze.

105 We examined the rate of transitions from the cage to the maze and how it depends on time  
106 spent in the cage (*Figure 1–figure supplement 3A*). Surprisingly the rate of entry into the  
107 maze is highest immediately after the animal returns to the cage. Then it declines gradually  
108 by a factor of 4 over the first minute in the cage and remains steady thereafter. This is a large  
109 effect, observed for every individual animal in both the rewarded and unrewarded groups. By  
110 contrast the opposite transition, namely exit from the maze, occurs at an essentially constant  
111 rate throughout the visit (*Figure 1–figure supplement 3B*).

112 The nature of the animal's forays into the maze changed over time. We call each foray  
113 from entrance to exit a "bout". After a few hesitant entries into the main corridor, the mouse  
114 engaged in one or more long bouts that dove deep into the binary tree to most or all of the  
115 leaf nodes (*Figure 2A*). For a water-deprived animal, this typically led to discovery of the  
116 reward port. After ~10 bouts, the trajectories became more focused, involving travel to the  
117 reward port and some additional exploration (*Figure 2B*). At a later stage still, the animal  
118 often executed perfect exploitation bouts that led straight to the reward port and back with no  
119 wrong turns (*Figure 2C*). Even at this late stage, however, the animal continued to explore  
120 other parts of the maze (*Figure 2D*). Similarly the unrewarded animals explored the maze  
121 throughout the night (*Figure 1-figure supplement 2*). While the length and structure of the  
122 animal's trajectories changed over time, the speed remained remarkably constant after ~50 s of  
123 adaptation (*Figure 2-figure supplement 1*).

124 Whereas *Figure 2* illustrates the trajectory of a mouse's nose in full spatio-temporal detail,  
125 a convenient reduced representation is the "node sequence". This simply marks the events  
126 when the animal arrives at one of the 127 nodes of the binary tree that describes the maze (see  
127 Methods and *Figure 3-figure supplement 1*). Among these nodes, 63 are T-junctions where  
128 the animal has 3 choices for the next node, and 64 are end nodes where the animal's only choice  
129 is to reverse course. We call the transition from one node to the next a "step". The following  
130 investigations all apply to an animal's node sequence.

### 131 **Few-shot learning of a reward location**

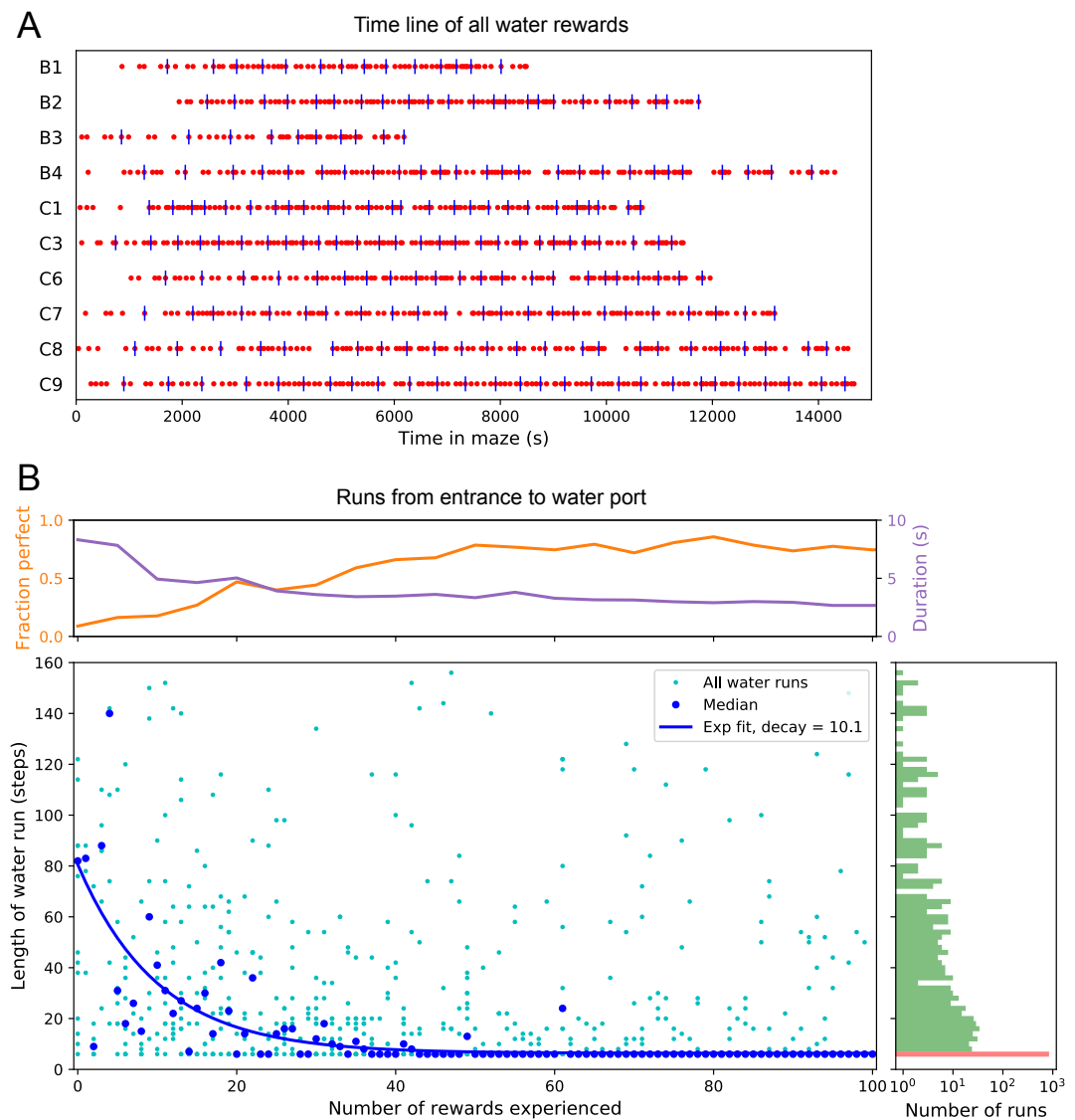
132 We now examine early changes in the animal's behavior that reveal how it rapidly acquires and  
133 remembers information needed for navigation. First we focus on navigation to the water port.

134 The ten water-deprived animals had no indication that water would be found in the maze.  
135 Yet all 10 discovered the water port in less than 2000 s and requiring fewer than 17 bouts  
136 (*Figure 3A*). The port dispensed only a drop of water followed by a 90-s timeout before  
137 rearming. During the timeout the animals generally left the port location to explore other parts  
138 of the maze or return home. For each of the water-deprived animals, the frequency at which it  
139 consumed rewards in the maze increased rapidly as it learned how to find the water port, then  
140 settled after a few reward experiences (*Figure 3A*).

141 How many reward experiences are sufficient to teach the animal reliable navigation to the  
142 water port? To establish a learning curve one wants to compare performance on the identical  
143 task over successive trials. Recall that this experiment has no imposed trial structure. Yet  
144 the animals naturally segmented their behavior through discrete visits to the maze. Thus we  
145 focused on all the instances when the animal started at the maze entrance and walked to the  
146 water port (*Figure 3B*).

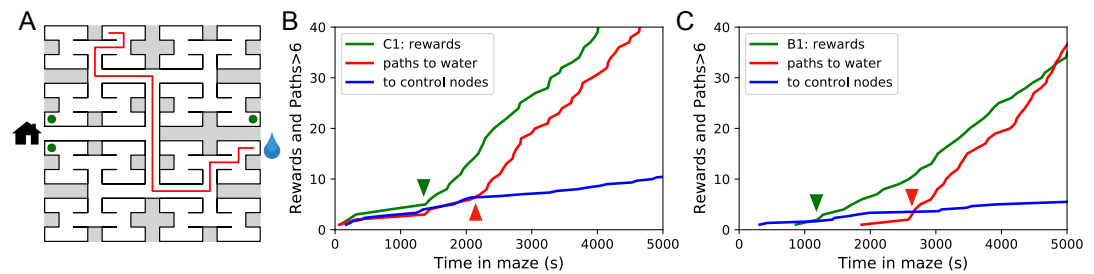
147 On the first few occasions these paths to water can involve hundreds of steps between nodes  
148 and their length scatters over a wide range. However, after a few rewards, the animals began  
149 taking the perfect path without detours (length 6, *Figure 3-figure supplement 1*), and soon  
150 that became the norm. Note the path length plotted here is directly related to the number of  
151 "turning errors": Every time the mouse turns away from the shortest path to the water port that  
152 adds two steps to the path length (*Equation 7*). The rate of these errors declined over time, by  
153 a factor of  $e$  after ~10 rewards consumed (*Figure 3B*). Late in the night ~75% of the paths to  
154 water were perfect. The animals executed them with increasing speed; eventually these fast  
155 "water runs" took as little as 2 s (*Figure 3B*). Many of these visits went unrewarded owing to  
156 the 90-s timeout period on the water port.

157 In summary, after ~10 reward experiences on average the mice learn to navigate efficiently  
158 to the water port, which requires making 6 correct decisions, each among 3 options. Note that  
159 even at late times, long after they have perfected the "water run", the animals continue to take



**Figure 3. Few-shot learning of path to water.** (A) Time line of all water rewards collected by 10 water-deprived mice (red dots, every fifth reward has a blue tick mark). (B) The length of runs from the entrance to the water port, measured in steps between nodes, and plotted against the number of rewards experienced. Main panel: All individual runs (cyan dots) and median over 10 mice (blue circles). Exponential fit decays by  $1/e$  over 10.1 rewards. Right panel: Histogram of the run length, note log axis. Red: perfect runs with the minimum length 6; green: longer runs. Top panel: The fraction of perfect runs (length 6) plotted against the number of rewards experienced, along with the median duration of those perfect runs.

**Figure 3—figure supplement 1.** Definition of node trajectories.



**Figure 4. Sudden changes in behavior.** (A) An example of a long uninterrupted path through 11 junctions to the water port (drop icon). Circles mark control nodes related by symmetry to the water port to assess the frequency of long paths occurring by chance. (B) For one animal (named C1) the cumulative number of rewards (green); of long paths (>6 junctions) to the water port (red); and of similar paths to the 3 control nodes (blue, divided by 3). All are plotted against the time spent in the maze. Arrowheads indicate the time of sudden changes, obtained from fitting a step function to the rates. (C) Same as B for animal B1.

**Figure 4–figure supplement 1.** Long direct paths for all animals.

**Figure 4–figure supplement 2.** Statistics of sudden changes in behavior.

160 some extremely long paths: a subject for a later section (*Figure 6*).

### 161 **Discontinuous learning**

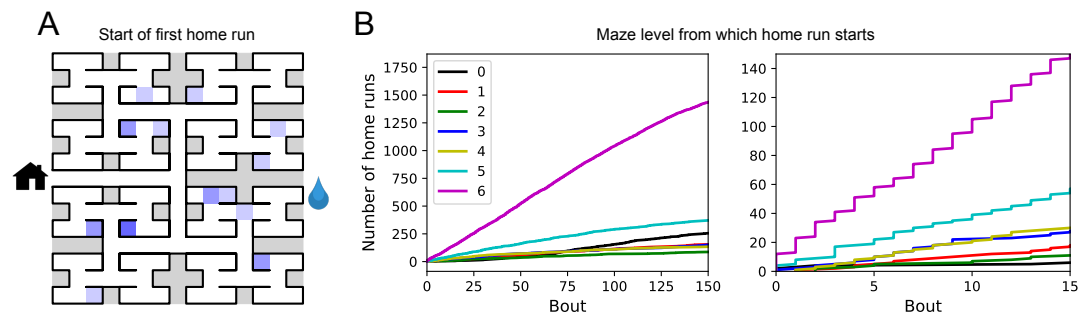
162 While an average across animals shows evidence of rapid learning (*Figure 3*) one wonders  
163 whether the knowledge is acquired gradually or through moments of “sudden insight”. To  
164 explore this we scrutinized more closely the time line of individual water-deprived animals in  
165 their experience with the maze. The discovery of the water port and the subsequent collection of  
166 water drops at a regular rate is one clear change in behavior that relies on new knowledge. This  
167 rate of water rewards can increase rather suddenly (*Figure 3A*), suggesting an instantaneous  
168 step in knowledge.

169 Over time, the animals learned the path to water not only from the entrance of the maze but  
170 from many locations scattered throughout the maze. The largest distance between the water  
171 port and an end node in the opposite half of the maze involves 12 steps through 11 intersections  
172 (*Figure 4A*). Thus we included as another behavioral variable the occurrence of long direct  
173 paths to the water port which reflects how directly the animals navigate within the maze.

174 *Figure 4B* shows for one animal the cumulative occurrence of water rewards and that of  
175 long direct paths to water. The animal discovers the water port early on at 75 s, but at 1380  
176 s the rate of water rewards jumps suddenly by a factor of 5. The long paths to water follow  
177 a rather different time line. At first they occur randomly, at the same rate as the paths to the  
178 unrewarded control nodes. At 2070 s the long paths suddenly increase in frequency by a factor  
179 of 5. Given the sudden change in rates of both kinds of events there is little ambiguity about  
180 when the two steps happen and they are well separated in time (*Figure 4B*).

181 The animal behaves as though it gains a new insight at the time of the second step that  
182 allows it to travel to the water port directly from elsewhere in the maze. Note that the two  
183 behavioral variables are independent: The long paths don’t change when the reward rate steps  
184 up, and the reward rate doesn’t change when the rate of long paths steps up. Another animal  
185 (*Figure 4C*) similarly showed an early step in the reward rate (at 860 s) and a dramatic step in  
186 the rate of long paths (at 2580 s). In this case the emergence of long paths coincided with a  
187 modest increase (factor of 2) in the reward rate.

188 Similar discontinuities in behavior were seen in at least 5 of the 10 water-deprived animals  
189 (*Figure 4–figure supplement 1*, *Figure 4–figure supplement 2*), and their timing could be



**Figure 5. Homing succeeds on first attempt.** (A) Locations in the maze where the 19 animals started their first return to the exit (home run). Some locations were used by 2 or 3 animals (darker color). (B) Left: The cumulative number of home runs from different levels in the maze, summed over all animals, and plotted against the bout number. Level 0 = first T-junction, level 6 = end nodes. Right: Zoom of (Left) into early bouts.

**Figure 5–figure supplement 1.** Entry paths do not retrace exit paths.

190 identified to a precision of ~200 s. We varied the criterion of performance by asking for even  
191 longer error-free paths, but the results were largely unchanged and no additional discontinuity  
192 appeared. These observations suggest that mice can acquire a complex decision-making skill  
193 rather suddenly. A mouse may have multiple moments of sudden insight that affect different  
194 aspects of its behavior. The exact time of the insight cannot be predicted but is easily identified  
195 post-hoc. Future neurophysiological studies of the phenomenon will face the interesting  
196 challenge of capturing these singular events.

### 197 **One-shot learning of the home path**

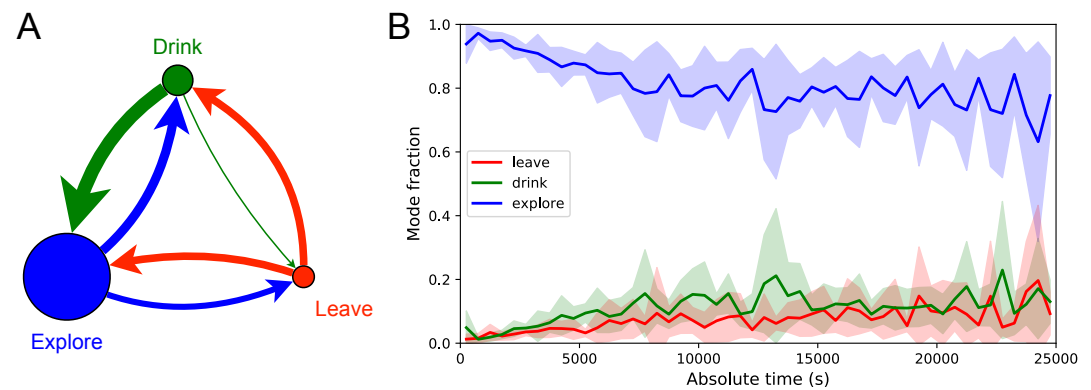
198 For an animal entering an unfamiliar environment, the most important path to keep in memory  
199 may be the escape route. In the present case that is the route to the maze entrance, from which  
200 the tunnel leads home to the cage. We expected that the mice would begin by penetrating into  
201 the maze gradually and return home repeatedly so as to confirm the escape route. This might  
202 help build a memory of the home path gradually level-by-level into the binary tree. Nothing  
203 could be further from the truth.

204 At the end of any given bout into the maze, there is a “home run”, namely the direct  
205 path without reversals that takes the animal to the exit (see *Figure 3–figure supplement 1*).  
206 *Figure 5A* shows the nodes where each animal started its first home run, following the first  
207 penetration into the maze. With few exceptions that first home run began from an end node,  
208 as deep into the maze as possible. Recall that this involves making the correct choice at six  
209 successive 3-way intersections, an outcome that is unlikely to happen by chance.

210 The above hypothesis regarding gradual practice of home runs would predict that short  
211 home runs should appear before long ones in the course of the experiment. The opposite is the  
212 case (*Figure 5B*). In fact, the end nodes (level 6 of the maze) are by far the favorite place from  
213 which to return to the exit, and those maximal-length home runs systematically appear before  
214 shorter ones. This conclusion was confirmed for each individual animal, whether rewarded or  
215 unrewarded.

216 Clearly the animals do not practice the home path or build it up gradually. Instead they  
217 seem to possess an Ariadne’s thread (*Pseudo-Apollodorus, 1st Century AD?*) starting with  
218 their first excursion into the maze, long before they might have acquired any general knowledge  
219 of the maze layout. On the other hand the mouse does not follow the strategy of Theseus,  
220 namely to precisely retrace the path that led it into the labyrinth. In that case the animal’s home  
221 path should be the reverse of the path into the maze that started the bout. Instead the entry path





**Figure 6. Exploration is a dominant and persistent mode of behavior.** (A) Ethogram for rewarded animals. Area of the circle reflects the fraction of time spent in each behavioral mode averaged over animals and duration of the experiment. Width of the arrow reflects the probability of transitioning to another mode. ‘Drink’ involves travel to the water port and time spent there. Transitions from ‘Leave’ represent what the animal does at the start of the next bout into the maze. (B) The fraction of time spent in each mode as a function of absolute time throughout the night. Mean  $\pm$  SD across the 10 rewarded animals.

**Figure 6–figure supplement 1. Three modes of behavior.**

222 and the home path have very little overlap (*Figure 5–figure supplement 1*). It appears that the  
223 animal acquires a homing strategy over the course of a single bout, and in a manner that allows  
224 a direct return home from locations not previously encountered.

### 225 **Structure of behavior in the maze**

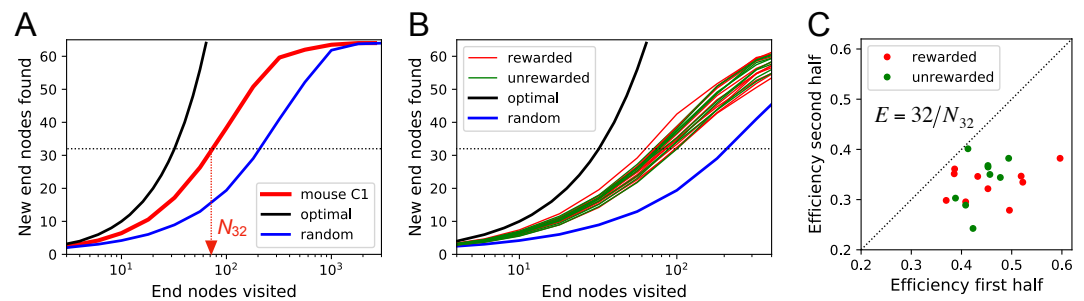
226 Here we focus on rules and patterns that govern the animal’s activity in the maze on both large  
227 and small scales.

#### 228 **Behavioral states**

229 Once the animal has learned to perform long uninterrupted paths to the water port, one can  
230 categorize its behavior by three states: (1) walking to the water port; (2) walking to the exit;  
231 and (3) exploring the maze. Operationally we define exploration as all periods in which the  
232 animal is in the maze but not on a direct path to water or to the exit. For the ten satiated animals  
233 this includes all times in the maze except for the walks to the exit.

234 *Figure 6* illustrates the occupancies and transition probabilities between these states. The  
235 animals spent most of their time by far in the exploration state: 84% for rewarded and 95%  
236 for unrewarded mice. Across animals there was very little variation in the balance of the 3  
237 modes (*Figure 6–figure supplement 1*). The rewarded mice began about half their bouts into  
238 the maze with a trip to the water port and the other half by exploring (*Figure 6A*). After a  
239 drink, the animals routinely continued exploring, about 90% of the time.

240 For water-deprived animals the dominance of exploration persisted even at a late stage of  
241 the night when they routinely executed perfect exploitation bouts to and from the water port:  
242 Over the duration of the night the ‘explore’ fraction dropped slightly from 0.92 to 0.75, with the  
243 balance accrued to the ‘drink’ and ‘leave’ modes as the animals executed many direct runs to the  
244 water port and back. The unrewarded group of animals also explored the maze throughout the  
245 night even though it offered no overt rewards (*Figure 6–figure supplement 1*). One suspects  
246 that the animals derive some intrinsic reward from the act of patrolling the environment itself.



**Figure 7. Exploration covers the maze efficiently.** (A) The number of distinct end nodes encountered as a function of the number of end nodes visited for: mouse C1 (red); the optimal explorer agent (black); an unbiased random walk (blue). Arrowhead: the value  $N_{32} = 76$  by which mouse C1 discovered half of the end nodes. (B) An expanded section of the graph in A including curves from 10 rewarded (red) and 9 unrewarded (green) animals. The efficiency of exploration, defined as  $E = 32/N_{32}$ , is  $0.385 \pm 0.050$  (SD) for rewarded and  $0.384 \pm 0.039$  (SD) for unrewarded mice. (C) The efficiency of exploration for the same animals, comparing the values in the first and second halves of the time in the maze. The decline is a factor of  $0.74 \pm 0.12$  (SD) for rewarded and  $0.81 \pm 0.13$  (SD) for unrewarded mice.

**Figure 7–figure supplement 1. Efficiency of exploration**

#### 247 Efficiency of exploration

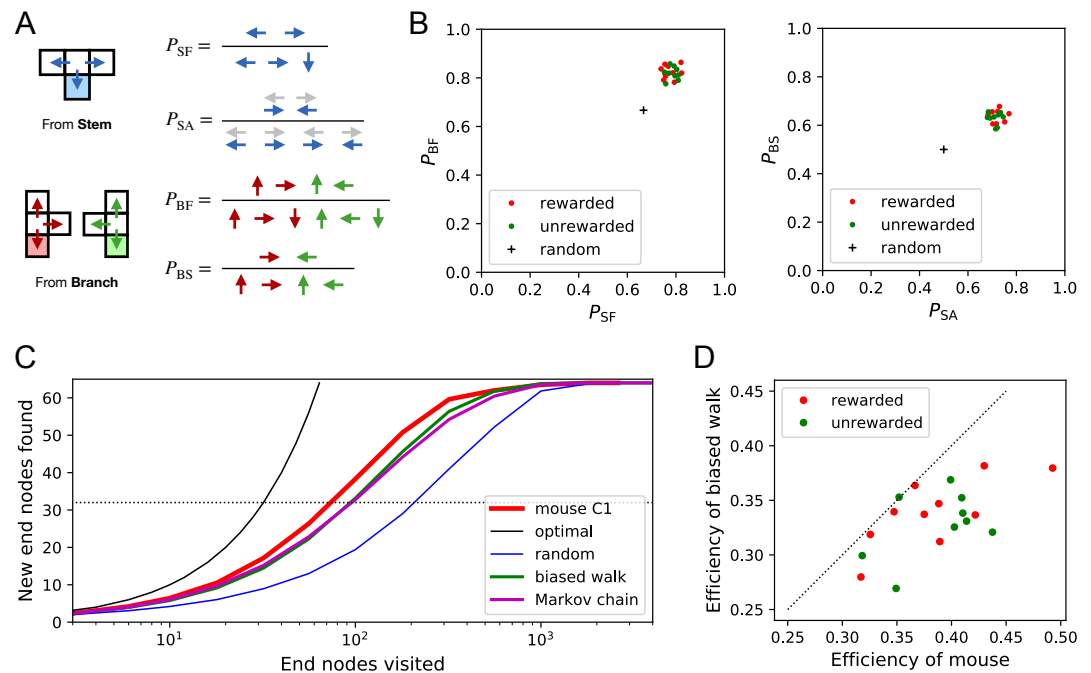
248 During the direct paths to water and to the exit the animal behaves deterministically, whereas  
 249 the exploration behavior appears stochastic. Here we delve into the rules that govern the  
 250 exploration component of behavior.

251 One can presume that a goal of the exploratory mode is to rapidly survey all parts of the  
 252 environment for the appearance of new resources or threats. We will measure the efficiency of  
 253 exploration by how rapidly the animal visits all end nodes of the binary maze, starting at any  
 254 time during the experiment. The optimal agent with perfect memory and complete knowledge  
 255 of the maze – including the absence of any loops – could visit the end nodes systematically  
 256 one after another without repeats, thus encountering all of them after just 64 visits. A less  
 257 perfect agent, on the other hand, will visit the same node repeatedly before having encountered  
 258 all of them. *Figure 7A* plots for one exploring mouse the number of distinct end nodes it  
 259 encountered as a function of the number of end nodes visited. The number of new nodes rises  
 260 monotonically; 32 of the end nodes have been discovered after the mouse checked 76 times;  
 261 then the curve gradually asymptotes to 64. We will characterize the efficiency of the search by  
 262 the number of visits  $N_{32}$  required to survey half the end nodes, and define

$$E = \frac{32}{N_{32}}. \quad (1)$$

263 This mouse explores with efficiency  $E = 32/76 = 0.42$ . For comparison, *Figure 7A* plots the  
 264 performance of the optimal agent ( $E = 1.0$ ) and that of a random walker that makes random  
 265 decisions at every 3-way junction ( $E = 0.23$ ). Note the mouse is about half as efficient as the  
 266 optimal agent, but twice as efficient as a random walker.

267 The different mice were remarkably alike in this component of their exploratory behavior  
 268 (*Figure 7B*): across animals the efficiency varied by only 11% of the mean ( $0.387 \pm 0.044$  SD).  
 269 Furthermore there was no detectable difference in efficiency between the rewarded animals and  
 270 the sated unrewarded animals. Over the course of the night the efficiency declined significantly  
 271 for almost every animal – whether rewarded or not – by an average of 23% (*Figure 7C*).



**Figure 8. Turning biases favor exploration.** (A) Definition of four turning biases at a T-junction based on the ratios of actions taken. Top: An animal arriving from the stem of the T (shaded) may either reverse or turn left or right.  $P_{SF}$  is the probability that it will move forward rather than reversing. Given that it moves forward,  $P_{SA}$  is the probability that it will take an alternating turn from the preceding one (gray), i.e. left-right or right-left. Bottom: An animal arriving from the bar of the T may either reverse or go straight, or turn into the stem of the T.  $P_{BF}$  is the probability that it will move forward through the junction rather than reversing. Given that it moves forward,  $P_{BS}$  is the probability that it turns into the stem. (B) Scatter graph of the biases  $P_{SF}$  and  $P_{BF}$  (left) and  $P_{SA}$  and  $P_{BS}$  (right). Every dot represents a mouse. Cross: values for an unbiased random walk. (C) Exploration curve of new end nodes discovered vs end nodes visited, displayed as in *Figure 7A*, including results from a biased random walk with the 4 turning biases derived from the same mouse, as well as a more elaborate Markov-chain model (see *Figure 10C*). (D) Efficiency of exploration (*Equation 1*) in 19 mice compared to the efficiency of the corresponding biased random walk.

**Figure 8-figure supplement 1. Bias statistics.**

## 272 Rules of exploration

273 What allows the mice to search much more efficiently than a random walking agent? We  
274 inspected more closely the decisions that the animals make at each 3-way junction. It emerged  
275 that these decisions are governed by strong biases (*Figure 8*). The probability of choosing  
276 each arm of a T-junction depends crucially on how the animal entered the junction. The animal  
277 can enter a T-junction from 3 places and exit it in 3 directions (*Figure 8A*). By tallying the  
278 frequency of all these occurrences across all T-junctions in the maze one finds clear deviations  
279 from an unbiased random walk (*Figure 8B, Figure 8-figure supplement 1*).

280 First, the animals have a strong preference for proceeding through a junction rather than  
281 returning to the preceding node ( $P_{SF}$  and  $P_{BF}$  in *Figure 8B*). Second there is a bias in favor  
282 of alternating turns left and right rather than repeating the same direction turn ( $P_{SA}$ ). Finally,  
283 the mice have a mild preference for taking a branch off the straight corridor rather than  
284 proceeding straight ( $P_{BS}$ ). A comparison across animals again revealed a remarkable degree  
285 of consistency even in these local rules of behavior: The turning biases varied by only 3%  
286 across the population and even between the rewarded and unrewarded groups (*Figure 8B,*  
287 *Figure 8-figure supplement 1*).

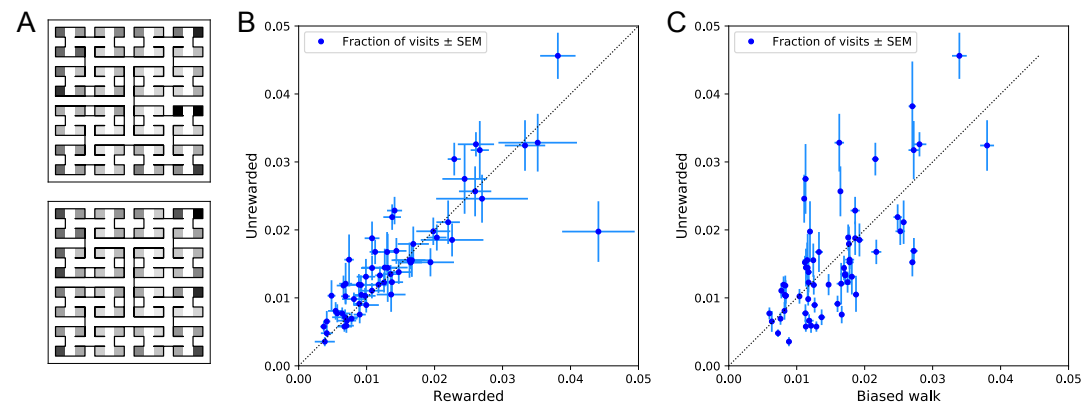
288 Qualitatively, one can see that these turning biases will improve the animal's search strategy.  
289 The forward biases  $P_{SF}$  and  $P_{BF}$  keep the animal from re-entering territory it has covered already.  
290 The bias  $P_{BS}$  favors taking a branch that leads out of the maze. This allows the animal to rapidly  
291 cross multiple levels during an outward path and then enter a different territory. By comparison,  
292 the unbiased random walk tends to get stuck in the tips of the tree and revisits the same end  
293 nodes many times before escaping. To test this intuition we simulated a biased random agent  
294 whose turning probabilities at a T-junction followed the same biases as measured from the  
295 animal (*Figure 8C*). These biased agents did in fact search with much higher efficiency than  
296 the unbiased random walk. They did not fully explain the behavior of the mice (*Figure 8D*),  
297 accounting for ~87% of the animal's efficiency (compared to 60% for the random walk). A more  
298 sophisticated model of the animal's behavior - involving many more parameters (*Figure 10C*) -  
299 failed to get any closer to the observed efficiency (*Figure 8C, Figure 7-figure supplement 1C*).  
300 Clearly some components of efficient search in these mice remain to be understood.

## 301 Systematic node preferences

302 A surprising aspect of the animals' explorations is that they visit certain end nodes of the  
303 binary tree much more frequently than others (*Figure 9*). This effect is large: more than a  
304 factor of 10 difference between the occupancy of the most popular and least popular end nodes  
305 (*Figure 9A-B*). This was surprising given our efforts to design the maze symmetrically, such  
306 that in principle all end nodes should be equivalent. Furthermore the node preferences were  
307 very consistent across animals and even across the rewarded and unrewarded groups. Note that  
308 the standard error across animals of each node's occupancy is much smaller than the differences  
309 between the nodes (*Figure 9B*).

310 The nodes on the periphery of the maze are systematically preferred. Comparing the  
311 outermost ring of 26 end nodes (excluding the water port and its neighbor) to the innermost 16  
312 end nodes, the outer ones are favored by a large factor of 2.2. This may relate to earlier reports  
313 of a "centrifugal tendency" among rats patrolling a maze (*Uster et al., 1976*).

314 Interestingly, the biased random walk using four bias numbers (*Figure 8, Figure 10D*)  
315 replicates a good amount of the pattern of preferences. For unrewarded animals, where the  
316 maze symmetry is not disturbed by the water port, the biased random walk predicts 51% of  
317 the observed variance across nodes (*Figure 9C*), and an outer/inner node preference of 1.97,  
318 almost matching the observed ratio of 2.20. The far more complex Markov-chain model of  
319 behavior (*Figure 10C*) performed slightly better, explaining 66% of the variance in port visits



**Figure 9. Preference for certain end nodes during exploration.** (A) The number of visits to different end nodes encoded by a gray scale. Top: rewarded, bottom: unrewarded animals. Gray scale spans a factor of 12 (top) or 13 (bottom). (B) The fraction of visits to each end node, comparing the rewarded vs unrewarded group of animals. Each data point is for one end node, the error bar is the SEM across animals in the group. The outlier on the bottom right is the neighbor of the water port, a frequently visited end node among rewarded animals. The water port is off scale and not shown. (C) As in panel B but comparing the unrewarded animals to their simulated 4-bias random walks. These biases explain 51% of the variance in the observed preference for end nodes.

320 and matching the outer/inner node preference of 2.20.

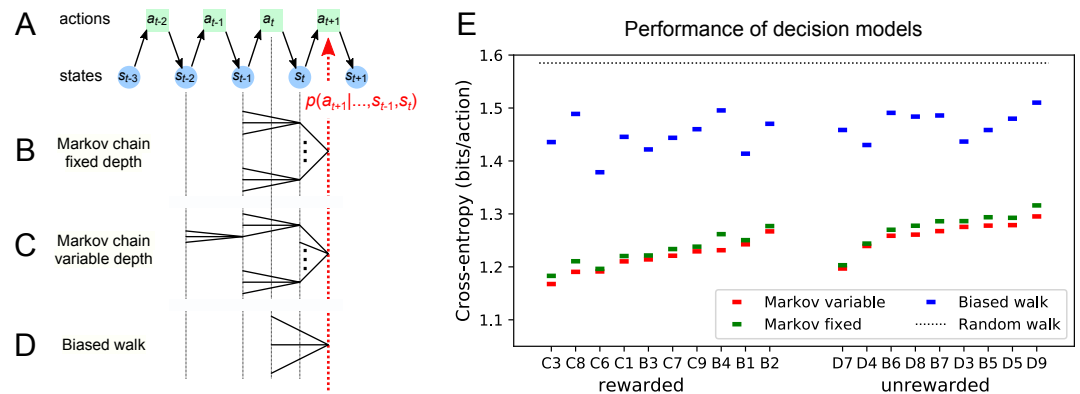
### 321 Models of maze behavior

322 Moving beyond the efficiency of exploration one may ask more broadly: How well do we really  
323 understand what the mouse does in the maze? Can we predict its action at the next junction?  
324 Once the predictable component is removed, how much intrinsic randomness remains in the  
325 mouse's behavior? Here we address these questions using more sophisticated models that  
326 predict the probability of the mouse's future actions based on the history of its trajectory.

327 At a formal level, the mouse's trajectory through the maze is a string of numbers standing  
328 for the nodes the animal visited (*Figure 10A* and *Figure 3-figure supplement 1*). We want to  
329 predict the next action of the mouse, namely the step that takes it to the next node. The quality  
330 of the model will be assessed by the cross-entropy between the model's predictions and the  
331 mouse's observed actions, measured in bits per action. This is the uncertainty that remains  
332 about the mouse's next action given the prediction from the model. The ultimate lower limit is  
333 the true source entropy of the mouse, namely that component of its decisions that cannot be  
334 explained by the history of its actions.

335 One family of models we considered are fixed-depth Markov chains (*Figure 10B*). Here  
336 the probability of the next action  $a_{t+1}$  is specified as a function of the history stretching over  
337 the  $k$  preceding nodes  $(s_{t-k+1}, \dots, s_t)$ . In fitting the model to the mouse's actual node sequence  
338 one tallies how often each history leads to each action, and uses those counts to estimate  
339 the conditional probabilities  $p(a_{t+1}|s_{t-k+1}, \dots, s_t)$ . Given a new node sequence, the model  
340 will then use the history strings  $(s_{t-k+1}, \dots, s_t)$  to predict the outcome of the next action. In  
341 practice we trained the model on 80% of the animal's trajectory and tested it by evaluating the  
342 cross-entropy on the remaining 20%.

343 Ideally, the depth  $k$  of these action trees would be very large, so as to take as much of the  
344 prior history into account as possible. However, one soon runs into a problem of over-fitting:  
345 Because each T-junction in the maze has 3 neighboring junctions, the number of possible  
346 histories grows as  $3^k$ . As  $k$  increases, this quickly exceeds the length of the measured node  
347 sequence, so that every history appears only zero or one times in the data. At this point one



**Figure 10. Recent history constrains the mouse's decisions.** (A) The mouse's trajectory through the maze produces a sequence of states  $s_t$  = node occupied after step  $t$ . From each state, up to 3 possible actions lead to the next state (end nodes allow only one action). We want to predict the animal's next action,  $a_{t+1}$ , based on the prior history of states or actions. (B-D) Three possible models to make such a prediction. (B) A fixed-depth Markov chain where the probability of the next action depends only on the current state  $s_t$  and the preceding state  $s_{t-1}$ . The branches of the tree represent all  $3 \times 127$  possible histories  $(s_{t-1}, s_t)$ . (C) A variable-depth Markov chain where only certain branches of the tree of histories contribute to the action probability. Here one history contains only the current state, some others reach back three steps. (D) A biased random walk model, as defined in *Figure 8*, in which the probability of the next action depends only on the preceding action, not on the state. (E) Performance of the models in (B,C,D) when predicting the decisions of the animal at T-junctions. In each case we show the cross-entropy between the predicted action probability and the real actions of the animal (lower values indicate better prediction, perfect prediction would produce zero). Dotted line represents an unbiased random walk with 1/3 probability of each action.

**Figure 10-figure supplement 1. Markov model fits.**

348 can no longer estimate any probabilities, and cross-validation on a different segment of data  
349 fails catastrophically. In practice we found that this limitation sets in already beyond  $k = 2$   
350 (*Figure 10–figure supplement 1A*). To address this issue of data-limitation we developed a  
351 variable-depth Markov chain (*Figure 10C*). This model retains longer histories, but only if  
352 they occur frequently enough to allow a reliable probability estimate (see Methods, *Figure 10–*  
353 *figure supplement 1B-C*). In addition, we explored different schemes of pooling the counts  
354 across certain T-junctions that are related by the symmetry of the maze (see Methods).

355 With these methods we focused on the portions of trajectory when the mouse was in ‘explore’  
356 mode, because the segments in ‘drink’ and ‘leave’ mode are fully predictable. Furthermore, we  
357 evaluated the models only at nodes corresponding to T-junctions, because the decision from an  
358 end node is again fully predictable. *Figure 10E* compares the performance of various  
359 models of mouse behavior. The variable-depth Markov chains routinely produced the best fits,  
360 although the improvement over fixed-depth models was modest. Across all 19 animals in this  
361 study the remaining uncertainty about the animal’s action at a T-junction is  $1.237 \pm 0.035$  (SD)  
362 bits/action, compared to the prior uncertainty of  $\log_2 3 = 1.585$  bits. The rewarded animals  
363 have slightly lower entropy than the unrewarded ones (1.216 vs 1.261 bits/action). The Markov  
364 chain models that produced the best fits to the behavior used history strings with an average  
365 length of  $\sim 4$ .

366 We also evaluated the predictions obtained from the simple biased random walk model  
367 (*Figure 10D*). Recall that this attempts to capture the history-dependence with just 4 bias  
368 parameters (*Figure 8A*). As expected this produced considerably higher cross-entropies than  
369 the more sophisticated Markov chains (by about 18%, *Figure 10E*). Finally we used several  
370 professional file compression routines to try and compress the mouse’s node sequence. In  
371 principle, this sets an upper bound on the true source entropy of the mouse, even if the  
372 compression algorithm has no understanding of animal behavior. The best such algorithm (bzip2  
373 compression (*Seward, 2019*)) far under-performed all the other models of mouse behavior,  
374 giving 43% higher cross-entropy on average, and thus offered no additional useful bounds.

375 We conclude that during exploration of the maze the mouse’s choice behavior is strongly  
376 influenced by its current location and  $\sim 3$  locations preceding it. There are minor contributions  
377 from states further back. By knowing the animal’s history one can narrow down its action  
378 plan at a junction from the *a priori* 1.59 bits (one of three possible actions) to just  $\sim 1.24$  bits.  
379 This finally is a quantitative answer to the question, “How well can one predict the animal’s  
380 behavior?” Whether the remainder represents an irreducible uncertainty – akin to “free will”  
381 of the mouse – remains to be seen. Readers are encouraged to improve on this number by  
382 applying their own models of behavior to our published data set.

## 383 Discussion

### 384 Summary of contributions

385 We present a new approach to the study of learning and decision-making in mice. We give  
386 the animal access to a complex labyrinth and leave it undisturbed for a night while monitoring  
387 its movements. The result is a rich data set that reveals new aspects of learning and the  
388 structure of exploratory behavior. With these methods we find that mice learn a complex  
389 task that requires 6 correct 3-way decisions after only  $\sim 10$  experiences of success (*Figure 2,*  
390 *Figure 3*). Along the way the animal gains task knowledge in discontinuous steps that can be  
391 localized to within a few minutes of resolution (*Figure 4*). Underlying the learning process is  
392 an exploratory behavior that occupies 90% of the animal’s time in the maze, persists long after  
393 the task has been mastered, and even in complete absence of an extrinsic reward (*Figure 6*).  
394 The decisions the animal makes at choice points in the labyrinth are constrained in part by

395 the history of its actions (*Figure 8, Figure 10*), in a way that favors efficient searching of the  
396 maze (*Figure 7*). This microstructure of behavior is surprisingly consistent across mice, with  
397 variation in parameters of only a few percent (*Figure 8*). Our most expressive models to predict  
398 the animal's choices still leave a remaining uncertainty of  $\sim 1.24$  bits per decision (*Figure 10*),  
399 a quantitative benchmark by which competing models can be tested. Finally – as discussed  
400 below – some of the observations constrain what algorithms the animals might use for learning  
401 and navigation.

## 402 **Historical context**

403 Mazes have been a staple of animal psychology for well over 100 years. The early versions  
404 were true labyrinths. For example, *Small (1901)* built a model of the maze in Hampton Court  
405 gardens scaled to rat size. Subsequent researchers felt less constrained by Victorian landscapes  
406 and began to simplify the maze concept. Most commonly the maze offered one standard path  
407 from a starting location to a food reward box. A few blind alleys would branch from the  
408 standard path, and researchers would tally how many errors the animal committed by briefly  
409 turning into a blind (*Tolman and Honzik, 1930*). Later on, the design was further reduced to  
410 a single T-junction. After all, the elementary act of maze navigation is whether to turn left  
411 or right at a junction (*Tolman, 1938*), so why not study that process in isolation? One can  
412 avoid the tedium of carrying the animal from the exit to the entrance by closing off all branches  
413 of the T, forcing the animal to retrace its steps, which leads to the W-track (*Kim and Frank,*  
414 *2009*). And reducing the concept even further, one can ask the animal to refrain from walking  
415 altogether, and instead poke its nose into a hole on the left or the right side of a box (*Uchida*  
416 *and Mainen, 2003*). This led to the popular behavior boxes now found in rodent neuroscience  
417 laboratories everywhere. While each of these reductions of the “maze” concept enabled a new  
418 type of experiment, the essence of a “confusing network of paths” has been lost entirely, and  
419 with it the behavioral richness of the animals navigating those paths.

420 On this background our experimental design is a step back to complex labyrinths, enhanced  
421 with fully automated animal tracking and new methods of data analysis. In fact, our labyrinth  
422 is considerably more complex than Hampton Court or most of the mazes employed by Tolman  
423 and others (*Tolman and Honzik, 1930; Buel, 1934; Munn, 1950a*): The blind alleys in those  
424 mazes are all short and unbranched. When an animal strays from the target path it receives  
425 feedback quickly and can correct. By contrast our binary tree maze has 64 equally deep  
426 branches, only one of which contains the reward port. If the animal makes a mistake at any  
427 level of the tree it can find out only after traveling all the way to the last node.

428 Another unusual but crucial aspect of our experimental design is the absence of any human  
429 interference. Most studies of animal navigation and learning involve some kind of trial structure.  
430 For example the experimenter puts the rat in the start box, watches it make its way through the  
431 maze, coaxes it back on the path if necessary, and picks it up once it reaches the target box.  
432 Then another trial starts. In modern experiments with two-alternative-forced-choice (2AFC)  
433 behavior boxes the animal doesn't have to be picked up, but a trial starts with appearance of  
434 a cue, and then proceeds through some strict protocol through delivery of the reward. The  
435 argument in favor of imposing a trial structure is that it creates reproducible conditions, so that  
436 one can gather comparable data and average them suitably over many trials.

437 Our experiments had no imposed structure whatsoever; in fact it may be inappropriate to  
438 call them experiments. The investigator opened the entry to the maze in the evening and did  
439 not return until the morning. A potential advantage of leaving the animals to themselves is  
440 that they are more likely to engage in mouse-like behavior, rather than constantly responding  
441 to the stress of human interference or the pressures from being a cog in a behavior machine.  
442 The result was a rich data set, with the typical animal delivering  $\sim 15,000$  decisions in a single



443 night, even if one only counts the nodes of the binary tree as decision points. Since the mice  
444 made all the choices, the scientific effort lay primarily in adapting methods of data analysis to  
445 the nature of mouse trajectories. Somewhat surprisingly, the absence of experimental structure  
446 was no obstacle to making precise and reproducible measurements of the animal's behavior.

#### 447 **How fast do animals learn?**

448 Among the wide range of phenomena of animal learning, one can distinguish easy and hard  
449 tasks by some measure of task complexity. In a simple picture of a behavioral task the animal  
450 needs to recognize several different contexts and based on that express one of several different  
451 actions. One can draw up a contingency table between contexts and actions, and measure the  
452 complexity of the task by the mutual information in that table. This ignores any task difficulties  
453 associated with sensing the context at all or with producing the desired actions. However,  
454 in all the examples discussed here the stimuli are discriminated easily and the actions come  
455 naturally, thus the learning difficulty lies only in forming the associations, not in sharpening  
456 the perceptual mechanisms or practicing complex motor output.

457 Many well-studied behaviors have a complexity of 1 bit or less, and often animals can learn  
458 these associations after a single experience. For example, the Bruce effect (*Bruce, 1959*) is  
459 a learning phenomenon in which the female mouse forms an olfactory memory of her mate.  
460 This allows her to identify a foreign male in a subsequent encounter, in which case she aborts  
461 the pregnancy. In effect the female maps two different contexts (smell of mate vs non-mate)  
462 onto two kinds of pregnancy outcomes (carry to term vs abort). The mutual information in that  
463 contingency table is at most 1 bit, and may be considerably lower, for example if non-mate  
464 males are very rare or very frequent. Mice form the correct association after a single instance  
465 of mating, although proper memory formation requires several hours of exposure to the mate  
466 odor (*Rosser and Keverne, 1985*).

467 Similarly fear learning under the common electroshock paradigm establishes a mapping  
468 between two kinds of contexts (paired with shock vs innocuous) and two actions (freeze vs  
469 proceed), again with an upper bound of 1 bit of complexity. Rats and mice will form the  
470 association after a single experience lasting only seconds, and alter their behavior over several  
471 hours (*Fanselow and Bolles, 1979; Bourchuladze et al., 1994*). This is an adaptive warning  
472 system to deal with life-threatening events, and rapid learning here has a clear survival value.

473 Animals are particularly adept at learning a new association between an odor and food. For  
474 example bees will extend their proboscis in response to a new odor after just one pairing trial  
475 where the odor appeared together with sugar (*Bitterman et al., 1983*). Similarly rodents will  
476 start digging for food in a scented bowl after just a few pairings with that odor (*Cleland et al.,*  
477 *2009*). Again, these are 1-bit tasks learned rapidly after one or a few experiences.

478 By comparison the tasks that a mouse performs in the labyrinth are more complex. For  
479 example, the path from the maze entrance to the water port involves 6 junctions, each with 3  
480 options. At a minimum 6 different contexts must be mapped correctly into one of 3 actions  
481 each, which involves  $6 \cdot \log_2 3 = 9.5$  bits of complexity. The animals begin to execute perfect  
482 paths from the entrance to the water port well within the first hour (*Figure 2C, Figure 3B*).  
483 At a later stage during the night the animal learns to walk direct paths to water from many  
484 different locations in the maze (*Figure 4*); by this time it has consumed 10-20 rewards. In  
485 the limit, if the animal could turn correctly towards water from each of 63 junctions in the  
486 maze, it would have learned  $63 \cdot \log_2 3 = 100$  bits. Conservatively we estimate that the animals  
487 have mastered 10-20 bits of complexity based on 10-20 reward experiences within an hour of  
488 time spent in the maze. Note this considers only information about the water port and ignores  
489 whatever else the animals are learning about the maze during their incessant exploratory forays.  
490 These numbers align well with classic experiments on rats in diverse mazes and problem boxes

491 *Munn (1950a)*. Although those tasks come in many varieties, a common theme is that ~10  
492 successful trials are sufficient to learn ~10 decisions (*Woodrow, 1942*).

493 In a different corner of the speed-complexity space are the many 2-alternative-forced-choice  
494 (2AFC) tasks in popular use today. These tend to be 1-bit tasks, for example the monkey should  
495 flick its eyes to the left when visual motion is to the left (*Newsome and Pare, 1988*), or the  
496 mouse should turn a steering wheel to the right when a light appears on the left (*Burgess et al.,*  
497 *2017*). Yet the animals take a long time to learn these simple tasks. For example, the mouse  
498 with the steering wheel requires about 10,000 experiences before performance saturates. It  
499 never gets particularly good, with a typical hit rate only 2/3 of the way from random to perfect.  
500 All this training takes 3-6 weeks; in the case of monkeys several months. The rate of learning,  
501 measured in task complexity per unit time, is surprisingly low: < 1 bit/month compared to ~10  
502 bits/h observed in the labyrinth. The difference is a factor of 6,000. Similarly when measured  
503 in complexity learned per reward experience: The 2AFC mouse may need 5,000 rewards to  
504 learn a contingency table with 1 bit complexity, the mouse in the maze needs ~10 rewards to  
505 learn 10 bits. Given these differences in learning rate spanning many orders of magnitude,  
506 it seems likely that whatever neural process underlies ultra-slow 2AFC learning is different  
507 from the implementation of fast learning in the labyrinth and other complex environments.  
508 Furthermore, the ultra-slow mode of learning may have little relevance for an animal's natural  
509 condition. In the month that the 2AFC mouse requires to finally report the location of a light,  
510 its relative in the wild has developed from a baby to having its own babies. Along the way, that  
511 wild mouse had to make many decisions, often involving high stakes, without the benefit of  
512 10,000 trials of practice.

### 513 **Sudden insight**

514 The dynamics of the learning process are often conceived as a continuously growing associ-  
515 ation between stimuli and actions, with each reinforcing experience making an infinitesimal  
516 contribution. The reality can be quite different. When a child first learns to balance on a bicycle,  
517 performance goes from abysmal to astounding within a few seconds. The timing of such a  
518 discontinuous step in performance seems impossible to predict but easy to recognize after the  
519 fact.

520 From the early days of animal learning experiments there have been warnings against the  
521 tendency to average learning curves across subjects (*Krechevsky, 1932; Estes, 1956*). The  
522 average of many discontinuous curves will certainly look continuous and incremental, but that  
523 reassuring shape may miss the essence of the learning process. A recent reanalysis of many  
524 Pavlovian conditioning experiments suggested that discontinuous steps in performance are the  
525 rule rather than the exception (*Gallistel et al., 2004*). Here we found that the same applies to  
526 navigation in a complex labyrinth. While the average learning curve presents like a continuous  
527 function (*Figure 3B*), the individual records of water rewards show that each animal improves  
528 rather quickly but at different times (*Figure 3A*).

529 Owing to the unstructured nature of the experiment, the mouse may adopt different policies  
530 for getting to the water port. In at least half the animals we observed a discontinuous change  
531 in that policy, namely when the animal started using efficient direct paths within the maze  
532 (*Figure 4, Figure 4-figure supplement 2*). This second switch happened considerably after  
533 the animal started collecting rewards, and did not greatly affect the reward rate. Furthermore,  
534 the animals never reverted to the less efficient policy, just as a child rarely unlearns to balance  
535 a bicycle.

536 Presumably this switch in performance reflects some discontinuous change in the animal's  
537 internal model of the maze, what Tolman called the "cognitive map" (*Tolman, 1948; Behrens*  
538 *et al., 2018*). In the unrewarded animals we could not detect any discontinuous change in the

539 use of long paths. However, as Tolman argued, those animals may well acquire a sophisticated  
540 cognitive map that reveals itself only when presented with a concrete task, like finding water.  
541 Future experiments will need to address this. The discontinuous changes in performance pose  
542 a challenge to conventional models of reinforcement learning, in which reward events are the  
543 primary driver of learning and each event contributes an infinitesimal update to the action  
544 policy. It will also be important to model the acquisition of distinct kinds of knowledge that  
545 contribute to the same behavior, like the location of the target and efficient routes to approach  
546 it.

### 547 **Exploratory behavior**

548 By all accounts the animals spent a large fraction of the night exploring the maze (*Figure 1–*  
549 *figure supplement 2*). The water-deprived animals continued their forays into the depths of  
550 the maze long after they had found the water port and learned to exploit it regularly. The sated  
551 animals experienced no overt reward from the maze, yet they likewise spent nearly half their  
552 time in that environment. As has been noted many times, animals – like humans – derive some  
553 form of intrinsic reward from exploration (*Berlyne, 1960*). Some have suggested that there  
554 exists a homeostatic drive akin to hunger and thirst that elicits the information-seeking activity,  
555 and that the drive is in turn sated by the act of exploration (*Hughes, 1997*). If this were the  
556 case, then the drive to explore should be weakest just after an episode of exploration, much as  
557 the drive for food-seeking is weaker after a big meal.

558 Our observations are in conflict with this notion. The animal is most likely to enter the maze  
559 within the first minute of its return to the cage (*Figure 1–figure supplement 3*), a strong trend  
560 that runs opposite to the prediction from satiation of curiosity. Several possible explanations  
561 come to mind: (1) On these very brief visits to the cage the animal may just want to certify  
562 that the exit route to the safe environment still exists, before continuing with exploration of the  
563 maze. (2) The temporal contrast between the boredom of the cage and the mystery of the maze  
564 is highest right at the moment of exit from the maze, and that may exert pressure to re-enter the  
565 maze. Understanding this in more detail will require dedicated experiments. For example, one  
566 could deliberately deprive the animals of access to the maze for some hours, and test whether  
567 that results in an increased drive to explore, as observed for other homeostatic drives around  
568 eating, drinking, and sleeping.

569 When left to their own devices, mice choose to spend much of their time engaged in  
570 exploration. In the maze, mice follow the collection of a water drop with a bout of exploration  
571 90% of the time (*Figure 6B*). One wonders how that affects their actions when they are strapped  
572 into a rigid behavior machine, like a 2AFC choice box. Presumably the drive to explore persists,  
573 perhaps more so because the forced environment is so unpleasant. And within the confines  
574 of the two alternatives, the only act of exploration the mouse has left is to give the wrong  
575 answer. This would manifest as an unexpectedly high error rate on unambiguous stimuli,  
576 sometimes called the "lapse rate" (*Carandini and Churchland, 2013*). The fact that the lapse  
577 rate decreases only gradually over weeks to months of training (*Burgess et al., 2017*) suggests  
578 that it is difficult to crush the animal's drive to explore.

579 The animals in our experiments had never been presented with a maze environment, yet they  
580 quickly settled into a steady mode of exploration. Once a mouse progressed beyond the first  
581 intersection it typically entered deep into the maze to one or more end nodes (*Figure 5*). Within  
582 50 s of the first entry the animals adopted a steady speed of locomotion that they would retain  
583 throughout the night (*Figure 2–figure supplement 1*). Within 250 s of first contact with the  
584 maze the average animal already spent 50% of its time there (*Figure 1–figure supplement 2*).  
585 Contrast this with a recent study of "free exploration" in an exposed arena: Those animals  
586 required several hours before they even completed one walk around the perimeter (*Fonio et al.,*

587 2009). Here the drive to explore is clearly pitted against fear of the open space, which may not  
588 be conducive to observing exploration *per se*.

589 The persistence of exploration throughout the entire duration of the experiment suggests  
590 that the animals are continuously surveying the environment, perhaps expecting new features  
591 to arise. These surveys are quite efficient: The animals cover all parts of the maze much faster  
592 than expected from a random walk (Figure 7). Effectively they avoid re-entering territory they  
593 surveyed just recently. It is often assumed that this requires some global memory of places  
594 visited in the environment (Nagy *et al.*, 2020; Olton, 1979). Such memory would have to  
595 persist for a long time: Surveying half of the available end nodes typically required 450 turning  
596 decisions. However, we found that a global long-term memory is not needed to explain the  
597 efficient search. The animals seem to be governed by a set of local turning biases that require  
598 memory only of the most recent decision and no knowledge of location (Figure 8). These local  
599 biases alone can explain most of the character of exploration without any global understanding  
600 or long-term memory. Incidentally, they also explain other seemingly global aspects of the  
601 behavior, for example the systematic preference that the mice have for the outer rather than the  
602 inner regions of the maze (Figure 9). Of course, this argument does not exclude the presence  
603 of a long-term memory, which may reveal itself in some other feature of the behavior.

604 Perhaps the most remarkable aspect of these biases is how similar they are across all 19 mice  
605 studied here, regardless of whether the animal experienced water rewards or not (Figure 8B,  
606 Figure 8-figure supplement 1), and independent of the sex of the mouse. The four decision  
607 probabilities were identical across individuals to within a standard deviation of  $<0.03$ . We  
608 cannot think of a trivial reason why this should be so. For example the two biases for forward  
609 motion (Figure 8B left) are poised halfway between the value for a random walk ( $p = 2/3$ ) and  
610 certainty ( $p = 1$ ). At either of those extremes, simple saturation might lead to a reproducible  
611 value, but not in the middle of the range. Why do different animals follow the exact same  
612 decision rules at an intersection between tunnels? Given that tunnel systems are part of the  
613 mouse's natural ecology, it is possible that those rules are innate and determined genetically.  
614 Indeed the rules by which mice build tunnels have a strong genetic component (Weber *et al.*,  
615 2013), so the rules for using tunnels may be written in the genes as well. The high precision  
616 with which one can measure those behaviors even in a single night of activity opens the way to  
617 efficient comparisons across genotypes, and also across animals with different developmental  
618 experience.

619 Finally, after mice discover the water port and learn to access it from many different points  
620 in the maze (Figure 4) they are presumably eager to discover other things. In ongoing work we  
621 installed three water ports (visible in the videos accompanying this article) and implemented a  
622 rule that activates the three ports in a cyclic sequence. Mice discovered all three ports rapidly  
623 and learned to visit them in the correct order. Future experiments will have to raise the bar on  
624 what the mice are expected to learn in a night.

## 625 Mechanisms of navigation

626 How do the animals navigate when they perform direct paths to the water port or to the exit?  
627 This question will require future directed experiments, but we can exclude a few candidate  
628 explanations based on observations so far. Early workers already concluded that rodents in a  
629 maze will use whatever sensory cues and tricks are available to accomplish their tasks (Munn,  
630 1950b). Our maze was designed to restrict those options somewhat.

631 To limit the opportunity for visual navigation, the floor and walls of the maze are visually  
632 opaque. The ceiling is transparent, but the room is kept dark except for infrared illuminators.  
633 Even if the animal finds enough light, the goals (water port or exit) are invisible within the  
634 maze except from the immediately adjacent corridor. There are no visible beacons that would

635 identify the goal.

636 With regard to the sense of touch and kinesthetics, the maze was constructed for maximal  
637 symmetry. At each level of the binary tree all the junctions have locally identical geometry,  
638 with intersecting corridors of the same length. In practice the animals may well detect some  
639 inadvertent cues, like an unusual drop of glue, that could identify one node from another.

640 The role of odors deserves particular attention because the mouse may use them both  
641 passively and actively. Does the animal first find the water port by following the smell of water?  
642 Probably not. For one, the port only emits a single drop of water when triggered by a nose  
643 poke. Second, we observed many instances where the animal is in the final corridor adjacent to  
644 the water port yet fails to discover it. The initial discovery seems to occur via touch. Finally, in  
645 exploratory experiments with a different labyrinth we used an open bowl of water as a reward;  
646 the correct path required traveling away from the water for many turns. The mice discovered  
647 this path readily.

648 Do the animals lay down an odor trail to mark the location of the water port? This could  
649 be a simple algorithm for externalized cognition: After finding a location to which you want  
650 to return, urinate on the floor on the way back, gradually reducing the emissions along the  
651 way. This would establish an odor gradient that you can follow uphill on the next foray. Taking  
652 advantage of the symmetric layout of the maze, we rotated the entire maze by 180 degrees  
653 between one foray and the next, while leaving both the entrance and the water port in the  
654 same absolute location. Following that rotation the animal did in fact make a few visits to the  
655 rotated port location, but then quickly exploited the real water port again, without any apparent  
656 re-learning. Through further experiments of this type one can hope to test not only the role of  
657 odor tracks, but also that of inadvertent construction details that might identify each node.

658 Finally we considered an algorithm that is often invoked for animals moving in an open  
659 arena: vector-based navigation (*Wehner et al., 1996*). Once the animal discovers a target, it  
660 keeps track of that target's heading and distance using a path integrator. When it needs to return  
661 to the target it follows the heading vector and updates heading and distance until it arrives. This  
662 is clearly not what the mice do in the labyrinth, as seen most easily by considering the "home  
663 runs" back to the exit at the end of a bout. Here the target, namely the exit, is known from  
664 the start of the bout, because the animal enters through the same hole. At the end of the bout,  
665 when the mouse decides to exit from the maze, does it follow the heading vector to the exit?  
666 *Figure 5A* shows the 13 locations from which mice returned in a direct path to the exit on their  
667 very first foray. None of these locations is compatible with heading-based navigation: In each  
668 case an animal following the heading to the exit would get stuck in a different end node first.

669 Future directed experiments will serve to further exclude candidate mechanisms of naviga-  
670 tion. Since the animals normally solve the task within an hour, one can test a new hypothesis  
671 quite efficiently. Understanding what mechanisms they use will then inform thinking about the  
672 algorithm for learning, and about the neuronal mechanisms that implement it.

## 673 **Methods and Materials**

### 674 **Experimental design**

675 The goal of the study was to observe mice as they explored a complex environment for the  
676 first time, with little or no human interference and no specific instructions. In preliminary  
677 experiments we tested several labyrinth designs and water reward schedules. Eventually we  
678 settled on the protocol described here, and tested 20 mice in rapid succession. Each mouse was  
679 observed only over a 7-hour period during the first night it encountered the labyrinth.

### 680 **Maze construction**

681 The maze measured ~24 x 24 x 2 inches; for manufacture we used materials specified in inches,  
682 so dimensions are quoted in those non-SI units where appropriate. The ceiling was made of 0.5  
683 inch clear acrylic. Slots of 1/8 inch width were cut into this plate on a 1.5 inch grid. Pegged  
684 walls made of 1/8 inch infrared-transmitting acrylic (opaque in the visible spectrum, ePlastics)  
685 were inserted into these slots and secured with a small amount of hot glue. The floor was a sheet  
686 of infrared-transmitting acrylic, supported by a thicker sheet of clear acrylic. The resulting  
687 corridors (1-1/8 inches wide) formed a 6-level binary tree with T-junctions and progressive  
688 shortening of each branch, ranging from ~12 inch to 1.5 inch (*Figure 1* and *Figure 2*). A single  
689 end node contained a 1.5 cm circular opening with a water delivery port (described below).  
690 The maze included provision for two additional water ports not used in the present report. In  
691 unrewarded experiments the water port was covered with an acrylic plate. Once per week the  
692 maze was submerged in cage cleaning solution. Between different animals the floor and walls  
693 were cleaned with ethanol.

### 694 **Reward delivery system**

695 The water reward port was controlled by a Matlab script on the main computer through an  
696 interface (Sanworks Bpod State Machine r1). Rewards were triggered when the animal's nose  
697 broke the IR beam in the water port (Sanworks Port interface + valve). The interface briefly  
698 opened the water valve to deliver ~30  $\mu$ L of water and flashed an infrared LED mounted outside  
699 the maze for 1 s. This served to mark reward events on the video recording. Following each  
700 reward, the system entered a time-out period for 90 s, during which the port did not provide  
701 further reward.

### 702 **Cage and connecting passage**

703 The entrance to the maze was connected to an otherwise normal mouse cage by red plastic  
704 tubing (3 cm dia, 1 m long). The cage contained food, bedding, nesting material, and in the  
705 case of unrewarded experiments also a normal water bottle.

### 706 **Animals and treatments**

707 All mice were C57Bl6/J animals (Jackson Labs) between the ages of 45 and 98 days (mean 62  
708 days). Both sexes were used: 4 males and 6 females in the rewarded experiments, 5 males and  
709 4 females in the unrewarded experiments. For water deprivation, the animal was transferred  
710 from its home cage (generally group-housed) to the maze cage ~22 h before the start of the  
711 experiment. Non-deprived animals were transferred minutes before the start. All procedures  
712 were performed in accordance with institutional guidelines and approved by the Caltech IACUC.

### 713 **Video recording**

714 All data reported here were collected over the course of 7 hours during the dark portion of  
715 the animal's light cycle. Video recording was initiated a few seconds prior to connecting the  
716 tunnel to the maze. Videos were recorded by an OpenCV python script controlling a single

717 webcam (Logitech C920) located ~1 m below the floor of the maze. The maze and access tube  
718 were illuminated by multiple infrared LED arrays (center wavelength 850 nm). Three of these  
719 lights illuminated the maze from below at a 45 degree angle, producing contrast to resolve  
720 the animal's foot pads. The remaining lights pointed at the ceiling of the room to produce  
721 backlight for a sharp outline of the animal.

### 722 **Animal tracking**

723 A version of DeepLabCut (*Nath et al., 2019*) modified to support gray-scale processing was  
724 used to track the animal's trajectory, using key points at the nose, feet, tail base and mid-body.  
725 All subsequent analysis was based on the trajectory of the animal's nose, consisting of positions  
726  $x(t)$  and  $y(t)$  in every video frame.

### 727 **Rates of transition between cage and maze**

728 This section relates to *Figure 1–figure supplement 3*. We entertained the hypothesis that the  
729 animals become “thirsty for exploration” as they spend more time in the cage. In that case one  
730 would predict that the probability of entering the maze in the next second will increase with  
731 time spent in the cage. One can compute this probability from the distribution of residency  
732 times in the cage, as follows:

733 Say  $t = 0$  when the animal enters the cage. The probability density that the animal will  
734 next leave the cage at time  $t$  is

$$p(t) = e^{-\int_0^t r(t') dt'} r(t) \quad (2)$$

735 where  $r(t)$  is the instantaneous rate for entering the maze. So

$$\int_0^t p(t') dt' = 1 - e^{-\int_0^t r(t') dt'} \quad (3)$$

$$\int_0^t r(t') dt' = -\ln\left(1 - \int_0^t p(t') dt'\right) \quad (4)$$

736 This relates the cumulative of the instantaneous rate function to the cumulative of the  
737 observed transition times. In this way we computed the rates

$$r_m(t) = \text{rate of entry into the maze as a function of time spent in the cage} \quad (5)$$

$$r_c(t) = \text{rate of entry into the cage as a function of time spent in the maze} \quad (6)$$

738 The rate of entering the maze is highest at short times in the cage (*Figure 1–figure supple-*  
739 *ment 3A*). It peaks after ~15 s in the cage and then declines gradually by a factor of 4 over  
740 the first minute. So the mouse is most likely to enter the maze just after it returns from there.  
741 This runs opposite to the expectation from a homeostatic drive for exploration, which should  
742 be sated right after the animal returns. We found no evidence for an increase in the rate at  
743 late times. These effects were very similar in rewarded and unrewarded groups and in fact the  
744 tendency to return early was seen in every animal.

745 By contrast the rate of exiting the maze is almost perfectly constant over time (*Figure 1–*  
746 *figure supplement 3B*). In other words the exit from the maze appears like a constant rate

747 Poisson process. There is a slight elevation of the rate at short times among rewarded animals  
748 (*Figure 1–figure supplement 3B* top). This may come from the occasional brief water runs  
749 they perform. Another strange deviation is an unusual number of very short bouts (duration  
750 2-12 s) among unrewarded animals (*Figure 1–figure supplement 3B* bottom). These are brief  
751 excursions in which the animal runs to the central junction, turns around, and runs to the exit.  
752 Several animals exhibited these, often several bouts in a row, and at all times of the night.

### 753 **Reduced trajectories**

754 From the raw nose trajectory we computed two reduced versions. First we divided the maze  
755 into discrete “cells”, namely the squares the width of a corridor that make up the grid of the  
756 maze. At any given time the nose is in one of these cells and that time series defines the **cell**  
757 **trajectory**.

758 At a coarser level still one can ask when the animal passes through the nodes of the binary  
759 tree, which are the decision points in the maze. The special cells that correspond to the nodes  
760 of the tree are those at the center of a T-junction and those at the leaves of the tree. We marked  
761 all the times when the trajectory  $(x(t), y(t))$  entered a new node cell. If the animal leaves a  
762 node cell and returns to it before entering a different node cell, that is not considered a new  
763 node. This procedure defines a discrete **node sequence**  $s_i$  and corresponding arrival times at  
764 those nodes  $t_i$ . We call the transition between two nodes a “step”. Much of the analysis in this  
765 paper is derived from the animal’s node sequence. The median mouse performed 16,192 steps  
766 in the 7 h period of observation (mean = 15,257; SD = 3,340).

767 In *Figure 4* and *Figure 5* we count the occurrence of **direct paths** leading to the water  
768 port (a “water run”) or to the exit (a “home run”). A direct path is a node sequence without any  
769 reversals. *Figure 3–figure supplement 1* illustrates some examples.

770 If the animal makes one wrong step from the direct path, that step needs to be backtracked,  
771 adding a total of two steps to the length of the path. If further errors occur during backtracking  
772 they need to be corrected as well. The binary maze contains no loops, so the number of errors  
773 is directly related to the length of the path:

$$\text{Errors} = (\text{Length of path} - \text{Length of direct path})/2. \quad (7)$$

### 774 **Statistics of sudden insight**

775 In *Figure 4* one can distinguish two events: First the animal finds the water port and begins  
776 to collect rewards at a steady rate: this is when the green curve rises up. At a later time the  
777 long direct paths to the water port become much more frequent than to the comparable control  
778 nodes: this is when the red and blue curves diverge. For almost all animals these two events are  
779 well separated in time (*Figure 4–figure supplement 1*). In many cases the rate of long paths  
780 seems to change discontinuously: a sudden change in slope of the curve.

781 A sudden increase in the rate of certain behavioral events would suggest a sudden change  
782 in the animal’s strategy. Here we analyze the evidence for a sudden change in the time series  
783 of events. We focused on the occurrence of long paths to water, and compared two models:  
784 The “step model” supposes that the events are generated at a constant rate  $r_{\text{before}}$  up to some  
785 time  $t_s$  and then with a higher rate  $r_{\text{after}}$ ,

$$r_{\text{step}}(t) = \begin{cases} r_{\text{before}}, & t < t_s \\ r_{\text{after}}, & t > t_s \end{cases} \quad (8)$$

786 The “ramp model” supposes instead that the rate of the events increases in a graded fashion as



787 a quadratic function of time,

$$r_{\text{ramp}}(t) = a + bt + ct^2 \quad (9)$$

788 Note that each of the models has 3 free parameters. If the ramp model provides a better fit that  
789 speaks against an interpretation in terms of “sudden insight”.

790 The data are a set of  $n$  event times  $t_i$  in the observation interval  $[0, T]$ . We model the event  
791 train as an inhomogeneous Poisson point process with instantaneous rate  $r(t)$ . The likelihood  
792 of the data given the rate function  $r(t)$  is

$$L[r(t)] = e^{-\int_0^T r(t) dt} \prod_i r(t_i) \quad (10)$$

793 and the log likelihood is

$$\ln L = \sum_i \ln r(t_i) - \int_0^T r(t) dt \quad (11)$$

794 For each of the 10 rewarded mice, we maximized  $\ln L$  over the parameters under both the  
795 step model and the ramp model (*Figure 4-figure supplement 2*). For the step model, one can  
796 optimize analytically with respect to the initial rate  $r_{\text{before}}$  and final rate  $r_{\text{after}}$ , and express the  
797 log likelihood purely as a function of the step time  $t_s$ :

$$\ln L(t_s) = m(t_s) \ln m(t_s) + (n - m(t_s)) \ln (n - m(t_s)) - m(t_s) \ln t_s \\ - (n - m(t_s)) \ln (T - t_s) - n \quad (12)$$

798 where  $m(t_s)$  is the number of events prior to the step time  $t_s$ . For the ramp model the optimization  
799 was done by numerical search.

800 For 8 of 10 mice, the step model had a higher likelihood than the ramp model. For 5 mice,  
801 the step model was more than 25 times as likely as the ramp model. In these 5 cases the change  
802 in the frequency of long paths was large (5- to 1640-fold). The time of this step change in the  
803 animal’s behavior ranged from 1280 to 2580 s in the maze, and could be identified with an  
804 uncertainty of 80 to 310 s. This uncertainty is quoted as the standard deviation of the likelihood  
805 expressed as a function of the step time (*Equation 12*).

## 806 Efficiency of exploration

807 The goal of this analysis is to measure how effectively the animal surveys all the end nodes of  
808 the maze. The specific question is: In a string of  $n$  end nodes that the animal samples, how  
809 many of these are distinct? On average how does the number of distinct nodes  $d$  increase with  
810  $n$ ? This was calculated as follows:

811 We restricted the animal’s node trajectory ( $s_i$ ) to clips of exploration mode, excluding the  
812 direct paths to the water port or the exit. All subsequent steps were applied to these clips, then  
813 averaged over clips. Within each clip we marked the sequence of end nodes ( $e_i$ ). We slid a  
814 window of size  $n$  across this sequence and counted the number of distinct nodes  $d$  in each  
815 window. Then we averaged  $d$  over all windows in all clips. Then we repeated that for a wide  
816 range of  $n$ . The resulting  $d(n)$  is plotted in the figures reporting new nodes vs nodes visited  
817 (*Figure 7A,B and Figure 8C*).

818 For a summary analysis we fitted the curves of  $d(n)$  with a 2-parameter function:

$$d(n) \approx 64 \left( 1 - \frac{1}{1 + \frac{z+bz^3}{1+b}} \right) \quad (13)$$

819 where

$$z = n/a. \quad (14)$$

820 The parameter  $a$  is the number of visits  $n$  required to survey half of the end nodes, whereas  $b$   
 821 reflects a relative acceleration in discovering the last few end nodes. This function was found  
 822 by trial and error and produces absurdly good fits to the data (*Figure 7–figure supplement 1*).  
 823 The values quoted in the text for efficiency of exploration are  $E = 32/a$  (*Equation 1*).

824 The value of  $b$  was generally small ( $\sim 0.1$ ) with no difference between rewarded and unre-  
 825 warded animals. It declined slightly over the night (*Figure 7–figure supplement 1B*), along  
 826 with the decline in  $a$  (*Figure 7C*).

### 827 Biased random walk

828 For the analysis of *Figure 8* we considered only the parts of the trajectory during ‘exploration’  
 829 mode. Then we parsed every step between two nodes in terms of the type of action it represents.  
 830 Note that every link between nodes in the maze is either a ‘left branch’ or a ‘right branch’,  
 831 depending on its relationship to the parent T-junction. Therefore there are 4 kinds of action:

- 832 •  $a = 0$ : ‘in left’, take a left branch into the maze
- 833 •  $a = 1$ : ‘in right’, take a right branch into the maze
- 834 •  $a = 2$ : ‘out left’, take a left branch out of the maze
- 835 •  $a = 3$ : ‘out right’, take a right branch out of the maze

836 At any given node some actions are not available, for example from an end node one can  
 837 only take one of the ‘out’ actions.

838 To compute the turning biases we considered every T-junction along the trajectory and  
 839 correlated the action  $a_0$  that led into that node with the subsequent action  $a_1$ . By tallying the  
 840 action pairs  $(a_0, a_1)$  we computed the conditional probabilities  $p(a_1|a_0)$ . Then the 4 biases are  
 841 defined as

$$P_{SF} = \frac{p(0|0) + p(0|1) + p(1|0) + p(1|1)}{p(0|0) + p(0|1) + p(1|0) + p(1|1) + p(2|0) + p(3|1)} \quad (15)$$

$$P_{SA} = \frac{p(0|1) + p(1|0)}{p(0|0) + p(0|1) + p(1|0) + p(1|1)} \quad (16)$$

$$P_{BF} = \frac{p(0|3) + p(1|2) + p(2|2) + p(2|3) + p(3|2) + p(3|3)}{p(0|3) + p(1|2) + p(2|2) + p(2|3) + p(3|2) + p(3|3) + p(0|2) + p(1|3)} \quad (17)$$

$$P_{BS} = \frac{p(2|2) + p(2|3) + p(3|2) + p(3|3)}{p(0|3) + p(1|2) + p(2|2) + p(2|3) + p(3|2) + p(3|3)} \quad (18)$$

### 842 Models of decisions during exploration

843 The general approach is to develop a model that assigns probabilities to the animal’s next  
 844 action, namely which node it will move to next, based on its recent history of actions. All the  
 845 analysis was restricted to the animal’s ‘exploration’ mode and to the 63 nodes in the maze that  
 846 are T-junctions. During the ‘drink’ and ‘leave’ modes the animal’s next action is predictable.  
 847 Similarly when it finds itself at one of the 64 end nodes it only has one action available.

848 For every mouse trajectory we split the data into 5 segments, trained the model on 80% of  
849 the data, and tested it on 20%, averaging the resulting cross-entropy over the 5 possible splits.  
850 Each segment was in turn composed of parts of the trajectory sampled evenly throughout the  
851 7-h experiment, so as to average over the small changes in the course of the night. The model  
852 was evaluated by the cross-entropy between the predictions and the animal's true actions. If  
853 one had an optimal model of behavior, the result would reveal the animal's true source entropy.

#### 854 Fixed depth Markov chain

855 To fit a model with fixed history depth  $k$  to a measured node sequence  $(s_t)$ , we evaluated  
856 all the substrings in that sequence of length  $(k + 1)$ . At any given time  $t$ , the  $k$ -string  $\mathbf{h}_t =$   
857  $(s_{t-k+1}, \dots, s_t)$  identifies the history of the animal's  $k$  most recent locations. The current state  
858  $s_t$  is one of 63 T-junctions. Each state is preceded by one of 3 possible states. So the number  
859 of history strings is  $63 \cdot 3^{k-1}$ . The 2-string  $(s_t, s_{t+1})$  identifies the next action  $a_{t+1}$ , which can  
860 be 'in left', 'in right', or 'out', corresponding to the 3 branches of the T junction. Tallying  
861 the history strings with the resulting actions leads to a contingency table of size  $63 \cdot 3^{k-1} \times 3$ ,  
862 containing

$$n(\mathbf{h}, a) = \text{number of times history } \mathbf{h} \text{ leads to action } a \quad (19)$$

863 Based on these sample counts we estimated the probability of each action  $a$  conditional on the  
864 history  $\mathbf{h}$  as

$$p(a | \mathbf{h}) = \frac{n(\mathbf{h}, a) + 1}{\sum_{a'} n(\mathbf{h}, a') + 3} \quad (20)$$

865 This amounts to additive smoothing with a pseudocount of 1, also known as "Laplace smooth-  
866 ing". These conditional probabilities were then used in the testing phase to predict the action  
867 at time  $t$  based on the preceding history  $\mathbf{h}_t$ . The match to the actually observed actions  $a_t$  was  
868 measured by the cross-entropy

$$H = - \sum_t \log_2 p(a_t | \mathbf{h}_t) \quad (21)$$

#### 869 Variable depth Markov chain

870 As one pushes to longer histories, i.e. larger  $k$ , the analysis quickly becomes data-limited,  
871 because the number of possible histories grows exponentially with  $k$ . Soon one finds that  
872 the counts for each history-action combination drop to where one can no longer estimate  
873 probabilities correctly. In an attempt to offset this problem we pruned the history tree such that  
874 each surviving branch had more than some minimal number of counts in the training data.

875 As expected, this model is less prone to over-fitting and degrades more gently as one  
876 extends to longer histories (*Figure 10-figure supplement 1A*). The lowest cross-entropy was  
877 obtained with an average history length of  $\sim 4.0$  but including some paths of up to length 6.  
878 Of all the algorithms we tested, this produced the lowest cross-entropies, although the gains  
879 relative to the fixed-depth model were modest (*Figure 10-figure supplement 1C*).

#### 880 Pooling across symmetric nodes in the maze

881 Another attempt to increase the counts for each history involved pooling counts over multiple  
882 T-junctions in the maze that are closely related by symmetry. For example, all the T-junctions at  
883 the same level of the binary tree look locally similar, in that they all have corridors of identical  
884 length leading from the junction. If one supposes that the animal acts the same way at each  
885 of those junctions, one would be justified in pooling across these nodes, leading to a better

886 estimate of the action probabilities, and perhaps less over-fitting. This particular procedure  
887 was unsuccessful, in that it produced higher cross-entropy than without pooling.

888 However, one may want to distinguish two types of junctions within a given level: L-nodes  
889 are reached by a left branch from their parent junction one level lower in the tree, R-nodes  
890 by a right branch. For example, in *Figure 3–figure supplement 1*, node 1 is L-type and node  
891 2 is R-type. When we pooled histories over all the L-nodes at a given level and separately  
892 over all the R-nodes the cross-entropy indeed dropped, by about 5% on average. This pooling  
893 greatly reduced the amount of over-fitting (*Figure 10–figure supplement 1B*), which allowed  
894 the use of longer histories, which in turn improved the predictions on test data. The benefit of  
895 distinguishing L- and R-nodes probably relates to the animal’s tendency to alternate left and  
896 right turns.

897 All the Markov model results we report are obtained using pooling over L-nodes and  
898 R-nodes at each maze level.

### 899 **Data availability**

900 All data and code needed to reproduce the figures and quoted results are available in this public  
901 repository: <https://github.com/markusmeister/Rosenberg-2021-Repository>.

### 902 **Acknowledgments**

903 Funding: This work was supported by the Simons Collaboration on the Global Brain (grant  
904 543015 to MM and 543025 to PP), by NSF award 1564330 to PP, and by a gift from Google to  
905 PP.

906 Author contributions: Conception of the study MR, TZ, PP, MM; Data collection MR, TZ;  
907 Analysis and interpretation MR, TZ, PP, MM; Drafting the manuscript MM; Revision and  
908 approval MR, TZ, PP, MM.

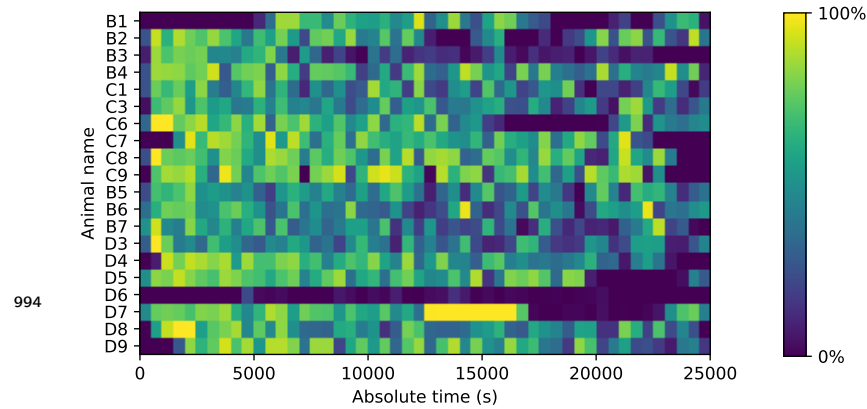
909 Competing interests: The authors declare no competing interests.

910 Data and code availability: Data and code will be available in a public repository following  
911 acceptance of the manuscript.

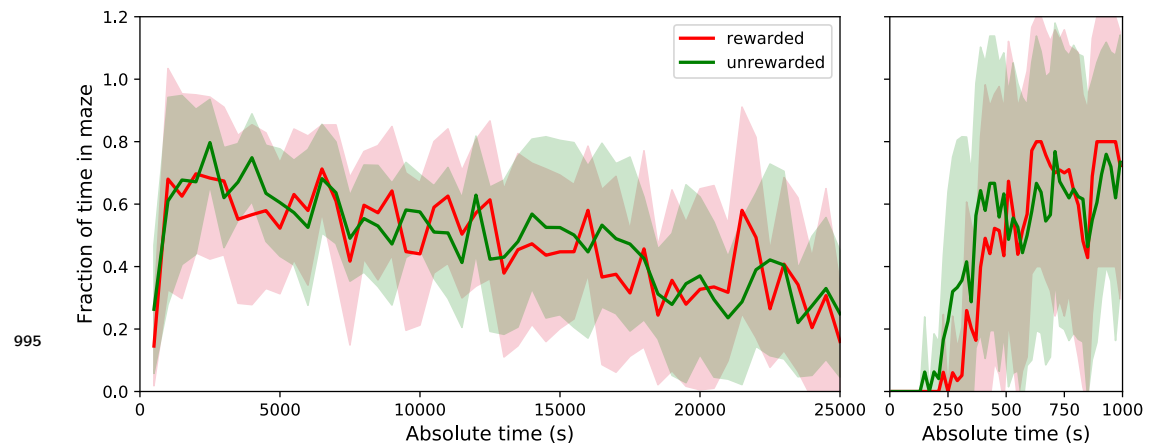
## 909 References

- 910 **Behrens TEJ**, Muller TH, Whittington JCR, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson  
911 Z. What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*. 2018 Oct;  
912 100(2):490–509. doi: [10.1016/j.neuron.2018.10.002](https://doi.org/10.1016/j.neuron.2018.10.002).
- 913 **Berlyne DE**. Conflict, Arousal, and Curiosity. New York, NY, US: McGraw-Hill Book Company; 1960.  
914 doi: [10.1037/11164-000](https://doi.org/10.1037/11164-000).
- 915 **Bitterman ME**, Menzel R, Fietz A, Schäfer S. Classical Conditioning of Proboscis Extension in Honey-  
916 bees (*Apis Mellifera*). *Journal of Comparative Psychology*. 1983; 97(2):107–119. doi: [10.1037/0735-](https://doi.org/10.1037/0735-7036.97.2.107)  
917 [7036.97.2.107](https://doi.org/10.1037/0735-7036.97.2.107).
- 918 **Bourtchuladze R**, Frenquelli B, Blendy J, Cioffi D, Schutz G, Silva AJ. Deficient Long-Term Memory  
919 in Mice with a Targeted Mutation of the cAMP-Responsive Element-Binding Protein. *Cell*. 1994  
920 Oct; 79(1):59–68. doi: [10.1016/0092-8674\(94\)90400-6](https://doi.org/10.1016/0092-8674(94)90400-6).
- 921 **Brennan PA**, Keverne EB. Neural Mechanisms of Mammalian Olfactory Learning. *Progress in*  
922 *Neurobiology*. 1997 Mar; 51(4):457–481. doi: [10.1016/s0301-0082\(96\)00069-x](https://doi.org/10.1016/s0301-0082(96)00069-x).
- 923 **Bruce HM**. An Exteroceptive Block to Pregnancy in the Mouse. *Nature*. 1959 Jul; 184:105. doi:  
924 [10.1038/184105a0](https://doi.org/10.1038/184105a0).
- 925 **Buel J**. The Linear Maze. I. "Choice-Point Expectancy," "Correctness," and the Goal Gradient. *Journal*  
926 *of Comparative Psychology*. 1934; 17(2):185–199. doi: [10.1037/h0072346](https://doi.org/10.1037/h0072346).
- 927 **Burgess CP**, Lak A, Steinmetz NA, Zátka-Haas P, Bai Reddy C, Jacobs EAK, Linden JF, Paton JJ,  
928 Ranson A, Schröder S, Soares S, Wells MJ, Wool LE, Harris KD, Carandini M. High-Yield Methods  
929 for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. *Cell Reports*. 2017 Sep;  
930 20(10):2513–2524. doi: [10.1016/j.celrep.2017.08.047](https://doi.org/10.1016/j.celrep.2017.08.047).
- 931 **Carandini M**, Churchland AK. Probing Perceptual Decisions in Rodents. *Nature Neuroscience*. 2013  
932 Jul; 16:824–31. doi: [10.1038/nn.3410](https://doi.org/10.1038/nn.3410).
- 933 **Cleland TA**, Narla VA, Boudadi K. Multiple Learning Parameters Differentially Regulate Olfactory  
934 Generalization. *Behavioral Neuroscience*. 2009 Feb; 123(1):26–35. doi: [10.1037/a0013991](https://doi.org/10.1037/a0013991).
- 935 **Estes W**. The Problem of Inference from Curves Based on Group Data. *Psychological Bulletin*. 1956;  
936 53(2):134–140. doi: [10.1037/h0045156](https://doi.org/10.1037/h0045156).
- 937 **Fanselow M**, Bolles R. Naloxone and Shock-Elicited Freezing in the Rat. *Journal of comparative and*  
938 *physiological psychology*. 1979 Sep; 93:736–44. doi: [10.1037/h0077609](https://doi.org/10.1037/h0077609).
- 939 **Fonio E**, Benjamini Y, Golani I. Freedom of Movement and the Stability of Its Unfolding in Free  
940 Exploration of Mice. *Proceedings of the National Academy of Sciences of the United States of*  
941 *America*. 2009 Dec; 106(50):21335–21340. doi: [10.1073/pnas.0812513106](https://doi.org/10.1073/pnas.0812513106).
- 942 **Gallistel CR**, Fairhurst S, Balsam P. The Learning Curve: Implications of a Quantitative Analysis.  
943 *Proceedings of the National Academy of Sciences of the United States of America*. 2004 Sep;  
944 101(36):13124–13131. doi: [10.1073/pnas.0404965101](https://doi.org/10.1073/pnas.0404965101).
- 945 **Guo ZV**, Li N, Huber D, Ophir E, Gutnisky D, Ting JT, Feng G, Svoboda K. Flow of Corti-  
946 cal Activity Underlying a Tactile Decision in Mice. *Neuron*. 2014 Jan; 81(1):179–194. doi:  
947 [10.1016/j.neuron.2013.10.020](https://doi.org/10.1016/j.neuron.2013.10.020).
- 948 **Hughes RN**. Intrinsic Exploration in Animals: Motives and Measurement. *Behavioural Processes*.  
949 1997 Dec; 41(3):213–226. doi: [10.1016/S0376-6357\(97\)00055-7](https://doi.org/10.1016/S0376-6357(97)00055-7).
- 950 **Kim SM**, Frank LM. Hippocampal Lesions Impair Rapid Learning of a Continuous Spatial Alternation  
951 Task. *PLOS ONE*. 2009 May; 4(5):e5494. doi: [10.1371/journal.pone.0005494](https://doi.org/10.1371/journal.pone.0005494).

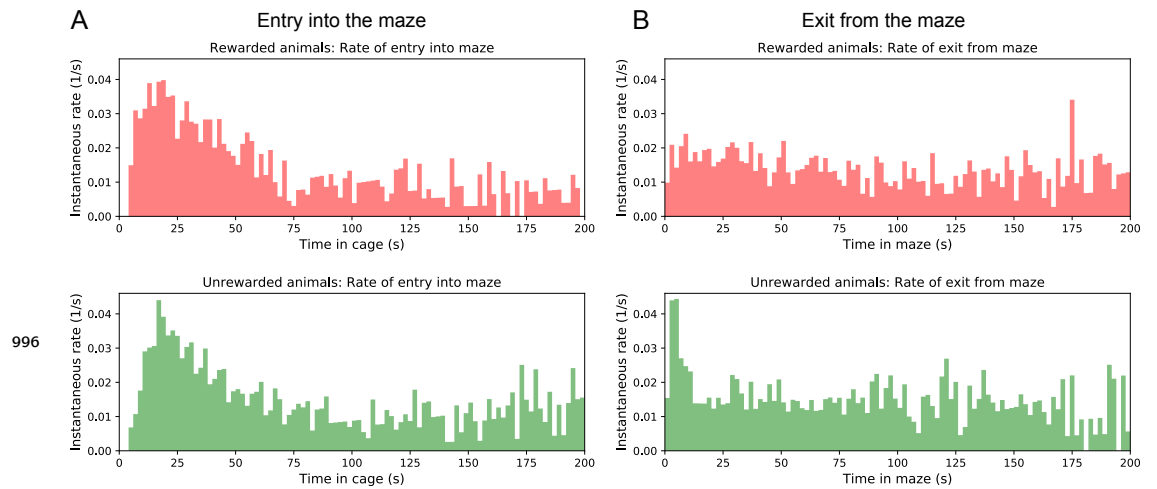
- 952 **Krechevsky I.** "Hypotheses" in Rats. *Psychological Review*. 1932; 39(6):516–532. doi:  
953 10.1037/h0073500.
- 954 **LeDoux JE.** Emotion Circuits in the Brain. *Annual Review of Neuroscience*. 2000; 23:155–184. doi:  
955 [10.1146/annurev.neuro.23.1.155](https://doi.org/10.1146/annurev.neuro.23.1.155).
- 956 **Munn NL.** The Learning Process. In: *Handbook of Psychological Research on the Rat; an Introduction*  
957 *to Animal Psychology* Oxford, England: Houghton Mifflin; 1950.p. 226–288.
- 958 **Munn NL.** The Role of Sensory Processes in Maze Behavior. In: *Handbook of Psychological Research*  
959 *on the Rat; an Introduction to Animal Psychology* Oxford, England: Houghton Mifflin; 1950.p.  
960 181–225.
- 961 **Nagy M, Horicsányi A, Kubinyi E, Couzin ID, Vásárhelyi G, Flack A, Vicsek T.** Synergistic Benefits  
962 of Group Search in Rats. *Current Biology*. 2020 Sep; doi: [10.1016/j.cub.2020.08.079](https://doi.org/10.1016/j.cub.2020.08.079).
- 963 **Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW.** Using DeepLabCut for 3D Markerless  
964 Pose Estimation across Species and Behaviors. *Nature Protocols*. 2019 Jul; 14(7):2152–2176. doi:  
965 [10.1038/s41596-019-0176-0](https://doi.org/10.1038/s41596-019-0176-0).
- 966 **Newsome WT, Pare EB.** A Selective Impairment of Motion Perception Following Lesions of the  
967 Middle Temporal Visual Area (MT). *Journal of Neuroscience*. 1988 Jun; 8(6):2201–2211. doi:  
968 [10.1523/JNEUROSCI.08-06-02201.1988](https://doi.org/10.1523/JNEUROSCI.08-06-02201.1988).
- 969 **Olton D.** Mazes, Maps, and Memory. *American Psychologist*. 1979; 34(7):583–596. doi: [10.1037/0003-](https://doi.org/10.1037/0003-066X.34.7.583)  
970 [066X.34.7.583](https://doi.org/10.1037/0003-066X.34.7.583).
- 971 **Pseudo-Apollodorus.** *Epitome*. In: *Library and Epitome*; 1st Century AD?.p. Ch 1 Sec 9.
- 972 **Rosser AE, Keverne EB.** The Importance of Central Noradrenergic Neurones in the Formation of an  
973 Olfactory Memory in the Prevention of Pregnancy Block. *Neuroscience*. 1985 Aug; 15(4):1141–1147.  
974 doi: [10.1016/0306-4522\(85\)90258-1](https://doi.org/10.1016/0306-4522(85)90258-1).
- 975 **Seward J, Bzip2;** 2019.
- 976 **Small WS.** Experimental Study of the Mental Processes of the Rat. II. *The American Journal of*  
977 *Psychology*. 1901; 12(2):206–239. doi: [10.2307/1412534](https://doi.org/10.2307/1412534).
- 978 **Tolman EC.** The Determiners of Behavior at a Choice Point. *Psychological Review*. 1938; 45:1–41.  
979 doi: [10.1037/h0062733](https://doi.org/10.1037/h0062733).
- 980 **Tolman EC.** Cognitive Maps in Rats and Men. *Psychological Review*. 1948; 55(4):189–208. doi:  
981 [10.1037/h0061626](https://doi.org/10.1037/h0061626).
- 982 **Tolman E, Honzik C.** Degrees of Hunger, Reward and Non-Reward, and Maze Learning in Rats.  
983 *University of California Publications in Psychology*. 1930; 4:241–256.
- 984 **Uchida N, Mainen ZF.** Speed and Accuracy of Olfactory Discrimination in the Rat. *Nature Neuroscience*.  
985 2003 Nov; 6(11):1224–1229. doi: [10.1038/nn1142](https://doi.org/10.1038/nn1142).
- 986 **Uster HJ, Bättig K, Nägeli HH.** Effects of Maze Geometry and Experience on Exploratory Behavior in  
987 the Rat. *Animal Learning & Behavior*. 1976 Mar; 4(1):84–88. doi: [10.3758/BF03211992](https://doi.org/10.3758/BF03211992).
- 988 **Weber JN, Peterson BK, Hoekstra HE.** Discrete Genetic Modules Are Responsible for Complex Burrow  
989 Evolution in *Peromyscus* Mice. *Nature*. 2013 Jan; 493(7432):402–405. doi: [10.1038/nature11816](https://doi.org/10.1038/nature11816).
- 990 **Wehner R, Michel B, Antonsen P.** Visual Navigation in Insects: Coupling of Egocentric and Geocentric  
991 Information. *Journal of Experimental Biology*. 1996 Jan; 199(1):129–140.
- 992 **Woodrow H.** The Problem of General Quantitative Laws in Psychology. *Psychological Bulletin*. 1942;  
993 39(1):1–27. doi: [10.1037/h0058275](https://doi.org/10.1037/h0058275).



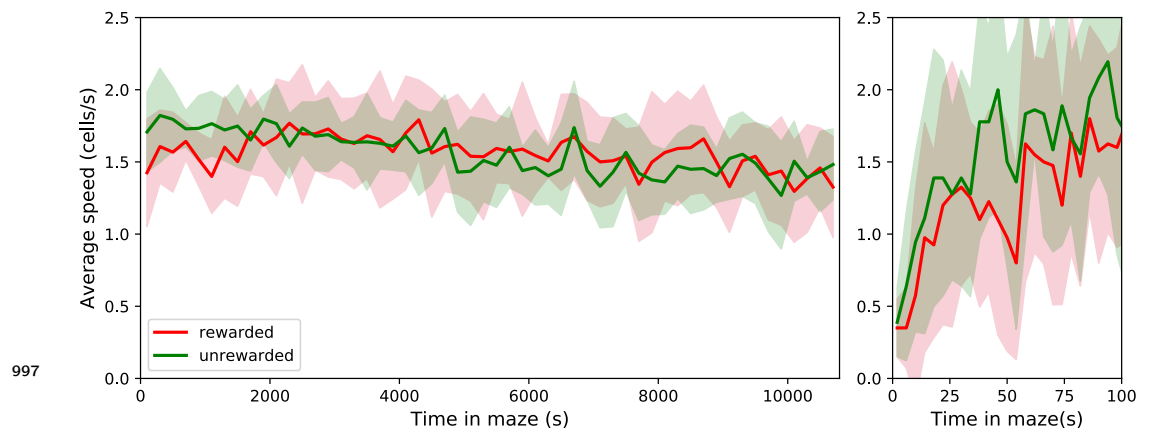
**Figure 1–figure supplement 1. Fraction of time spent in the maze.** Mice could move freely between the home cage and the maze. For each animal (vertical), the fraction of time in the maze (color scale) is plotted as a function of time since start of the experiment. Time bins are 500 s. Note that mouse D6 hardly entered the maze; it never progressed beyond the first junction. This animal was excluded from all subsequent analysis steps.



**Figure 1–figure supplement 2. Average fraction of time spent in the maze by group.** This shows the average fraction of time in the maze as Mean  $\pm$  SD over the population of 10 rewarded and 9 unrewarded animals. Right: expanded axis for early times. The tunnel to the maze opens at time 0. Rewarded and unrewarded animals used the maze in remarkably similar ways. Exploration of the maze began around 250 s after tunnel opening. Within the next 250 s the maze occupancy rose quickly to  $\sim$ 70%, then declined gradually over 7 h to  $\sim$ 30%.

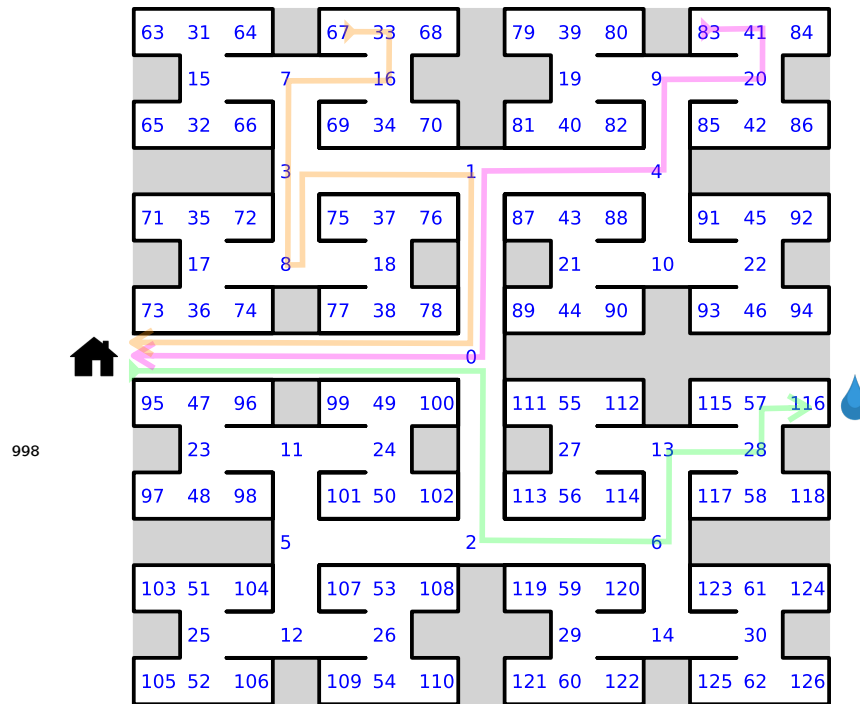


**Figure 1-figure supplement 3. Rates of transition between cage and maze.** (A) The instantaneous probability per unit time  $r_m(t)$  of entering the maze after having spent time  $t$  in the cage. Note this rate is highest immediately upon entering the cage, then declines by a large factor. (B) The instantaneous probability per unit time  $r_c(t)$  of exiting the maze after having spent time  $t$  in the maze.



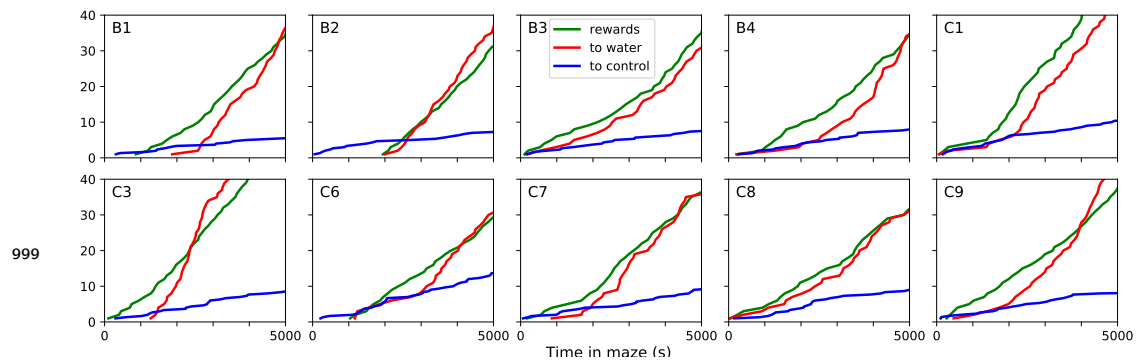
**Figure 2-figure supplement 1. The speed of locomotion in the maze is approximately constant.** Left: Speed plotted as Mean  $\pm$  SD over the population of rewarded and unrewarded animals. Right: expanded axis for early times. To assess the speed of locomotion we divided the maze into square cells as wide as the corridors and tracked how the nose of the animal moved through those cells. Then the speed was measured in number of cells traversed per unit time. Note that the speed is very similar across animals,  $\sim 1.56$  cells/s = 5.94 cm/s on average. It rises quickly over the first 50 s in the maze, then varies only little over the 7 h of the experiment.





998

**Figure 3—figure supplement 1. Definition of node trajectories.** A numbering scheme for all 127 nodes of the maze. Green: a direct path from the entrance to the water port (“water run”) with the node sequence  $(s_i) = (0, 2, 6, 13, 28, 57, 116)$ , involving 6 decisions. Magenta: a direct path from end node 83 to the exit (“home run”). Orange: a path from end node 67 to the exit that includes a reversal. Here the home run starts only from node 8, namely  $(8, 3, 1, 0)$ .

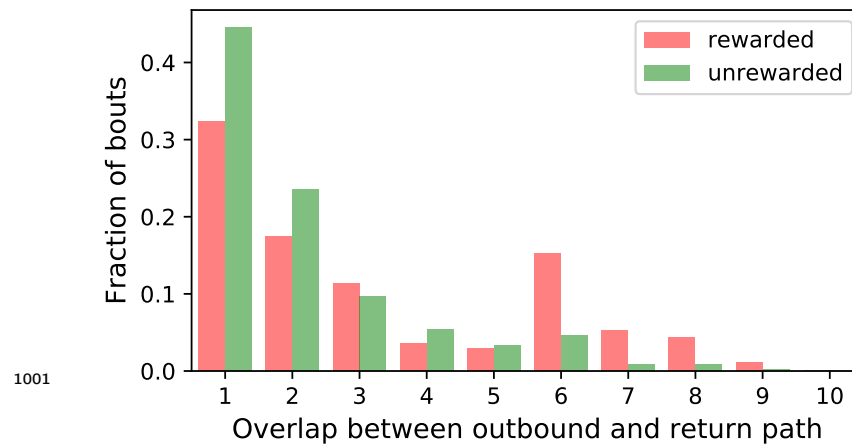


999

**Figure 4—figure supplement 1. Sudden changes in behavior for all rewarded animals.** For each of the 10 water-deprived animals this shows the cumulative rate of rewards, of long direct paths (>6 steps) to the water port, and of similar paths to 3 control nodes. Display as in **Figure 4**; panels B and C of that figure are included again here.

Animal	Likelihood ratio step/ramp	Time of step (s)	Ratio of rates after/before
<b>B1</b>	25.1	2580 ± 110	36.4
<b>B2</b>	36.2	2350 ± 220	30.3
B3	1.61	3090 ± 970	2.77
B4	0.67	2490 ± 590	7.36
<b>C1</b>	25.8	2070 ± 310	5.49
<b>C3</b>	32900	1280 ± 80	1640
C6	2.31	2960 ± 890	4.75
<b>C7</b>	41.3	1680 ± 280	16.9
C8	3.52	3070 ± 1000	2.50
C9	0.44	1980 ± 560	6.69

**Figure 4—figure supplement 2. Statistics of sudden changes in behavior.** Summary of ‘step’ and ‘ramp’ models fitted to the occurrence of long direct paths to the water port for all 10 rewarded animals. Boldface animals have a likelihood ratio > 25.



**Figure 5—figure supplement 1. Entry paths do not retrace exit paths.** For every bout we compared the start of the node sequence leading into the maze with the final portion leading back out to the exit. The number of nodes of the entry sequence that match the time-reverse of the exit sequence is called the “overlap”. This figure histograms the overlap for all bouts of all animals. Note the minimum overlap is 1, because all paths into and out of the maze have to pass through the central junction (node 0 in *Figure 3—figure supplement 1*). This is also the most frequent overlap. The peak at overlap 6 for rewarded animals results from the frequent direct paths to the water port and back, a sequence of 6 nodes in each direction.

Manuscript under review

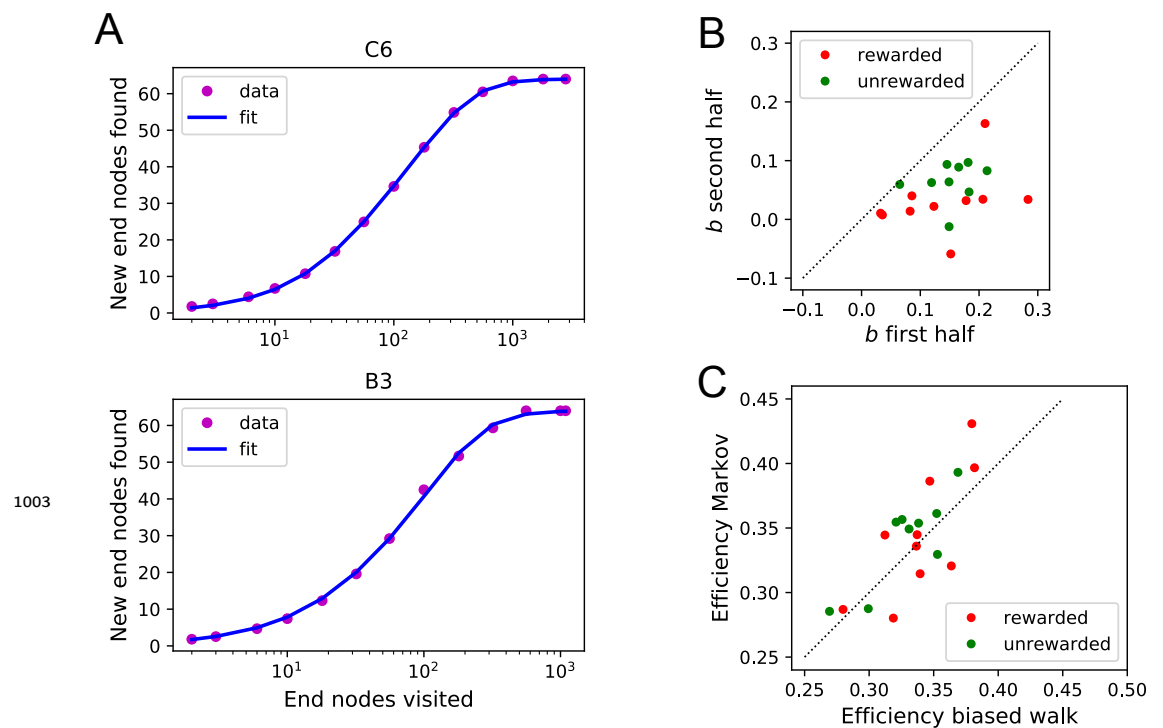
**A** Fraction of time in modes

Mode	rewarded	unrewarded
leave	0.053 ± 0.014	0.054 ± 0.013
drink	0.103 ± 0.026	
explore	0.844 ± 0.032	0.946 ± 0.013

**B** Transition probability between modes: rewarded animals

from / to:	leave	drink	explore
leave		0.51 ± 0.14	0.49 ± 0.14
drink	0.10 ± 0.05		0.90 ± 0.05
explore	0.40 ± 0.11	0.60 ± 0.11	

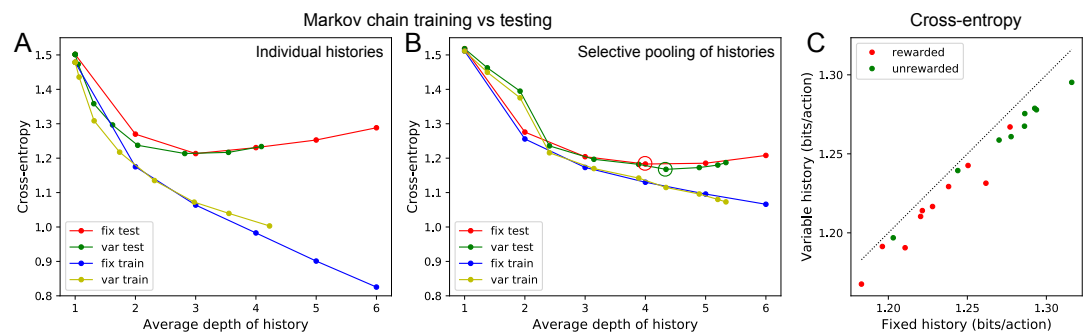
**Figure 6–figure supplement 1. Three modes of behavior.** (A) The fraction of time mice spent in each of the three modes while in the maze. Mean ± SD for 10 rewarded and 9 unrewarded animals. (B) Probability of transitioning from the mode on the left to the mode at the top. Transitions from ‘leave’ represent what the animal does at the start of the next bout into the maze.



**Figure 7–figure supplement 1. Functional fits to measure exploration efficiency** (A) Fitting Equation 13 to the data from the mouse’s exploration. Animals with best fit (top) and worst fit (bottom). The relative uncertainty in the two fit parameters  $a$  and  $b$  was only  $0.0038 \pm 0.0020$  (mean ± SD across animals). (B) The fit parameter  $b$  for all animals, comparing the first to the second half of the night. (C) The efficiency  $E$  (Equation 1) predicted from two models of the mouse’s trajectory: The 4-bias random walk (Figure 10D) and the optimal Markov chain (Figure 10C).

Bias	rewarded	unrewarded
$P_{SF}$	$0.77 \pm 0.03$	$0.78 \pm 0.02$
$P_{SA}$	$0.72 \pm 0.02$	$0.71 \pm 0.02$
$P_{BF}$	$0.82 \pm 0.03$	$0.81 \pm 0.03$
$P_{BS}$	$0.64 \pm 0.02$	$0.63 \pm 0.02$

**Figure 8–figure supplement 1. Statistics of the four turning biases.** Mean and standard deviation of the 4 biases of Figure 8A-B across animals in the rewarded and unrewarded groups.



**Figure 10–figure supplement 1. Fitting Markov models of behavior.** (A) Results of fitting the node sequence of a single animal (C3) with Markov models having a fixed depth ('fix') or variable depth ('var'). The cross-entropy of the model's prediction is plotted as a function of the average depth of history. In both cases we compare the results obtained on the training data ('train') vs those on separate testing data ('test'). Note that at larger depth the 'test' and 'train' estimates diverge, a sign of over-fitting the limited data available. (B) As in (A) but to combat the data limitation we pooled the counts obtained at all nodes that were equivalent under the symmetry of the maze (see Methods). Note considerably less divergence between 'train' and 'test' results, and a slightly lower cross-entropy during 'test' than in (A). (C) The minimal cross-entropy (circles in (B)) produced by variable vs fixed history models for each of the 19 animals. Note the variable history model always produces a better fit to the behavior.