

# <sup>1</sup> Mice in a labyrinth: Rapid learning, <sup>2</sup> sudden insight, and efficient exploration

<sup>3</sup> **Matthew Rosenberg<sup>1†</sup>, Tony Zhang<sup>1†</sup>, Pietro Perona<sup>2</sup>, Markus Meister<sup>1\*</sup>**

\*For correspondence:

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

<sup>†</sup>These authors contributed  
equally to this work

<sup>4</sup> <sup>1</sup>Division of Biology and Biological Engineering, California Institute of Technology;

<sup>5</sup> <sup>2</sup>Division of Engineering and Applied Science, California Institute of Technology

---

<sup>7</sup> **Abstract** Animals learn certain complex tasks remarkably fast, sometimes after a single  
<sup>8</sup> experience. What behavioral algorithms support this efficiency? Many contemporary studies  
<sup>9</sup> based on two-alternative-forced-choice (2AFC) tasks observe only slow or incomplete  
<sup>10</sup> learning. As an alternative, we study the unconstrained behavior of mice in a complex  
<sup>11</sup> labyrinth and measure the dynamics of learning and the behaviors that enable it. A mouse in  
<sup>12</sup> the labyrinth makes ~2000 navigation decisions per hour. The animal explores the maze,  
<sup>13</sup> quickly discovers the location of a reward, and executes correct 10-bit choices after only 10  
<sup>14</sup> reward experiences – a learning rate 1000-fold higher than in 2AFC experiments. Many mice  
<sup>15</sup> improve discontinuously from one minute to the next, suggesting moments of sudden insight  
<sup>16</sup> about the structure of the labyrinth. The underlying search algorithm does not require a global  
<sup>17</sup> memory of places visited and is largely explained by purely local turning rules.

<sup>18</sup>

---

## <sup>19</sup> Introduction

<sup>20</sup> How can animals or machines acquire the ability for complex behaviors from one or a few  
<sup>21</sup> experiences? Canonical examples include language learning in children, where new words are  
<sup>22</sup> learned after just a few instances of their use, or learning to balance a bicycle, where humans  
<sup>23</sup> progress from complete incompetence to near perfection after crashing once or a few times.  
<sup>24</sup> Clearly such rapid acquisition of new associations or of new motor skills can confer enormous  
<sup>25</sup> survival advantages.

<sup>26</sup> In laboratory studies, one prominent instance of one-shot learning is the Bruce effect  
<sup>27</sup> (*Bruce, 1959*). Here the female mouse forms an olfactory memory of her mating partner that  
<sup>28</sup> allows her to terminate the pregnancy if she encounters another male that threatens infanticide.  
<sup>29</sup> Another form of rapid learning accessible to laboratory experiments is fear conditioning, where  
<sup>30</sup> a formerly innocuous stimulus gets associated with a painful experience, leading to subsequent  
<sup>31</sup> avoidance of the stimulus (*Fanselow and Bolles, 1979; Bourchuladze et al., 1994*). These  
<sup>32</sup> learning systems appear designed for special purposes, they perform very specific associations,  
<sup>33</sup> and govern binary behavioral decisions. They are likely implemented by specialized brain  
<sup>34</sup> circuits, and indeed great progress has been made in localizing these operations to the accessory  
<sup>35</sup> olfactory bulb (*Brennan and Keverne, 1997*) and the cortical amygdala (*LeDoux, 2000*).

<sup>36</sup> In the attempt to identify more generalizable mechanisms of learning and decision making,  
<sup>37</sup> one route has been to train laboratory animals on abstract tasks with tightly specified sensory  
<sup>38</sup> inputs that are linked to motor outputs via arbitrary contingency rules. Canonical examples  
<sup>39</sup> are a monkey reporting motion in a visual stimulus by saccading its eyes (*Newsome and Pare,*  
<sup>40</sup> *1988*), and a mouse in a box classifying stimuli by moving its forelimbs or the tongue (*Burgess*  
<sup>41</sup> *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 (*Carandini and Churchland, 2013*).

45 What is needed therefore is a rodent behavior that involves complex decision making, with  
 46 many input variables and many possible choices. Ideally the animals would learn to perform  
 47 this task without excessive intervention by human shaping, so we may be confident that they  
 48 employ innate brain mechanisms rather than circuits created by the training. Obviously the  
 49 behavior should be easy to measure in the laboratory. Finally, it would be satisfying if this  
 50 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 the  
 57 environment in the form of a maze that defines straight paths and decision points. A maze of  
 58 tunnels is in fact a natural environment for a burrowing rodent. Early studies of rodent behavior  
 59 did place the animals into true labyrinths (*Small, 1901*), but their use gradually declined in  
 60 favor of linear tracks or boxes with a single choice point.

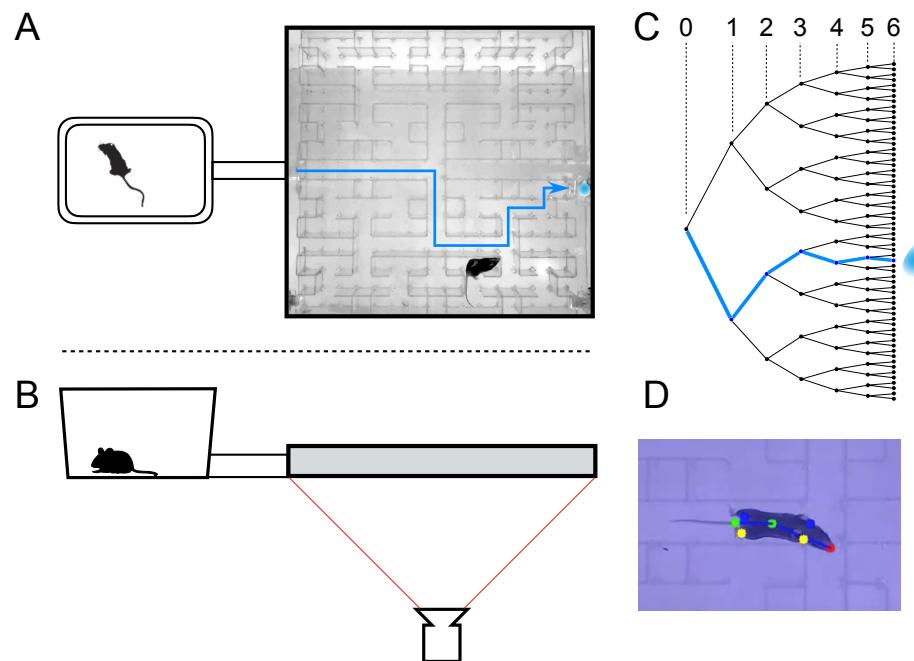
61 We report here on the behavior of laboratory mice in a complex labyrinth of tunnels. A  
 62 single mouse is placed in a home cage from which it has free access to the maze for one  
 63 night. No handling, shaping, or training by the investigators is involved. By continuous video-  
 64 recording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

## 73 Results

### 74 Adaptation to the maze

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

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

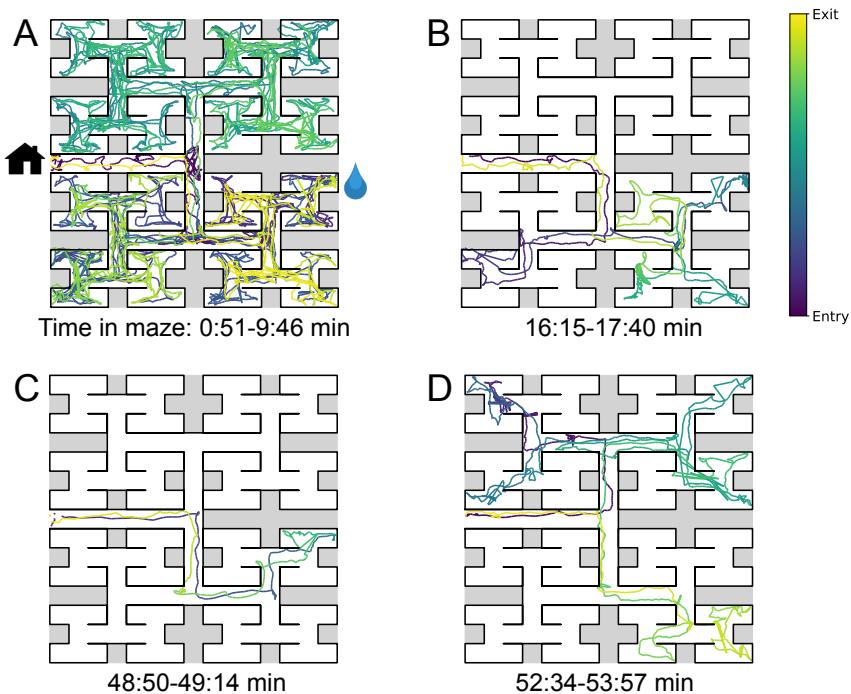


**Figure 1. The maze environment.** Top (A) and side (B) views of a home cage, connected via an entry tunnel to an enclosed labyrinth. The animal's actions in the maze are recorded via video from below using infrared illumination. (C) The maze is structured as a binary tree with 63 branch points (in 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). The trajectory of the animal's nose is shown; time is encoded by the color of the trace. The entrance from the home cage and the water port are indicated in panel A.

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

89 After an initial period of exploratory experiments we settled on a frozen protocol that was  
90 applied to 20 animals. Ten of these mice had been mildly water-deprived for up to 24 hours;  
91 they received food in the home cage and water only from the port hidden in the maze. Another  
92 ten mice had free access to food and water in the cage, and received no water from the port  
93 in the maze. Each animal's behavior in the maze was recorded continuously for 7 h during  
94 the first night of its experience with the maze, starting the moment the connection tunnel was  
95 opened (sample videos [here](#)). The investigator played no role during this period, and the animal  
96 was free to act as it wished including travel between the cage and the maze.

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

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

112 rate throughout the visit (*Figure 1–figure supplement 3B*).

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

125 Whereas *Figure 2* illustrates the trajectory of a mouse’s nose in full spatio-temporal detail,  
 126 a convenient reduced representation is the “node sequence”. This simply marks the events  
 127 when the animal enters each of the 127 nodes of the binary tree that describes the maze (see  
 128 Methods and *Figure 3–figure supplement 1*). Among these nodes, 63 are T-junctions where  
 129 the animal has 3 choices for the next node, and 64 are end nodes where the animal’s only choice  
 130 is to reverse course. We call the transition from one node to the next a “step”. The analysis in  
 131 the rest of the paper was carried out on the animal’s node sequence.

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

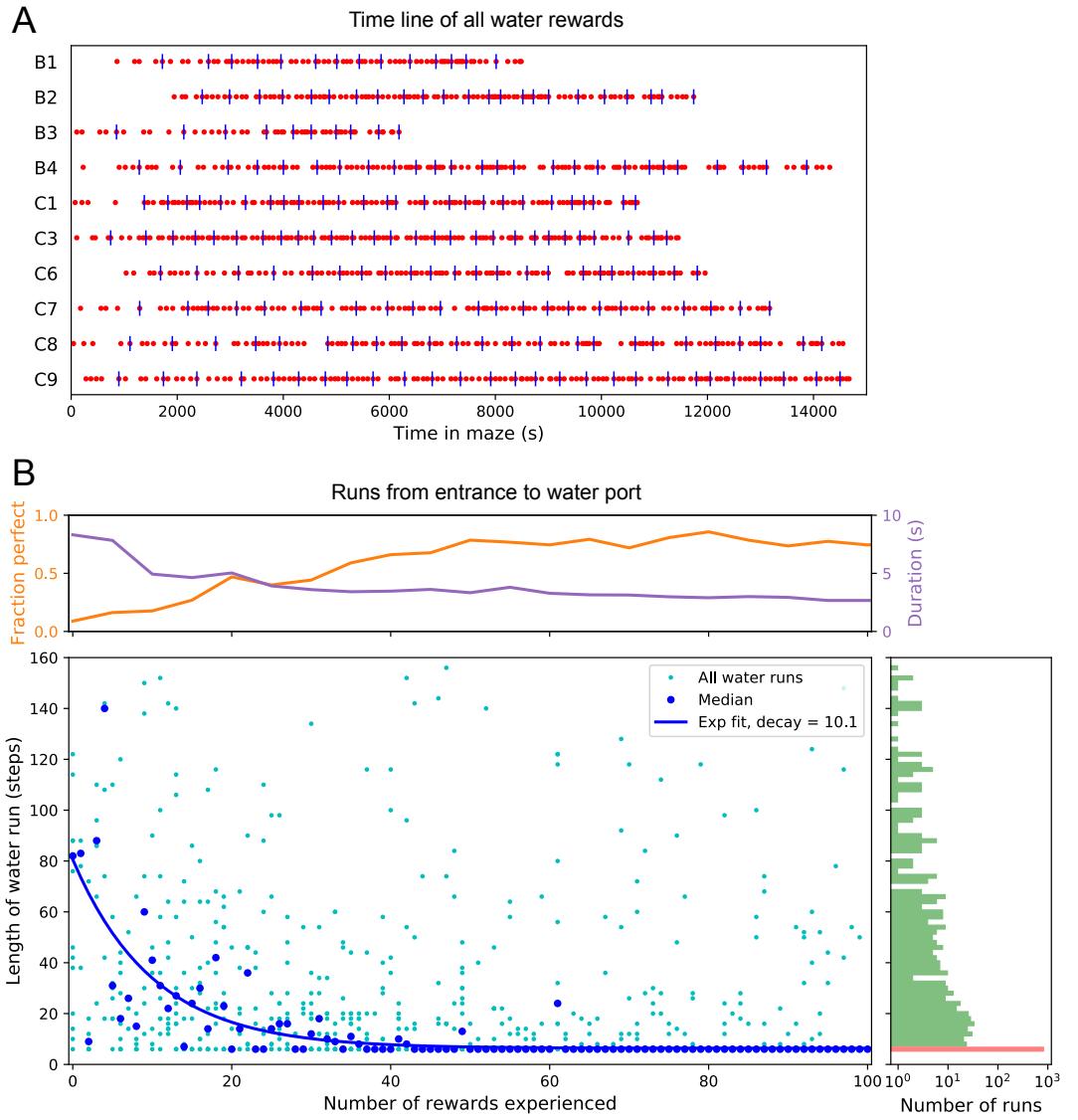
133 We now examine early changes in the animal’s behavior when it rapidly acquires and remembers  
 134 information needed for navigation. First we focus on navigation to the water port.

135 The ten water-deprived animals had no indication that water would be found in the maze.  
 136 Yet, all 10 discovered the water port in less than 2000 s and fewer than 17 bouts (*Figure 3A*).  
 137 The port dispensed only a drop of water followed by a 90-s timeout before rearming. During the  
 138 timeout the animals generally left the port location to explore other parts of the maze or return  
 139 home, even though they were not obliged to do so. For each of the water-deprived animals, the  
 140 frequency at which it consumed rewards in the maze increased rapidly as it learned how to find  
 141 the water port, then settled after a few reward experiences (*Figure 3A*).

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

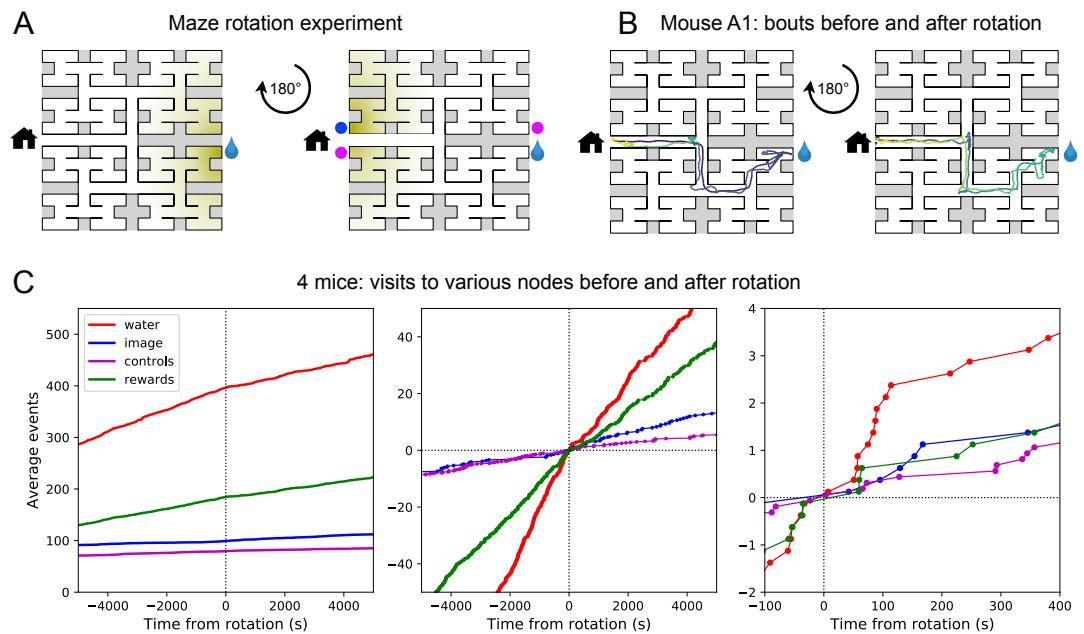
148 On the first few occasions these paths to water can involve hundreds of steps between nodes  
 149 and their length scatters over a wide range. However, after a few rewards, the animals began  
 150 taking the perfect path without detours (6 steps, *Figure 3–figure supplement 1*), and soon that 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

158 In summary, after ~10 reward experiences on average the mice learn to navigate efficiently  
 159 to the water port, which requires making 6 correct decisions, each among 3 options. Note that



**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. Navigation is robust to rotation of the maze.** (A) Logic of the experiment: The animal may have deposited an odorant in the maze (shading) that is centered on the water port. After 180 degree rotation of the maze, that gradient would lead to the image of the water port (blue dot). We also measure how often the mouse goes to two control nodes (magenta dots) that are related by symmetry. (B) Trajectory of mouse ‘A1’ in the bouts immediately before and after maze rotation. Time coded by color from dark to light as in *Figure 2*. (C) Left: Cumulative number of rewards as well as visits to the water port, the image of the water port, and the control nodes. All events are plotted vs time before and after the maze rotation. Average over 4 animals. Middle and right: Same data with the counts centered on zero and zoomed in for better resolution.

**Figure 4-figure supplement 1.** Navigation before and after maze rotation for each animal.

**Figure 4-figure supplement 2.** Speed before and after maze rotation.

even at late times, long after they have perfected the “water run”, the animals continue to take some extremely long paths: a subject for a later section (*Figure 7*).

### 162 The role of cues attached to the maze

163 These observations of rapid learning raise the question "How do the animals navigate?" In  
 164 particular, does the mouse build an internal representation that guides its action at every  
 165 junction? Or does it place marks in the external environment that signal the route to the water  
 166 port? In an extreme version of externalized cognition, the mouse leaves behind a trail of urine  
 167 marks or other secretions as it walks away from the water port, and on a subsequent bout simply  
 168 sniffs its way up the odor gradient (*Figure 4A*). This would require no internal representation.

169 The following experiment offers some partial insights. Owing to the design of the labyrinth  
 170 one can rotate the entire apparatus by 180 degrees, open one wall and close another, and obtain  
 171 a maze with the same structure (*Figure 4A*). Alternatively one can also rotate only the floor.  
 172 After such a modification, all the physical cues attached to the rotated parts now point in the  
 173 wrong direction, namely to the end node 180 degrees opposite the water port (the "image  
 174 location"). If the animal navigated to the goal following cues previously deposited in the maze  
 175 it should end up at that image location.

176 We performed a maze rotation on four animals after several hours of exposure, when  
 177 they had acquired the perfect route to water. Immediately after rotation, 3 of the 4 animals

178 went to the correct water port on their first entry into the maze, and before ever visiting the  
 179 image location (e.g. *Figure 4B*). The fourth mouse visited the image location once and then  
 180 the correct water port (*Figure 4–figure supplement 1*). The mice continued to collect water  
 181 rewards efficiently even immediately after the rotation.

182 Nonetheless, the maze rotation did introduce subtle changes in behavior that lasted for an  
 183 hour or more (*Figure 4C*). Visits to the image location were at chance levels prior to rotation,  
 184 then increased by a factor of 1.8. Visits to the water port declined in frequency, although they  
 185 still exceeded visits to the image location by a factor of 5. The reward rate declined by a factor  
 186 of 0.7. These effects could be verified for each animal (*Figure 4–figure supplement 1*). The  
 187 speed of the mice was not disturbed (*Figure 4–figure supplement 2*).

188 In summary, for navigation to the water port the experienced animals do not strictly depend  
 189 on physical cues that are attached to the maze. This includes any material they might have  
 190 deposited, but also pre-existing construction details by which they may have learned to identify  
 191 locations in the maze. The mice clearly notice a change in these cues, but continue to navigate  
 192 effectively to the goal. This conclusion applies to the time point of the rotation, a few hours  
 193 into the experiment. Conceivably the animal's navigation policy and its use of sensory cues  
 194 changes in the course of learning. This and many other questions regarding the mechanisms of  
 195 cognition will be taken up in a separate study.

### 196 Discontinuous learning

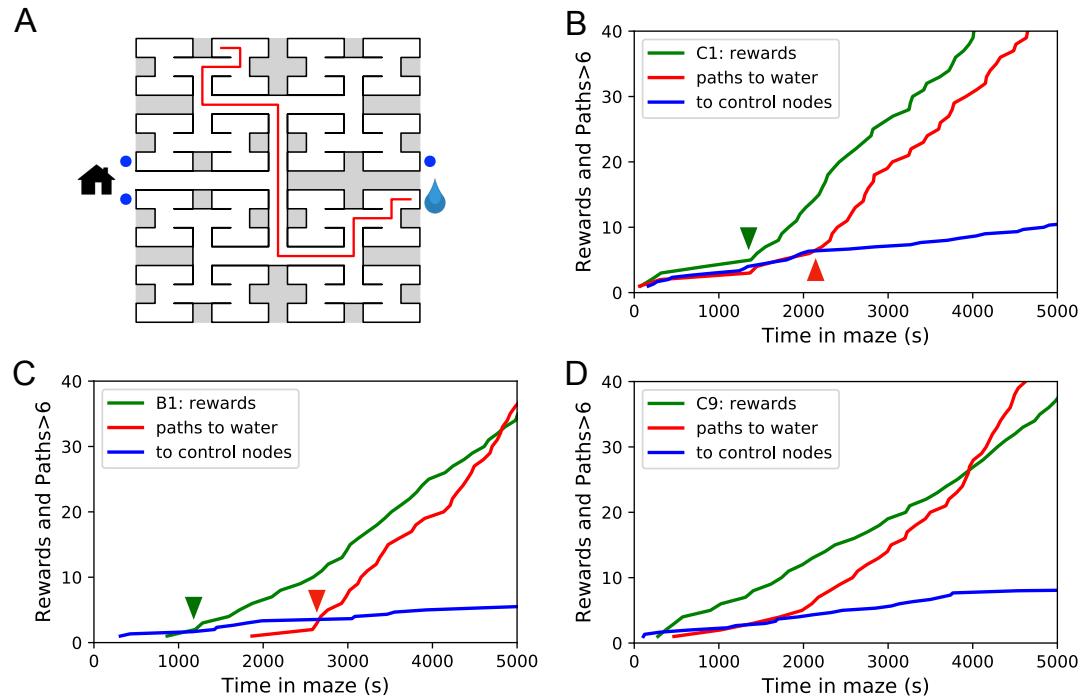
197 While an average across animals shows evidence of rapid learning (*Figure 3*) one wonders  
 198 whether the knowledge is acquired gradually or discontinuously, through moments of "sudden  
 199 insight". To explore this we scrutinized more closely the time line of individual water-deprived  
 200 animals in their experience with the maze. The discovery of the water port and the subsequent  
 201 collection of water drops at a regular rate is one clear change in behavior that relies on new  
 202 knowledge. Indeed, the rate of water rewards can increase rather suddenly (*Figure 3A*),  
 203 suggesting an instantaneous step in knowledge.

204 Over time, the animals learned the path to water not only from the entrance of the maze but  
 205 from many locations scattered throughout the maze. The largest distance between the water  
 206 port and an end node in the opposite half of the maze involves 12 steps through 11 intersections  
 207 (*Figure 5A*). Thus we included as another behavioral variable the occurrence of long direct  
 208 paths to the water port which reflects how directedly the animals navigate within the maze.

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

216 The animal behaves as though it gains a new insight at the time of the second step that  
 217 allows it to travel to the water port directly from elsewhere in the maze. Note that the two  
 218 behavioral variables are independent: The long paths don't change when the reward rate steps  
 219 up, and the reward rate doesn't change when the rate of long paths steps up. Another animal  
 220 (*Figure 5C*) similarly showed an early step in the reward rate (at 860 s) and a dramatic step in  
 221 the rate of long paths (at 2580 s). In this case the emergence of long paths coincided with a  
 222 modest increase (factor of 2) in the reward rate.

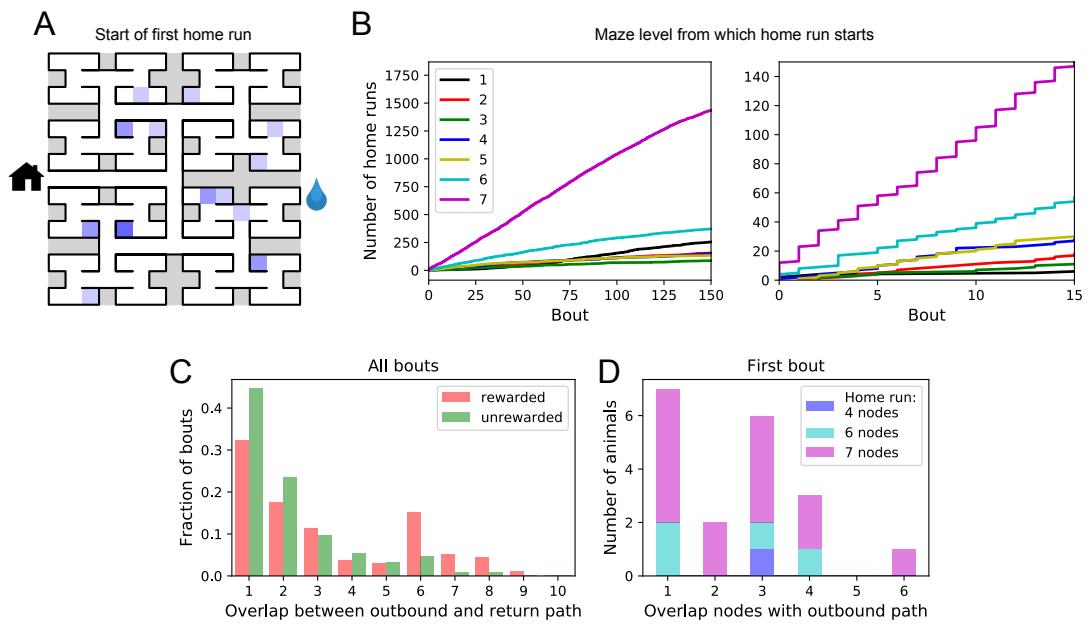
223 Similar discontinuities in behavior were seen in at least 5 of the 10 water-deprived animals  
 224 (*Figure 5B*, *Figure 5–figure supplement 1*, *Figure 5–figure supplement 2*), and their timing  
 225 could be identified to a precision of ~200 s. More gradual performance change was observed



**Figure 5. Sudden changes in behavior.** (A) An example of a long uninterrupted path through 11 junctions to the water port (drop icon). Blue 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. (D) Same as B for animal C9, an example of more continuous learning.

**Figure 5–figure supplement 1.** Long direct paths for all animals.

**Figure 5–figure supplement 2.** Statistics of sudden changes in behavior.



**Figure 6. 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 1 = first T-junction, level 7 = end nodes. Right: Zoom of (Left) into early bouts. (C) Overlap between the outbound and the home path. Histogram of the overlap for all bouts of all animals. (D) Same analysis for just the first bout of each animal. The length of the home run is color-coded as in panel B.

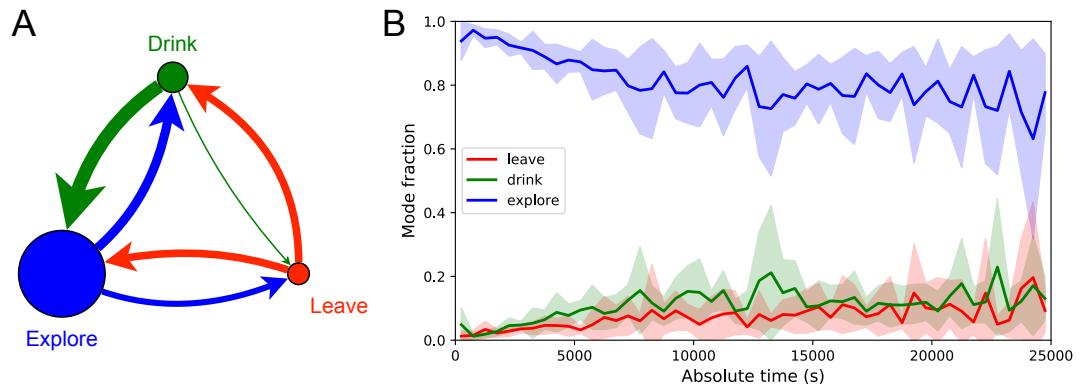
for the remaining animals (*Figure 5 D*). We varied the criterion of performance by asking for even longer error-free paths, and the results were largely unchanged and no additional discontinuity appeared. These observations suggest that mice can acquire a complex decision-making skill rather suddenly. A mouse may have multiple moments of sudden insight that affect different aspects of its behavior. The exact time of the insight cannot be predicted but is easily identified post-hoc. Future neurophysiological studies of the phenomenon will face the interesting challenge of capturing these singular events.

### One-shot learning of the home path

For an animal entering an unfamiliar environment, the most important path to keep in memory may be the escape route. In the present case that is the route to the maze entrance, from which the tunnel leads home to the cage. We expected that the mice would begin by penetrating into the maze gradually and return home repeatedly so as to confirm the escape route, a pattern previously observed for rodents in an open arena (*Tchernichovski et al., 1998; Fonio et al., 2009*). This might help build a memory of the home path gradually level-by-level into the binary tree. Nothing could be further from the truth.

At the end of any given bout into the maze, there is a “home run”, namely the direct path without reversals that takes the animal to the exit (see *Figure 3-figure supplement 1*). *Figure 6 A* shows the nodes where each animal started its first home run, following the first penetration into the maze. With few exceptions, that first home run began from an end node, as deep into the maze as possible. Recall that this involves making the correct choice at six successive 3-way intersections, an outcome that is unlikely to happen by chance.

The above hypothesis regarding gradual practice of home runs would predict that short



**Figure 7. 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 7-figure supplement 1.** Three modes of behavior.

248 home runs should appear before long ones in the course of the experiment. The opposite is the  
 249 case (*Figure 6 B*). In fact, the end nodes (level 7 of the maze) are by far the favorite place from  
 250 which to return to the exit, and those maximal-length home runs systematically appear before  
 251 shorter ones. This conclusion was confirmed for each individual animal, whether rewarded or  
 252 unrewarded.

253 Clearly the animals do not practice the home path or build it up gradually. Instead they  
 254 seem to possess an Ariadne’s thread (*Pseudo-Apollodorus, I-II Century AD*) starting with  
 255 their first excursion into the maze, long before they might have acquired any general knowledge  
 256 of the maze layout. On the other hand the mouse does not follow the strategy of Theseus,  
 257 namely to precisely retrace the path that led it into the labyrinth. In that case the animal’s home  
 258 path should be the reverse of the path into the maze that started the bout. Instead the entry  
 259 path and the home path tend to have little overlap (*Figure 6C*). Note the minimum overlap is 1,  
 260 because all paths into and out of the maze have to pass through the central junction (node 0 in  
 261 *Figure 3-figure supplement 1*). This is also the most frequent overlap. The peak at overlaps  
 262 6-8 for rewarded animals results from the frequent paths to the water port and back, a sequence  
 263 of at least 7 nodes in each direction. The separation of outbound and return path is seen even  
 264 on the very first home run (*Figure 6D*). Many home runs from the deepest level (7 nodes) have  
 265 only the central junction in common with the outbound path (overlap = 1).

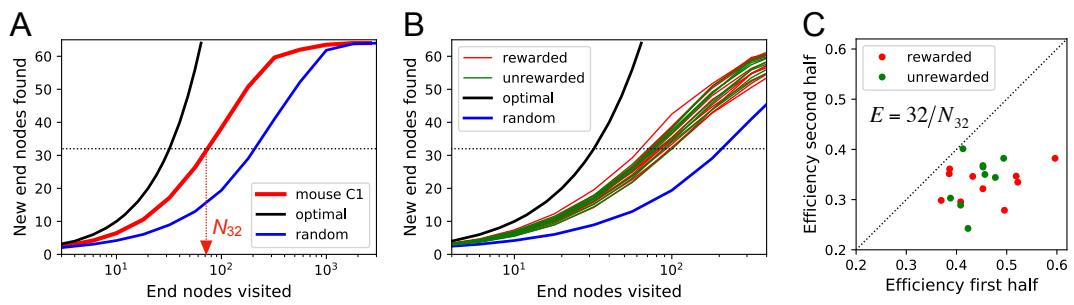
266 In summary it appears that the animal acquires a homing strategy over the course of a  
 267 single bout, and in a manner that allows a direct return home even from locations not previously  
 268 encountered.

### 269 Structure of behavior in the maze

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

#### 272 Behavioral states

273 Once the animal has learned to perform long uninterrupted paths to the water port, one can  
 274 categorize its behavior within the maze by three states: (1) walking to the water port; (2)



**Figure 8. 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 8-figure supplement 1.** Efficiency of exploration

walking to the exit; and (3) exploring the maze. Operationally we define exploration as all periods in which the animal is in the maze but not on a direct path to water or to the exit. For the ten sated animals this includes all times in the maze except for the walks to the exit.

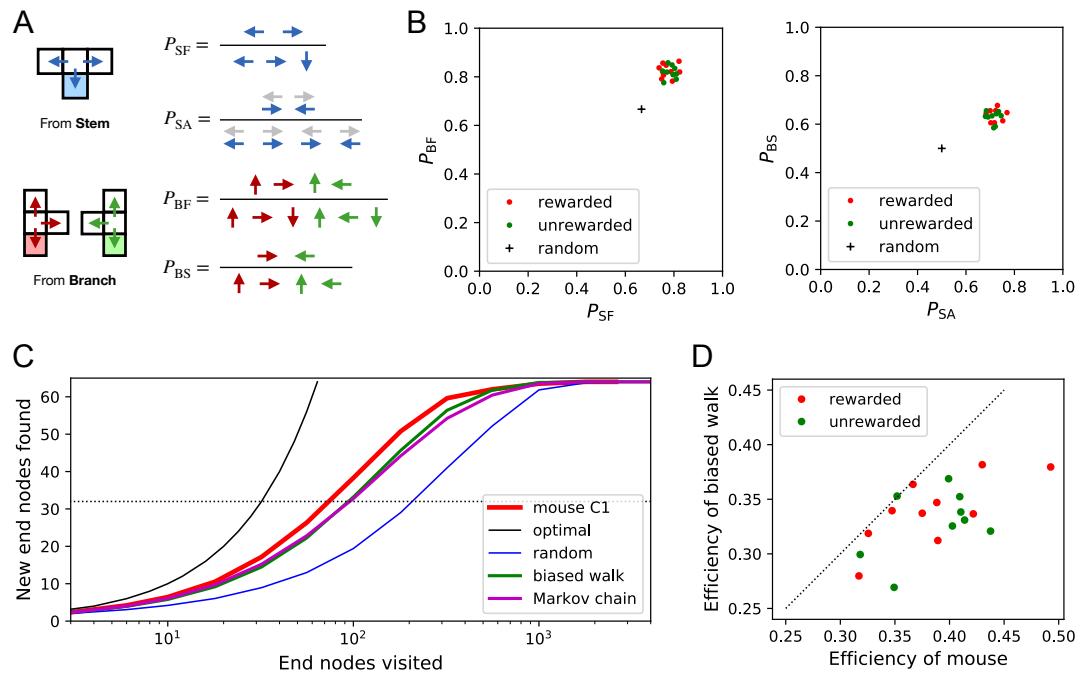
*Figure 7* illustrates the occupancies and transition probabilities between these states. The animals spent most of their time by far in the exploration state: 84% for rewarded and 95% for unrewarded mice. Across animals there was very little variation in the balance of the 3 modes (*Figure 7-figure supplement 1*). The rewarded mice began about half their bouts into the maze with a trip to the water port and the other half by exploring (*Figure 7A*). After a drink, the animals routinely continued exploring, about 90% of the time.

For water-deprived animals the dominance of exploration persisted even at a late stage of the night when they routinely executed perfect exploitation bouts to and from the water port: Over the duration of the night the ‘explore’ fraction dropped slightly from 0.92 to 0.75, with the balance accrued to the ‘drink’ and ‘leave’ modes as the animals executed many direct runs to the water port and back. The unrewarded group of animals also explored the maze throughout the night even though it offered no overt rewards (*Figure 7-figure supplement 1*). One suspects that the animals derive some intrinsic reward from the act of patrolling the environment itself.

#### 291 Efficiency of exploration

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

295 One can presume that a goal of the exploratory mode is to rapidly survey all parts of the  
296 environment for the appearance of new resources or threats. We will measure the efficiency of  
297 exploration by how rapidly the animal visits all end nodes of the binary maze, starting at any  
298 time during the experiment. The optimal agent with perfect memory and complete knowledge  
299 of the maze – including the absence of any loops – could visit the end nodes systematically  
300 one after another without repeats, thus encountering all of them after just 64 visits. A less  
301 perfect agent, on the other hand, will visit the same node repeatedly before having encountered  
302 all of them. *Figure 8A* plots for one exploring mouse the number of distinct end nodes it  
303 encountered as a function of the number of end nodes visited. The number of new nodes rises



**Figure 9. 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 8A*, 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 11C*). (D) Efficiency of exploration (*Equation 1*) in 19 mice compared to the efficiency of the corresponding biased random walk.

**Figure 9—figure supplement 1.** Bias statistics.

monotonically; 32 of the end nodes have been discovered after the mouse checked 76 times; then the curve gradually asymptotes to 64. We will characterize the efficiency of the search by the number of visits  $N_{32}$  required to survey half the end nodes, and define

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

This mouse explores with efficiency  $E = 32/76 = 0.42$ . For comparison, *Figure 8A* plots the performance of the optimal agent ( $E = 1.0$ ) and that of a random walker that makes random decisions at every 3-way junction ( $E = 0.23$ ). Note the mouse is about half as efficient as the optimal agent, but twice as efficient as a random walker.

The different mice were remarkably alike in this component of their exploratory behavior (*Figure 8B*): across animals the efficiency varied by only 11% of the mean ( $0.387 \pm 0.044$  SD). Furthermore there was no detectable difference in efficiency between the rewarded animals and the sated unrewarded animals. Over the course of the night the efficiency declined significantly for almost every animal – whether rewarded or not – by an average of 23% (*Figure 8C*).

316 Rules of exploration

317 What allows the mice to search much more efficiently than a random walking agent? We  
 318 inspected more closely the decisions that the animals make at each 3-way junction. It emerged  
 319 that these decisions are governed by strong biases (*Figure 9*). The probability of choosing  
 320 each arm of a T-junction depends crucially on how the animal entered the junction. The animal  
 321 can enter a T-junction from 3 places and exit it in 3 directions (*Figure 9A*). By tallying the  
 322 frequency of all these occurrences across all T-junctions in the maze one finds clear deviations  
 323 from an unbiased random walk (*Figure 9B, Figure 9-figure supplement 1*).

324 First, the animals have a strong preference for proceeding through a junction rather than  
 325 returning to the preceding node ( $P_{SF}$  and  $P_{BF}$  in *Figure 9B*). Second there is a bias in favor  
 326 of alternating turns left and right rather than repeating the same direction turn ( $P_{SA}$ ). Finally,  
 327 the mice have a mild preference for taking a branch off the straight corridor rather than  
 328 proceeding straight ( $P_{BS}$ ). A comparison across animals again revealed a remarkable degree  
 329 of consistency even in these local rules of behavior: The turning biases varied by only 3%  
 330 across the population and even between the rewarded and unrewarded groups (*Figure 9B, Figure 9-figure supplement 1*).

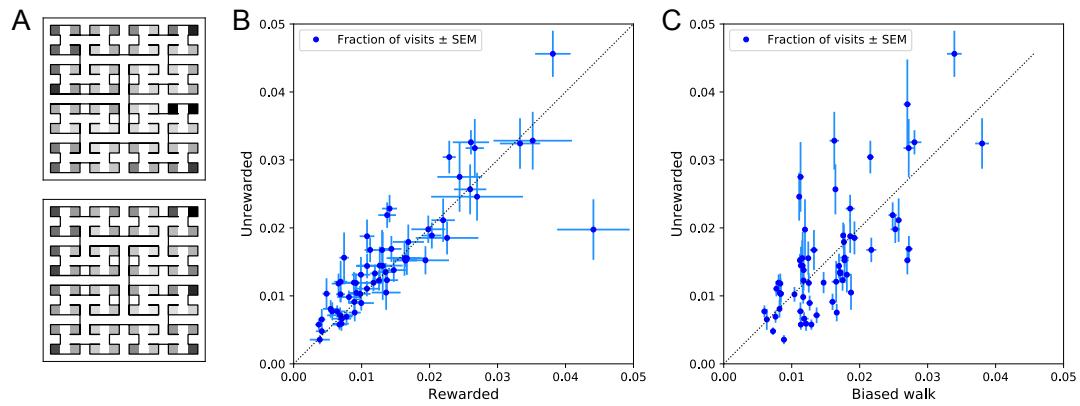
332 Qualitatively, one can see that these turning biases will improve the animal's search strategy.  
 333 The forward biases  $P_{SF}$  and  $P_{BF}$  keep the animal from re-entering territory it has covered already.  
 334 The bias  $P_{BS}$  favors taking a branch that leads out of the maze. This allows the animal to rapidly  
 335 cross multiple levels during an outward path and then enter a different territory. By comparison,  
 336 the unbiased random walk tends to get stuck in the tips of the tree and revisits the same end  
 337 nodes many times before escaping. To test this intuition we simulated a biased random agent  
 338 whose turning probabilities at a T-junction followed the same biases as measured from the  
 339 animal (*Figure 9C*). These biased agents did in fact search with much higher efficiency than  
 340 the unbiased random walk. They did not fully explain the behavior of the mice (*Figure 9D*),  
 341 accounting for ~87% of the animal's efficiency (compared to 60% for the random walk). A more  
 342 sophisticated model of the animal's behavior - involving many more parameters (*Figure 11C*) -  
 343 failed to get any closer to the observed efficiency (*Figure 9C, Figure 8-figure supplement 1C*).  
 344 Clearly some components of efficient search in these mice remain to be understood.

345 Systematic node preferences

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

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

358 Interestingly, the biased random walk using four bias numbers (*Figure 9, Figure 11D*)  
 359 replicates a good amount of the pattern of preferences. For unrewarded animals, where the  
 360 maze symmetry is not disturbed by the water port, the biased random walk predicts 51%  
 361 of the observed variance across nodes (*Figure 10C*), and an outer/inner node preference of  
 362 1.97, almost matching the observed ratio of 2.20. The more complex Markov-chain model of  
 363 behavior (*Figure 11C*) performed slightly better, explaining 66% of the variance in port visits



**Figure 10. 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.

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

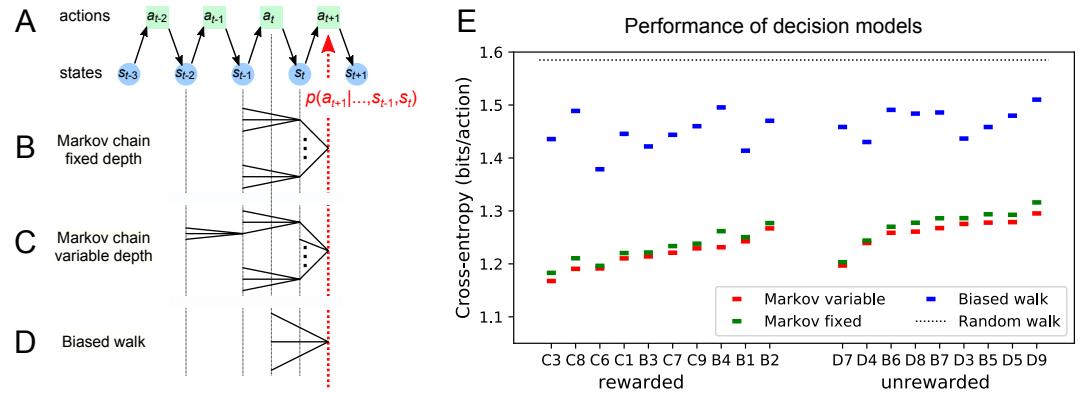
### 365 Models of maze behavior

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

371 At a formal level, the mouse's trajectory through the maze is a string of numbers standing  
 372 for the nodes the animal visited (*Figure 11A* and *Figure 3-figure supplement 1*). We want to  
 373 predict the next action of the mouse, namely the step that takes it to the next node. The quality  
 374 of the model will be assessed by the cross-entropy between the model's predictions and the  
 375 mouse's observed actions, measured in bits per action. This is the uncertainty that remains  
 376 about the mouse's next action given the prediction from the model. The ultimate lower limit is  
 377 the true source entropy of the mouse, namely that component of its decisions that cannot be  
 378 explained by the history of its actions.

379 One family of models we considered are fixed-depth Markov chains (*Figure 11B*). Here  
 380 the probability of the next action  $a_{t+1}$  is specified as a function of the history stretching over  
 381 the  $k$  preceding nodes  $(s_{t-k+1}, \dots, s_t)$ . In fitting the model to the mouse's actual node sequence  
 382 one tallies how often each history leads to each action, and uses those counts to estimate  
 383 the conditional probabilities  $p(a_{t+1}|s_{t-k+1}, \dots, s_t)$ . Given a new node sequence, the model  
 384 will then use the history strings  $(s_{t-k+1}, \dots, s_t)$  to predict the outcome of the next action. In  
 385 practice we trained the model on 80% of the animal's trajectory and tested it by evaluating the  
 386 cross-entropy on the remaining 20%.

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



**Figure 11. 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 9*, 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 11–figure supplement 1.** Markov model fits.

392 can no longer estimate any probabilities, and cross-validation on a different segment of data  
 393 fails catastrophically. In practice we found that this limitation sets in already beyond  $k = 2$   
 394 (*Figure 11–figure supplement 1A*). To address this issue of data-limitation we developed a  
 395 variable-depth Markov chain (*Figure 11C*). This model retains longer histories, but only if  
 396 they occur frequently enough to allow a reliable probability estimate (see Methods, *Figure 11–*  
 397 *figure supplement 1B–C*). In addition, we explored different schemes of pooling the counts  
 398 across certain T-junctions that are related by the symmetry of the maze (see Methods).

399 With these methods we focused on the portions of trajectory when the mouse was in ‘explore’  
 400 mode, because the segments in ‘drink’ and ‘leave’ mode are fully predictable. Furthermore,  
 401 we evaluated the models only at nodes corresponding to T-junctions, because the decision  
 402 from an end node is again fully predictable. *Figure 11E* compares the performance of various  
 403 models of mouse behavior. The variable-depth Markov chains routinely produced the best fits,  
 404 although the improvement over fixed-depth models was modest. Across all 19 animals in this  
 405 study the remaining uncertainty about the animal’s action at a T-junction is  $1.237 \pm 0.035$  (SD)  
 406 bits/action, compared to the prior uncertainty of  $\log_2 3 = 1.585$  bits. The rewarded animals  
 407 have slightly lower entropy than the unrewarded ones (1.216 vs 1.261 bits/action). The Markov  
 408 chain models that produced the best fits to the behavior used history strings with an average  
 409 length of ~4.

410 We also evaluated the predictions obtained from the simple biased random walk model  
 411 (*Figure 11D*). Recall that this attempts to capture the history-dependence with just 4 bias  
 412 parameters (*Figure 9A*). As expected this produced considerably higher cross-entropies than  
 413 the more sophisticated Markov chains (by about 18%, *Figure 11E*). Finally we used several  
 414 professional file compression routines to try and compress the mouse’s node sequence. In  
 415 principle, this sets an upper bound on the true source entropy of the mouse, even if the  
 416 compression algorithm has no understanding of animal behavior. The best such algorithm (bzip2  
 417 compression (*Seward, 2019*)) far under-performed all the other models of mouse behavior,  
 418 giving 43% higher cross-entropy on average, and thus offered no additional useful bounds.

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

## 427 Discussion

### 428 Summary of contributions

429 We present a new approach to the study of learning and decision-making in mice. We give the  
 430 animal access to a complex labyrinth and leave it undisturbed for a night while monitoring its  
 431 movements. The result is a rich data set that reveals new aspects of learning and the structure of  
 432 exploratory behavior. With these methods we find that mice learn a complex task that requires  
 433 6 correct 3-way decisions after only ~10 experiences of success (*Figure 2*, *Figure 3*). Along  
 434 the way the animal gains task knowledge in discontinuous steps that can be localized to within  
 435 a few minutes of resolution (*Figure 5*). Underlying the learning process is an exploratory  
 436 behavior that occupies 90% of the animal’s time in the maze and persists long after the task has  
 437 been mastered, even in complete absence of an extrinsic reward (*Figure 7*). The decisions the  
 438 animal makes at choice points in the labyrinth are constrained in part by the history of its actions

439 (*Figure 9, Figure 11*), in a way that favors efficient searching of the maze (*Figure 8*). This  
 440 microstructure of behavior is surprisingly consistent across mice, with variation in parameters of  
 441 only a few percent (*Figure 9*). Our most expressive models to predict the animal's choices still  
 442 leave a remaining uncertainty of ~1.24 bits per decision (*Figure 11*), a quantitative benchmark  
 443 by which competing models can be tested. Finally, some of the observations constrain what  
 444 algorithms the animals might use for learning and navigation (*Figure 4*).

445 **Historical context**

446 Mazes have been a staple of animal psychology for well over 100 years. The early versions  
 447 were true labyrinths. For example, *Small (1901)* built a model of the maze in Hampton Court  
 448 gardens scaled to rat size. Subsequent researchers felt less constrained by Victorian landscapes  
 449 and began to simplify the maze concept. Most commonly the maze offered one standard path  
 450 from a starting location to a food reward box. A few blind alleys would branch from the standard  
 451 path, and researchers would tally how many errors the animal committed by briefly turning  
 452 into a blind (*Tolman and Honzik, 1930*). Later on, the design was further reduced to a single  
 453 T-junction. After all, the elementary act of maze navigation is whether to turn left or right at a  
 454 junction (*Tolman, 1938*), so why not study that process in isolation? And reducing the concept  
 455 even further, one can ask the animal to refrain from walking altogether, and instead poke its  
 456 nose into a hole on the left or the right side of a box (*Uchida and Mainen, 2003*). This led to  
 457 the popular behavior boxes now found in rodent neuroscience laboratories everywhere. Each  
 458 of these reductions of the “maze” concept enabled a new type of experiment to study learning  
 459 and decision-making, for example limiting the number of choice points allows one to better  
 460 sample neural activity at each one. However, the essence of a “confusing network of paths”  
 461 has been lost along the way, and with it the behavioral richness of the animals navigating those  
 462 decisions.

463 Owing in part to the dissemination of user-friendly tools for animal tracking, one sees  
 464 a renaissance of experiments that embrace complex environments, including mazes with  
 465 many choice points (*Alonso et al., 2020; Wood et al., 2018; Sato et al., 2018; Nagy et al.,  
 466 2020; Rondi-Reig et al., 2006; Yoder et al., 2011; McNamara et al., 2014*), 3-dimensional  
 467 environments (*Grobéty and Schenk, 1992*), and infinite mazes (*Shokaku et al., 2020*). The  
 468 labyrinth in the present study is considerably more complex than Hampton Court or most of  
 469 the mazes employed by Tolman and others (*Tolman and Honzik, 1930; Buel, 1934; Munn,  
 470 1950a*). In those mazes the blind alleys are all short and unbranched; when an animal strays  
 471 from the target path it receives feedback quickly and can correct. By contrast our binary tree  
 472 maze has 64 equally deep branches, only one of which contains the reward port. If the animal  
 473 makes a mistake at any level of the tree it can find out only after traveling all the way to the last  
 474 node.

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

484 Our experiments had no imposed structure whatsoever; in fact it may be inappropriate to  
 485 call them experiments. The investigator opened the entry to the maze in the evening and did  
 486 not return until the morning. A potential advantage of leaving the animals to themselves is

487 that they are more likely to engage in mouse-like behavior, rather than constantly responding  
 488 to the stress of human interference or the alienation from being a cog in a behavior machine.  
 489 The result was a rich data set, with the typical animal delivering ~15,000 decisions in a single  
 490 night, even if one only counts the nodes of the binary tree as decision points. Since the mice  
 491 made all the choices, the scientific effort lay primarily in adapting methods of data analysis to  
 492 the nature of mouse trajectories. Somewhat surprisingly, the absence of experimental structure  
 493 was no obstacle to making precise and reproducible measurements of the animal's behavior.

#### 494 How fast do animals learn?

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

504 Many well-studied behaviors have a complexity of 1 bit or less, and often animals can  
 505 learn these associations after a single experience. For example, in the Bruce effect (*Bruce,*  
 506 *1959*) the female maps two different contexts (smell of mate vs non-mate) onto two kinds of  
 507 pregnancy outcomes (carry to term vs abort). The mutual information in that contingency table  
 508 is at most 1 bit, and may be considerably lower, for example if non-mate males are very rare or  
 509 very frequent. Mice form the correct association after a single instance of mating, although  
 510 proper memory formation requires several hours of exposure to the mate odor (*Rosser and*  
*Keverne, 1985*).

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

518 Animals are particularly adept at learning a new association between an odor and food. For  
 519 example bees will extend their proboscis in response to a new odor after just one pairing trial  
 520 where the odor appeared together with sugar (*Bitterman et al., 1983*). Similarly rodents will  
 521 start digging for food in a scented bowl after just a few pairings with that odor (*Cleland et al.,*  
 522 *2009*). Again, these are 1-bit tasks learned rapidly after one or a few experiences.

523 By comparison the tasks that a mouse performs in the labyrinth are more complex. For  
 524 example, the path from the maze entrance to the water port involves 6 junctions, each with 3  
 525 options. At a minimum 6 different contexts must be mapped correctly into one of 3 actions  
 526 each, which involves  $6 \cdot \log_2 3 = 9.5$  bits of complexity. The animals begin to execute perfect  
 527 paths from the entrance to the water port well within the first hour (*Figure 2C, Figure 3B*).  
 528 At a later stage during the night the animal learns to walk direct paths to water from many  
 529 different locations in the maze (*Figure 5*); by this time it has consumed 10-20 rewards. In  
 530 the limit, if the animal could turn correctly towards water from each of 63 junctions in the  
 531 maze, it would have learned  $63 \cdot \log_2 3 = 100$  bits. Conservatively we estimate that the animals  
 532 have mastered 10-20 bits of complexity based on 10-20 reward experiences within an hour of  
 533 time spent in the maze. Note this considers only information about the water port and ignores  
 534 whatever else the animals are learning about the maze during their incessant exploratory forays.

535 These numbers align well with classic experiments on rats in diverse mazes and problem boxes  
 536 *Munn (1950a)*. Although those tasks come in many varieties, a common theme is that ~10  
 537 successful trials are sufficient to learn ~10 decisions (*Woodrow, 1942*).

538 In a different corner of the speed-complexity space are the many 2-alternative-forced-choice  
 539 (2AFC) tasks in popular use today. These tend to be 1-bit tasks, for example the monkey should  
 540 flick its eyes to the left when visual motion is to the left (*Newsome and Pare, 1988*), or the  
 541 mouse should turn a steering wheel to the right when a light appears on the left (*Burgess et al.,*  
 542 *2017*). Yet, the animals take a long time to learn these simple tasks. For example, the mouse  
 543 with the steering wheel requires about 10,000 experiences before performance saturates. It  
 544 never gets particularly good, with a typical hit rate only 2/3 of the way from random to perfect.  
 545 All this training takes 3-6 weeks; in the case of monkeys several months. The rate of learning,  
 546 measured in task complexity per unit time, is surprisingly low: < 1 bit/month compared to ~10  
 547 bits/h observed in the labyrinth. The difference is a factor of 6,000. Similarly when measured in  
 548 complexity learned per reward experience: The 2AFC mouse may need 5,000 rewards to learn  
 549 a contingency table with 1 bit complexity, whereas the mouse in the maze needs ~10 rewards  
 550 to learn 10 bits. Given these enormous differences in learning rate, one wonders whether the  
 551 ultra-slow mode of learning has any relevance for an animal's natural condition. In the month  
 552 that the 2AFC mouse requires to finally report the location of a light, its relative in the wild has  
 553 developed from a baby to having its own babies. Along the way, that wild mouse had to make  
 554 many decisions, often involving high stakes, without the benefit of 10,000 trials of practice.

### 555 Sudden insight

556 The dynamics of the learning process are often conceived as a continuously growing associa-  
 557 tion between stimuli and actions, with each reinforcing experience making an infinitesimal  
 558 contribution. The reality can be quite different. When a child first learns to balance on a bicycle,  
 559 performance goes from abysmal to astounding within a few seconds. The timing of such a  
 560 discontinuous step in performance seems impossible to predict but easy to recognize after the  
 561 fact.

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

571 Owing to the unstructured nature of the experiment, the mouse may adopt different policies  
 572 for getting to the water port. In at least half the animals we observed a discontinuous change  
 573 in that policy, namely when the animal started using efficient direct paths within the maze  
 574 (*Figure 5, Figure 5-figure supplement 2*). This second switch happened considerably after  
 575 the animal started collecting rewards, and did not greatly affect the reward rate. Furthermore,  
 576 the animals never reverted to the less efficient policy, just as a child rarely unlearns to balance  
 577 a bicycle.

578 Presumably this switch in performance reflects some discontinuous change in the animal's  
 579 internal model of the maze, what Tolman called the "cognitive map" (*Tolman, 1948; Behrens*  
 580 *et al., 2018*). In the unrewarded animals we could not detect any discontinuous change in the  
 581 use of long paths. However, as Tolman argued, those animals may well acquire a sophisticated  
 582 cognitive map that reveals itself only when presented with a concrete task, like finding water.

583 Future experiments will need to address this. The discontinuous changes in performance pose  
 584 a challenge to conventional models of reinforcement learning, in which reward events are the  
 585 primary driver of learning and each event contributes an infinitesimal update to the action  
 586 policy. It will also be important to model the acquisition of distinct kinds of knowledge that  
 587 contribute to the same behavior, like the location of the target and efficient routes to approach  
 588 it.

### 589 **Exploratory behavior**

590 By all accounts the animals spent a large fraction of the night exploring the maze (*Figure 1–*  
*591 figure supplement 2*). The water-deprived animals continued their forays into the depths of  
 592 the maze long after they had found the water port and learned to exploit it regularly. After  
 593 consuming a water reward they wandered off into the maze 90% of the time (*Figure 7B*) instead  
 594 of lazily waiting in front of the port during the timeout period. The sated animals experienced  
 595 no overt reward from the maze, yet they likewise spent nearly half their time exploring that  
 596 environment. As has been noted many times, animals – like humans – derive some form of  
 597 intrinsic reward from exploration (*Berlyne, 1960*). Some have suggested that there exists a  
 598 homeostatic drive akin to hunger and thirst that elicits the information-seeking activity, and  
 599 that the drive is in turn sated by the act of exploration (*Hughes, 1997*). If this were the case,  
 600 then the drive to explore should be weakest just after an episode of exploration, much as the  
 601 drive for food-seeking is weaker after a big meal.

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

613 When left to their own devices, mice choose to spend much of their time engaged in  
 614 exploration. One wonders how that affects their actions when they are strapped into a rigid  
 615 behavior machine, like a 2AFC choice box. Presumably the drive to explore persists, perhaps  
 616 more so because the forced environment is so unpleasant. And within the confines of the two  
 617 alternatives, the only act of exploration the mouse has left is to give the wrong answer. This  
 618 would manifest as an unexpectedly high error rate on unambiguous stimuli, sometimes called  
 619 the "lapse rate" (*Carandini and Churchland, 2013; Pisupati et al., 2021*). The fact that the  
 620 lapse rate decreases only gradually over weeks to months of training (*Burgess et al., 2017*)  
 621 suggests that it is difficult to crush the animal's drive to explore.

622 The animals in our experiments had never been presented with a maze environment, yet they  
 623 quickly settled into a steady mode of exploration. Once a mouse progressed beyond the first  
 624 intersection it typically entered deep into the maze to one or more end nodes (*Figure 6*). Within  
 625 50 s of the first entry the animals adopted a steady speed of locomotion that they would retain  
 626 throughout the night (*Figure 2–figure supplement 1*). Within 250 s of first contact with the  
 627 maze the average animal already spent 50% of its time there (*Figure 1–figure supplement 2*).  
 628 Contrast this with a recent study of "free exploration" in an exposed arena: Those animals  
 629 required several hours before they even completed one walk around the perimeter (*Fonio et al.,  
 630 2009*). Here the drive to explore is clearly pitted against fear of the open space, which may not

631 be conducive to observing exploration *per se*.

632 The persistence of exploration throughout the entire duration of the experiment suggests  
 633 that the animals are continuously surveying the environment, perhaps expecting new features  
 634 to arise. These surveys are quite efficient: The animals cover all parts of the maze much faster  
 635 than expected from a random walk (*Figure 8*). Effectively they avoid re-entering territory they  
 636 surveyed just recently. It is often assumed that this requires some global memory of places  
 637 visited in the environment (Nagy *et al.*, 2020; Olton, 1979). Such memory would have to  
 638 persist for a long time: Surveying half of the available end nodes typically required 450 turning  
 639 decisions. However, we found that a global long-term memory is not needed to explain the  
 640 efficient search. The animals seem to be governed by a set of local turning biases that require  
 641 memory only of the most recent decision and no knowledge of location (*Figure 9*). These local  
 642 biases alone can explain most of the character of exploration without any global understanding  
 643 or long-term memory. Incidentally, they also explain other seemingly global aspects of the  
 644 behavior, for example the systematic preference that the mice have for the outer rather than the  
 645 inner regions of the maze (*Figure 10*). Of course, this argument does not exclude the presence  
 646 of a long-term memory, which may reveal itself in some other feature of the behavior.

647 Perhaps the most remarkable aspect of these biases is how similar they are across all 19 mice  
 648 studied here, regardless of whether the animal experienced water rewards or not (*Figure 9B*,  
*Figure 9-figure supplement 1*), and independent of the sex of the mouse. The four decision  
 649 probabilities were identical across individuals to within a standard deviation of <0.03. We  
 650 cannot think of a trivial reason why this should be so. For example the two biases for forward  
 651 motion (*Figure 9B* left) are poised halfway between the value for a random walk ( $p = 2/3$ ) and  
 652 certainty ( $p = 1$ ). At either of those extremes, simple saturation might lead to a reproducible  
 653 value, but not in the middle of the range. Why do different animals follow the exact same  
 654 decision rules at an intersection between tunnels? Given that tunnel systems are part of the  
 655 mouse's natural ecology, it is possible that those rules are innate and determined genetically.  
 656 Indeed the rules by which mice build tunnels have a strong genetic component (Weber *et al.*,  
 657 2013), so the rules for using tunnels may be written in the genes as well. The high precision  
 658 with which one can measure those behaviors even in a single night of activity opens the way to  
 659 efficient comparisons across genotypes, and also across animals with different developmental  
 660 experience.

661 Finally, after mice discover the water port and learn to access it from many different points  
 662 in the maze (*Figure 5*) they are presumably eager to discover other things. In ongoing work we  
 663 installed three water ports (visible in the videos accompanying this article) and implemented a  
 664 rule that activates the three ports in a cyclic sequence. Mice discovered all three ports rapidly  
 665 and learned to visit them in the correct order. Future experiments will have to raise the bar on  
 666 what the mice are expected to learn in a night.

## 668 Mechanisms of navigation

669 How do the animals navigate when they perform direct paths to the water port or to the exit?  
 670 The present study cannot resolve that, but one can gain some clues based on observations so  
 671 far. Early workers already concluded that rodents in a maze will use whatever sensory cues  
 672 and tricks are available to accomplish their tasks (Munn, 1950b). Our maze was designed to  
 673 restrict those options somewhat.

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

679 With regard to the sense of touch and kinesthetics, the maze was constructed for maximal  
 680 symmetry. At each level of the binary tree all the junctions have locally identical geometry,  
 681 with intersecting corridors of the same length. In practice the animals may well detect some  
 682 inadvertent cues, like an unusual drop of glue, that could identify one node from another. The  
 683 maze rotation experiment suggests that such cues are not essential for the animal's sense of  
 684 location in the maze, at least in the expert phase.

685 The role of odors deserves particular attention because the mouse may use them both  
 686 passively and actively. Does the animal first find the water port by following the smell of water?  
 687 Probably not. For one, the port only emits a single drop of water when triggered by a nose poke.  
 688 Second, we observed many instances where the animal is in the final corridor adjacent to the  
 689 water port yet fails to discover it. The initial discovery seems to occur via touch. The reader can  
 690 verify this in the videos accompanying this article. Regarding active use of odor markings in  
 691 the maze, the maze rotation experiment suggests that such cues are not required for navigation,  
 692 at least once the animals have adopted the shortest path to the water port (*Figure 4*).

693 Another algorithm that is often invoked for animals moving in an open arena is vector-based  
 694 navigation (*Wehner et al., 1996*). Once the animal discovers a target, it keeps track of that  
 695 target's heading and distance using a path integrator. When it needs to return to the target it  
 696 follows the heading vector and updates heading and distance until it arrives. Such a strategy has  
 697 limited appeal inside a labyrinth because the vectors are constantly blocked by walls. Consider,  
 698 for example, the "home runs" back to the exit at the end of a bout. Here the target, namely the  
 699 exit, is known from the start of the bout, because the animal enters through the same hole. At  
 700 the end of the bout, when the mouse decides to exit from the maze, can it follow the heading  
 701 vector to the exit? *Figure 6A* shows the 13 locations from which mice returned in a direct path  
 702 to the exit on their very first foray. None of these locations is compatible with heading-based  
 703 navigation: In each case an animal following the heading to the exit would get stuck in a  
 704 different end node first and would have to reverse from there, quite unlike what really happened.

705 Finally, a partial clue comes from errors the animals make. We found that the rotation image  
 706 of the water port, an end node diametrically across the entire maze, is one of the most popular  
 707 destinations for rewarded animals (*Figure 10A*). These errors would be highly unexpected  
 708 if the animals navigated from the entrance to the water by odor markings, or if they used an  
 709 absolute representation of heading and distance. On the other hand, if the animal navigates via  
 710 a remembered sequence of turns, then it will end up at that image node if it makes a single  
 711 mistake at just the first T-junction.

712 Future directed experiments will serve to narrow down how mice learn to navigate this  
 713 environment, and how their policy might change over time. Since the animals get to perfection  
 714 within an hour or so, one can test a new hypothesis quite efficiently. Understanding what  
 715 mechanisms they use will then inform thinking about the algorithm for learning, and about the  
 716 neuronal mechanisms that implement it.

## 717 Methods and Materials

### 718 Experimental design

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

### 724 Maze construction

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

### 737 Reward delivery system

738 The water reward port was controlled by a Matlab script on the main computer through an  
 739 interface (Sanworks Bpod State Machine r1). Rewards were triggered when the animal's nose  
 740 broke the IR beam in the water port (Sanworks Port interface + valve). The interface briefly  
 741 opened the water valve to deliver ~30  $\mu$ L of water and flashed an infrared LED mounted outside  
 742 the maze for 1 s. This served to mark reward events on the video recording. Following each  
 743 reward, the system entered a time-out period for 90 s, during which the port did not provide  
 744 further reward. In experiments with sated mice the water port was turned off.

### 745 Cage and connecting passage

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

### 749 Animals and treatments

750 All mice were C57BL/6J animals (Jackson Labs) between the ages of 45 and 98 days (mean 62  
 751 days). Both sexes were used: 4 males and 6 females in the rewarded experiments, 5 males and  
 752 4 females in the unrewarded experiments. For water deprivation, the animal was transferred  
 753 from its home cage (generally group-housed) to the maze cage ~22 h before the start of the  
 754 experiment. Non-deprived animals were transferred minutes before the start. All procedures  
 755 were performed in accordance with institutional guidelines and approved by the Caltech IACUC.

### 756 Video recording

757 All data reported here were collected over the course of 7 hours during the dark portion of  
 758 the animal's light cycle. Video recording was initiated a few seconds prior to connecting the  
 759 tunnel to the maze. Videos were recorded by an OpenCV python script controlling a single  
 760 webcam (Logitech C920) located ~1 m below the floor of the maze. The maze and access tube

were illuminated by multiple infrared LED arrays (center wavelength 850 nm). Three of these lights illuminated the maze from below at a 45 degree angle, producing contrast to resolve the animal's foot pads. The remaining lights pointed at the ceiling of the room to produce backlight for a sharp outline of the animal.

### 765 Animal tracking

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

### 770 Rates of transition between cage and maze

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

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

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

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

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

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

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

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

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

781 The rate of entering the maze is highest at short times in the cage (*Figure 1–figure supplement*  
 782 *3A*). It peaks after ~15 s in the cage and then declines gradually by a factor of 4 over  
 783 the first minute. So the mouse is most likely to enter the maze just after it returns from there.  
 784 This runs opposite to the expectation from a homeostatic drive for exploration, which should  
 785 be sated right after the animal returns. We found no evidence for an increase in the rate at  
 786 late times. These effects were very similar in rewarded and unrewarded groups and in fact the  
 787 tendency to return early was seen in every animal.

788 By contrast the rate of exiting the maze is almost perfectly constant over time (*Figure 1–*  
*789 figure supplement 3B*). In other words the exit from the maze appears like a constant rate  
 790 Poisson process. There is a slight elevation of the rate at short times among rewarded animals

791 (*Figure 1–figure supplement 3B top*). This may come from the occasional brief water runs  
 792 they perform. Another strange deviation is an unusual number of very short bouts (duration  
 793 2-12 s) among unrewarded animals (*Figure 1–figure supplement 3B bottom*). These are brief  
 794 excursions in which the animal runs to the central junction, turns around, and runs to the exit.  
 795 Several animals exhibited these, often several bouts in a row, and at all times of the night.

### 796 Reduced trajectories

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

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

810 In *Figure 5* and *Figure 6* we count the occurrence of **direct paths** leading to the water  
 811 port (a “water run”) or to the exit (a “home run”). A direct path is a node sequence without any  
 812 reversals. *Figure 3–figure supplement 1* illustrates some examples.

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

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

### 817 Maze rotation

818 The maze rotation experiment (*Figure 4*) was performed on 4 mice, all water-deprived. Two  
 819 of the animals (’D7’ and ’D9’) had experienced the maze before, and are part of the ‘rewarded’  
 820 group in other sections of the report. Two additional animals (’F2’ and ’A1’) had had no prior  
 821 contact with the maze.

822 The maze rotation occurred after at least 6 hours of exposure, by which time the animals  
 823 had all perfected the direct path to the water port.

824 For animals ’D7’ and ’D9’ we rotated only the floor of the maze, leaving the walls and  
 825 ceiling in the original configuration. For ’F2’ and ’A1’ we rotated the entire maze, moving one  
 826 wall segment at the central junction and the water port to attain the same shape. Navigation  
 827 remained intact for all animals. Note that ’A1’ performed a perfect path to the water port and  
 828 back immediately before and after a full maze rotation (*Figure 4B*).

829 The visits to the 4 locations in the maze (*Figure 4C*, *Figure 4–figure supplement 1*) were  
 830 limited to direct paths of length at least 2 steps. This avoids counting rapid flickers between  
 831 two adjacent nodes. In other words, the animal has to move at least 2 steps away from the target  
 832 node before another visit qualifies.

833 **Statistics of sudden insight**

834 In *Figure 5* one can distinguish two events: First the animal finds the water port and begins  
 835 to collect rewards at a steady rate: this is when the green curve rises up. At a later time the  
 836 long direct paths to the water port become much more frequent than to the comparable control  
 837 nodes: this is when the red and blue curves diverge. For almost all animals these two events are  
 838 well separated in time (*Figure 5-figure supplement 1*). In many cases the rate of long paths  
 839 seems to change discontinuously: a sudden change in slope of the curve.

840 Here we analyze the degree of "sudden change", namely how rapidly the rate changes in a  
 841 time series of events. We modeled the rate as a sigmoid function of time during the experiment:

$$r(t) = r_i + \frac{r_f - r_i}{2} \operatorname{erf}\left(\frac{t - t_s}{w}\right) \quad (8)$$

842 where

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2} dx$$

843 The rate begins at a low initial level  $r_i$ , reflecting chance occurrence of the event, and  
 844 saturates at a high final level  $r_f$ , limited for example by the animal's walking speed. The other  
 845 two parameters are the time  $t_s$  of half-maximal rate change, and the width  $w$  over which that  
 846 rate change takes place. A sudden change in the event rate would correspond to  $w = 0$ .

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

$$L[r(t)] = e^{-\int_0^T r(t) dt} \prod_i r(t_i) \quad (9)$$

850 and the log likelihood is

$$\ln L = \sum_i \ln r(t_i) - \int_0^T r(t) dt \quad (10)$$

851 For each of the 10 rewarded mice, we maximized  $\ln L$  over the 4 parameters of the rate  
 852 model, both for the reward events and the long paths to water. The resulting fits are plotted in  
 853 *Figure 5-figure supplement 1*.

854 Focusing on the learning of long paths to water, for 6 of the 10 animals the optimal width  
 855 parameter  $w$  was less than 300 s: B1, B2, C1, C3, C6, C7. These are the same animals one  
 856 would credit with a sudden kink in the cumulative event count based on visual inspection  
 857 (*Figure 5-figure supplement 1*).

858 To measure the uncertainty in the timing of this step, we refit the data for this subgroup of  
 859 mice with a model involving a sudden step in the rate,

$$r(t) = \begin{cases} r_i, & t < t_s \\ r_f, & t > t_s \end{cases} \quad (11)$$

860 and computed the likelihood of the data as a function of the step time  $t_s$ . We report the mean  
 861 and standard deviation of the step time over its likelihood in *Figure 5-figure supplement 2*.  
 862 Animal C6 was dropped from this "sudden step" group, because the uncertainty in the step  
 863 time was too large (~900 s).

864    **Efficiency of exploration**

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

869    We restricted the animal's node trajectory ( $s_i$ ) to clips of exploration mode, excluding the  
 870    direct paths to the water port or the exit. All subsequent steps were applied to these clips, then  
 871    averaged over clips. Within each clip we marked the sequence of end nodes ( $e_i$ ). We slid a  
 872    window of size  $n$  across this sequence and counted the number of distinct nodes  $d$  in each  
 873    window. Then we averaged  $d$  over all windows in all clips. Then we repeated that for a wide  
 874    range of  $n$ . The resulting  $d(n)$  is plotted in the figures reporting new nodes vs nodes visited  
 875    (*Figure 8A,B* and *Figure 9C*).

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

$$d(n) \approx 64 \left( 1 - \frac{1}{1 + \frac{z+bz^3}{1+b}} \right) \quad (12)$$

877    where

$$z = n / a \quad (13)$$

878    The parameter  $a$  is the number of visits  $n$  required to survey half of the end nodes, whereas  $b$   
 879    reflects a relative acceleration in discovering the last few end nodes. This function was found  
 880    by trial and error and produces absurdly good fits to the data (*Figure 8-figure supplement 1*).  
 881    The values quoted in the text for efficiency of exploration are  $E = 32/a$  (*Equation 1*).

882    The value of  $b$  was generally small (~0.1) with no difference between rewarded and unre-  
 883    warded animals. It declined slightly over the night (*Figure 8-figure supplement 1B*), along  
 884    with the decline in  $a$  (*Figure 8C*).

885    **Biased random walk**

886    For the analysis of *Figure 9* we considered only the parts of the trajectory during 'exploration'  
 887    mode. Then we parsed every step between two nodes in terms of the type of action it represents.  
 888    Note that every link between nodes in the maze is either a 'left branch' or a 'right branch',  
 889    depending on its relationship to the parent T-junction. Therefore there are 4 kinds of action:

- 890       •  $a = 0$ : 'in left', take a left branch into the maze
- 891       •  $a = 1$ : 'in right', take a right branch into the maze
- 892       •  $a = 2$ : 'out left', take a left branch out of the maze
- 893       •  $a = 3$ : 'out right', take a right branch out of the maze

894    At any given node some actions are not available, for example from an end node one can  
 895    only take one of the 'out' actions.

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

$$P_{SF} = \frac{p(0|0) + p(0|1) + p(1|0) + p(1|1)}{p(0|0) + p(0|1) + p(1|0) + p(1|1) + p(2|0) + p(3|1)} \quad (14)$$

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

$$P_{BF} = \frac{p(0|3) + p(1|2) + p(2|2) + p(2|3) + p(3|2) + p(3|3)}{p(0|3) + p(1|2) + p(2|2) + p(2|3) + p(3|2) + p(3|3) + p(0|2) + p(1|3)} \quad (16)$$

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

900 For the simulations of random agents (*Figure 8*, *Figure 9*) we used trajectories long enough  
 901 so the uncertainty in the resulting curves was smaller than the line width.

## 902 Models of decisions during exploration

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

908 For every mouse trajectory we split the data into 5 segments, trained the model on 80% of  
 909 the data, and tested it on 20%, averaging the resulting cross-entropy over the 5 possible splits.  
 910 Each segment was in turn composed of parts of the trajectory sampled evenly throughout the  
 911 7-h experiment, so as to average over the small changes in the course of the night. The model  
 912 was evaluated by the cross-entropy between the predictions and the animal's true actions. If  
 913 one had an optimal model of behavior, the result would reveal the animal's true source entropy.

### 914 Fixed depth Markov chain

915 To fit a model with fixed history depth  $k$  to a measured node sequence  $(s_t)$ , we evaluated  
 916 all the substrings in that sequence of length  $(k + 1)$ . At any given time  $t$ , the  $k$ -string  $\mathbf{h}_t =$   
 917  $(s_{t-k+1}, \dots, s_t)$  identifies the history of the animal's  $k$  most recent locations. The current state  
 918  $s_t$  is one of 63 T-junctions. Each state is preceded by one of 3 possible states. So the number  
 919 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  
 920 be 'in left', 'in right', or 'out', corresponding to the 3 branches of the T junction. Tallying  
 921 the history strings with the resulting actions leads to a contingency table of size  $63 \cdot 3^{k-1} \times 3$ ,  
 922 containing

$$n(\mathbf{h}, a) = \text{number of times history } \mathbf{h} \text{ leads to action } a \quad (18)$$

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

$$p(a | \mathbf{h}) = \frac{n(\mathbf{h}, a) + 1}{\sum_{a'} n(\mathbf{h}, a') + 3} \quad (19)$$

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

$$H = \langle -\log_2 p(a_t | \mathbf{h}_t) \rangle_t \quad (20)$$

929    Variable depth Markov chain

930    As one pushes to longer histories, i.e. larger  $k$ , the analysis quickly becomes data-limited,  
 931    because the number of possible histories grows exponentially with  $k$ . Soon one finds that  
 932    the counts for each history-action combination drop to where one can no longer estimate  
 933    probabilities correctly. In an attempt to offset this problem we pruned the history tree such that  
 934    each surviving branch had more than some minimal number of counts in the training data. As  
 935    expected, this model is less prone to over-fitting and degrades more gently as one extends to  
 936    longer histories (*Figure 11–figure supplement 1A*). The lowest cross-entropy was obtained  
 937    with an average history length of ~4.0 but including some paths of up to length 6. Of all the  
 938    algorithms we tested, this produced the lowest cross-entropies, although the gains relative to  
 939    the fixed-depth model were modest (*Figure 11–figure supplement 1C*).

940    Pooling across symmetric nodes in the maze

941    Another attempt to increase the counts for each history involved pooling counts over multiple  
 942    T-junctions in the maze that are closely related by symmetry. For example, all the T-junctions at  
 943    the same level of the binary tree look locally similar, in that they all have corridors of identical  
 944    length leading from the junction. If one supposes that the animal acts the same way at each  
 945    of those junctions, one would be justified in pooling across these nodes, leading to a better  
 946    estimate of the action probabilities, and perhaps less over-fitting. This particular procedure  
 947    was unsuccessful, in that it produced higher cross-entropy than without pooling.

948    However, one may want to distinguish two types of junctions within a given level: L-nodes  
 949    are reached by a left branch from their parent junction one level lower in the tree, R-nodes  
 950    by a right branch. For example, in *Figure 3–figure supplement 1*, node 1 is L-type and node  
 951    2 is R-type. When we pooled histories over all the L-nodes at a given level and separately  
 952    over all the R-nodes the cross-entropy indeed dropped, by about 5% on average. This pooling  
 953    greatly reduced the amount of over-fitting (*Figure 11–figure supplement 1B*), which allowed  
 954    the use of longer histories, which in turn improved the predictions on test data. The benefit of  
 955    distinguishing L- and R-nodes probably relates to the animal’s tendency to alternate left and  
 956    right turns.

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

959    **Data availability**

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

962    **Acknowledgments**

963    **Funding:** This work was supported by the Simons Collaboration on the Global Brain (grant  
 964    543015 to MM and 543025 to PP), by NSF award 1564330 to PP, and by a gift from Google to  
 965    PP.

966    **Author contributions:** Conception of the study MR, TZ, PP, MM; Data collection MR, TZ;  
 967    Analysis and interpretation MR, TZ, PP, MM; Drafting the manuscript MM; Revision and  
 968    approval MR, TZ, PP, MM.

969    **Competing interests:** The authors declare no competing interests.

970    **Data and code availability:** Data and code will be available in a permanent public repository  
 971    following acceptance of the manuscript.

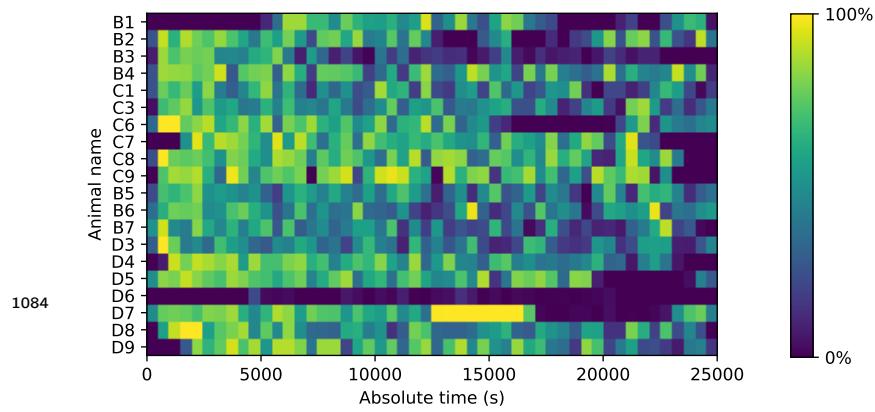
972    **Colleagues:** We thank Ben de Bivort, Loren Frank, Lisa Giocomo, Konrad Kording,  
 973    Clayton Lewis, Bence Ölveczky, and Xaq Pitkow for helpful discussions and comments.

## 974 References

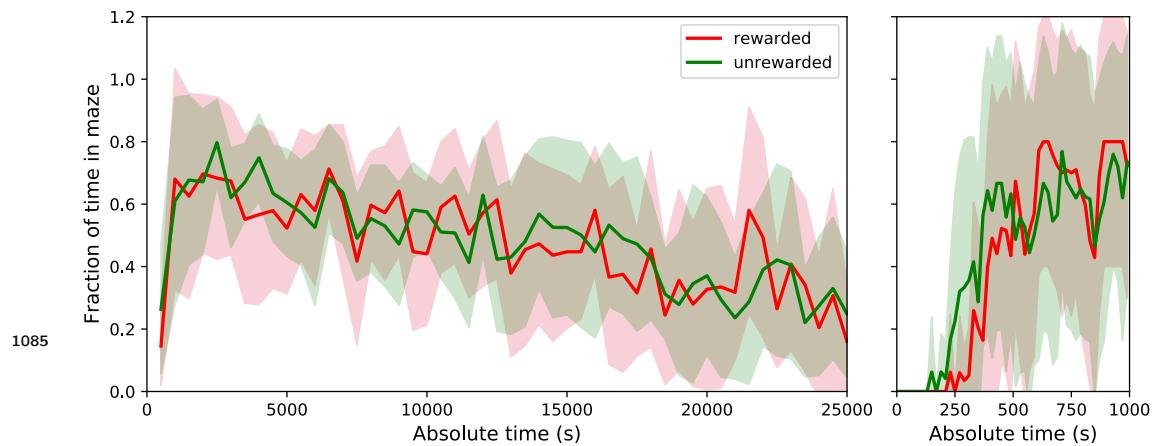
- 975 Alonso A, van der Meij J, Tse D, Genzel L. Naïve to Expert: Considering the Role of Previous Knowledge  
976 in Memory. *Brain and Neuroscience Advances*. 2020 Jan; 4:1–17. doi: [10.1177/2398212820948686](https://doi.org/10.1177/2398212820948686).
- 977 Behrens TEJ, Muller TH, Whittington JCR, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson  
978 Z. What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*. 2018 Oct;  
979 100(2):490–509. doi: [10.1016/j.neuron.2018.10.002](https://doi.org/10.1016/j.neuron.2018.10.002).
- 980 Berlyne DE. Conflict, Arousal, and Curiosity. New York, NY, US: McGraw-Hill Book Company; 1960.  
981 doi: [10.1037/11164-000](https://doi.org/10.1037/11164-000).
- 982 Bitterman ME, Menzel R, Fietz A, Schäfer S. Classical Conditioning of Proboscis Extension in Honey-  
983 bees (*Apis Mellifera*). *Journal of Comparative Psychology*. 1983; 97(2):107–119. doi: [10.1037/0735-7036.97.2.107](https://doi.org/10.1037/0735-7036.97.2.107).
- 985 Bourtchuladze R, Frenguelli B, Blendy J, Cioffi D, Schutz G, Silva AJ. Deficient Long-Term Memory  
986 in Mice with a Targeted Mutation of the cAMP-Responsive Element-Binding Protein. *Cell*. 1994  
987 Oct; 79(1):59–68. doi: [10.1016/0092-8674\(94\)90400-6](https://doi.org/10.1016/0092-8674(94)90400-6).
- 988 Brennan PA, Keverne EB. Neural Mechanisms of Mammalian Olfactory Learning. *Progress in  
989 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).
- 990 Bruce HM. An Exteroceptive Block to Pregnancy in the Mouse. *Nature*. 1959 Jul; 184:105. doi:  
991 [10.1038/184105a0](https://doi.org/10.1038/184105a0).
- 992 Buel J. The Linear Maze. I. "Choice-Point Expectancy," "Correctness," and the Goal Gradient. *Journal  
993 of Comparative Psychology*. 1934; 17(2):185–199. doi: [10.1037/h0072346](https://doi.org/10.1037/h0072346).
- 994 Burgess CP, Lak A, Steinmetz NA, Zatka-Haas P, Bai Reddy C, Jacobs EAK, Linden JF, Paton JJ,  
995 Ranson A, Schröder S, Soares S, Wells MJ, Wool LE, Harris KD, Carandini M. High-Yield Methods  
996 for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. *Cell Reports*. 2017 Sep;  
997 20(10):2513–2524. doi: [10.1016/j.celrep.2017.08.047](https://doi.org/10.1016/j.celrep.2017.08.047).
- 998 Carandini M, Churchland AK. Probing Perceptual Decisions in Rodents. *Nature Neuroscience*. 2013  
999 Jul; 16:824–31. doi: [10.1038/nn.3410](https://doi.org/10.1038/nn.3410).
- 1000 Cleland TA, Narla VA, Boudadi K. Multiple Learning Parameters Differentially Regulate Olfactory  
1001 Generalization. *Behavioral Neuroscience*. 2009 Feb; 123(1):26–35. doi: [10.1037/a0013991](https://doi.org/10.1037/a0013991).
- 1002 Estes W. The Problem of Inference from Curves Based on Group Data. *Psychological Bulletin*. 1956;  
1003 53(2):134–140. doi: [10.1037/h0045156](https://doi.org/10.1037/h0045156).
- 1004 Fanselow M, Bolles R. Naloxone and Shock-Elicited Freezing in the Rat. *Journal of comparative and  
1005 physiological psychology*. 1979 Sep; 93:736–44. doi: [10.1037/h0077609](https://doi.org/10.1037/h0077609).
- 1006 Fonio E, Benjamini Y, Golani I. Freedom of Movement and the Stability of Its Unfolding in Free  
1007 Exploration of Mice. *Proceedings of the National Academy of Sciences of the United States of  
1008 America*. 2009 Dec; 106(50):21335–21340. doi: [10.1073/pnas.0812513106](https://doi.org/10.1073/pnas.0812513106).
- 1009 Gallistel CR, Fairhurst S, Balsam P. The Learning Curve: Implications of a Quantitative Analysis.  
1010 *Proceedings of the National Academy of Sciences of the United States of America*. 2004 Sep;  
1011 101(36):13124–13131. doi: [10.1073/pnas.0404965101](https://doi.org/10.1073/pnas.0404965101).
- 1012 Grobéty MC, Schenk F. Spatial Learning in a Three-Dimensional Maze. *Animal Behaviour*. 1992 Jun;  
1013 43(6):1011–1020. doi: [10.1016/S0003-3472\(06\)80014-X](https://doi.org/10.1016/S0003-3472(06)80014-X).
- 1014 Guo ZV, Li N, Huber D, Ophir E, Gutnisky D, Ting JT, Feng G, Svoboda K. Flow of Corti-  
1015 cal Activity Underlying a Tactile Decision in Mice. *Neuron*. 2014 Jan; 81(1):179–194. doi:  
1016 [10.1016/j.neuron.2013.10.020](https://doi.org/10.1016/j.neuron.2013.10.020).

- 1017    Hughes RN. Intrinsic Exploration in Animals: Motives and Measurement. *Behavioural Processes*.  
 1018    1997 Dec; 41(3):213–226. doi: 10.1016/S0376-6357(97)00055-7.
- 1019    Krechevsky I. "Hypotheses" in Rats. *Psychological Review*. 1932; 39(6):516–532. doi:  
 1020    10.1037/h0073500.
- 1021    LeDoux JE. Emotion Circuits in the Brain. *Annual Review of Neuroscience*. 2000; 23:155–184. doi:  
 1022    10.1146/annurev.neuro.23.1.155.
- 1023    McNamara CG, Tejero-Cantero Á, Trouche S, Campo-Urriza N, Dupret D. Dopaminergic Neurons  
 1024    Promote Hippocampal Reactivation and Spatial Memory Persistence. *Nature Neuroscience*. 2014  
 1025    Dec; 17(12):1658–1660. doi: 10.1038/nn.3843.
- 1026    Munn NL. The Learning Process. In: *Handbook of Psychological Research on the Rat; an Introduction*  
 1027    to Animal Psychology Oxford, England: Houghton Mifflin; 1950.p. 226–288.
- 1028    Munn NL. The Role of Sensory Processes in Maze Behavior. In: *Handbook of Psychological Research*  
 1029    on the Rat; an Introduction to Animal Psychology Oxford, England: Houghton Mifflin; 1950.p.  
 1030    181–225.
- 1031    Nagy M, Horicsányi A, Kubinyi E, Couzin ID, Vásárhelyi G, Flack A, Vicsek T. Synergistic Benefits  
 1032    of Group Search in Rats. *Current Biology*. 2020 Sep; doi: 10.1016/j.cub.2020.08.079.
- 1033    Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW. Using DeepLabCut for 3D Markerless  
 1034    Pose Estimation across Species and Behaviors. *Nature Protocols*. 2019 Jul; 14(7):2152–2176. doi:  
 1035    10.1038/s41596-019-0176-0.
- 1036    Newsome WT, Pare EB. A Selective Impairment of Motion Perception Following Lesions of the  
 1037    Middle Temporal Visual Area (MT). *Journal of Neuroscience*. 1988 Jun; 8(6):2201–2211. doi:  
 1038    10.1523/JNEUROSCI.08-06-02201.1988.
- 1039    Olton D. Mazes, Maps, and Memory. *American Psychologist*. 1979; 34(7):583–596. doi: 10.1037/0003-  
 1040    066X.34.7.583.
- 1041    Pisupati S, Chartarifsky-Lynn L, Khanal A, Churchland AK. Lapses in Perceptual Decisions Reflect  
 1042    Exploration. *eLife*. 2021 Jan; 10:e55490. doi: 10.7554/eLife.55490.
- 1043    Pseudo-Apollodorus. Epitome. In: *Library and Epitome*; I-II Century AD.p. Ch 1 Sec 9.
- 1044    Rondi-Reig L, Petit GH, Tobin C, Tonegawa S, Mariani J, Berthoz A. Impaired Sequential Egocentric  
 1045    and Allocentric Memories in Forebrain-Specific-NMDA Receptor Knock-out Mice during a New  
 1046    Task Dissociating Strategies of Navigation. *The Journal of Neuroscience: The Official Journal of the*  
 1047    *Society for Neuroscience*. 2006 Apr; 26(15):4071–4081. doi: 10.1523/JNEUROSCI.3408-05.2006.
- 1048    Rosser AE, Keverne EB. The Importance of Central Noradrenergic Neurones in the Formation of an  
 1049    Olfactory Memory in the Prevention of Pregnancy Block. *Neuroscience*. 1985 Aug; 15(4):1141–1147.  
 1050    doi: 10.1016/0306-4522(85)90258-1.
- 1051    Sato N, Fujishita C, Yamagishi A. To Take or Not to Take the Shortcut: Flexible Spatial Behaviour of  
 1052    Rats Based on Cognitive Map in a Lattice Maze. *Behavioural Processes*. 2018 Jun; 151:39–43. doi:  
 1053    10.1016/j.beproc.2018.03.010.
- 1054    Seward J, Bzip2; 2019.
- 1055    Shokaku T, Moriyama T, Murakami H, Shinohara S, Manome N, Morioka K. Development of an  
 1056    Automatic Turntable-Type Multiple T-Maze Device and Observation of Pill Bug Behavior. *Review*  
 1057    of Scientific Instruments. 2020 Oct; 91(10):104104. doi: 10.1063/5.0009531.
- 1058    Small WS. Experimental Study of the Mental Processes of the Rat. II. *The American Journal of*  
 1059    *Psychology*. 1901; 12(2):206–239. doi: 10.2307/1412534.

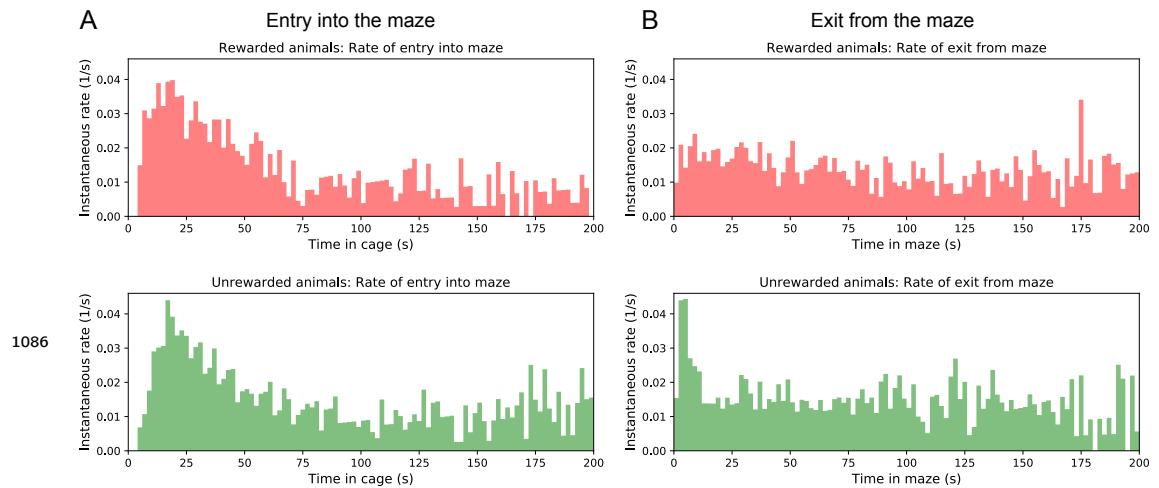
- 1060 Tchernichovski O, Benjamini Y, Golani I. The Dynamics of Long-Term Exploration in the Rat.  
1061 Biological Cybernetics. 1998 Jul; 78(6):423–432. doi: 10.1007/s004220050446.
- 1062 Tolman EC. The Determiners of Behavior at a Choice Point. Psychological Review. 1938; 45:1–41.  
1063 doi: 10.1037/h0062733.
- 1064 Tolman EC. Cognitive Maps in Rats and Men. Psychological Review. 1948; 55(4):189–208. doi:  
1065 10.1037/h0061626.
- 1066 Tolman E, Honzik C. Degrees of Hunger, Reward and Non-Reward, and Maze Learning in Rats.  
1067 University of California Publications in Psychology. 1930; 4:241–256.
- 1068 Uchida N, Mainen ZF. Speed and Accuracy of Olfactory Discrimination in the Rat. Nature Neuroscience.  
1069 2003 Nov; 6(11):1224–1229. doi: 10.1038/nn1142.
- 1070 Uster HJ, Bättig K, Nägeli HH. Effects of Maze Geometry and Experience on Exploratory Behavior in  
1071 the Rat. Animal Learning & Behavior. 1976 Mar; 4(1):84–88. doi: 10.3758/BF03211992.
- 1072 Weber JN, Peterson BK, Hoekstra HE. Discrete Genetic Modules Are Responsible for Complex Burrow  
1073 Evolution in Peromyscus Mice. Nature. 2013 Jan; 493(7432):402–405. doi: 10.1038/nature11816.
- 1074 Wehner R, Michel B, Antonsen P. Visual Navigation in Insects: Coupling of Egocentric and Geocentric  
1075 Information. Journal of Experimental Biology. 1996 Jan; 199(1):129–140.
- 1076 Wood RA, Bauza M, Krupic J, Burton S, Delecate A, Chan D, O’Keefe J. The Honeycomb Maze  
1077 Provides a Novel Test to Study Hippocampal-Dependent Spatial Navigation. Nature. 2018 Feb;  
1078 554(7690):102–105. doi: 10.1038/nature25433.
- 1079 Woodrow H. The Problem of General Quantitative Laws in Psychology. Psychological Bulletin. 1942;  
1080 39(1):1–27. doi: 10.1037/h0058275.
- 1081 Yoder RM, Clark BJ, Brown JE, Lamia MV, Valerio S, Shinder ME, Taube JS. Both Visual and  
1082 Idiothetic Cues Contribute to Head Direction Cell Stability during Navigation along Complex Routes.  
1083 Journal of Neurophysiology. 2011 Mar; 105(6):2989–3001. doi: 10.1152/jn.01041.2010.



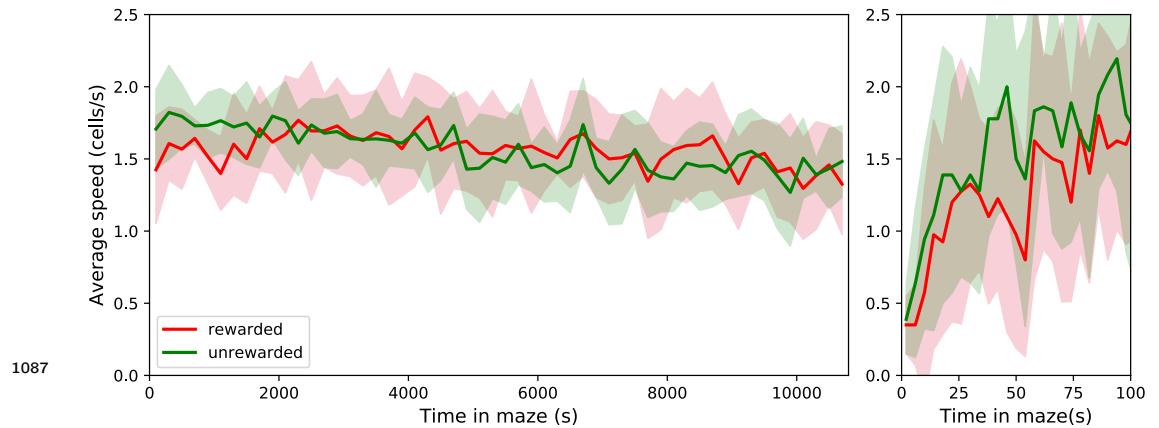
**Figure 1–figure supplement 1. Fraction of time spent in the maze.** Mice could move freely between the home cage and the maze. For each animal (vertical), the fraction of time in the maze (color scale) is plotted as a function of time since start of the experiment. Time bins are 500 s. Note that mouse 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%.



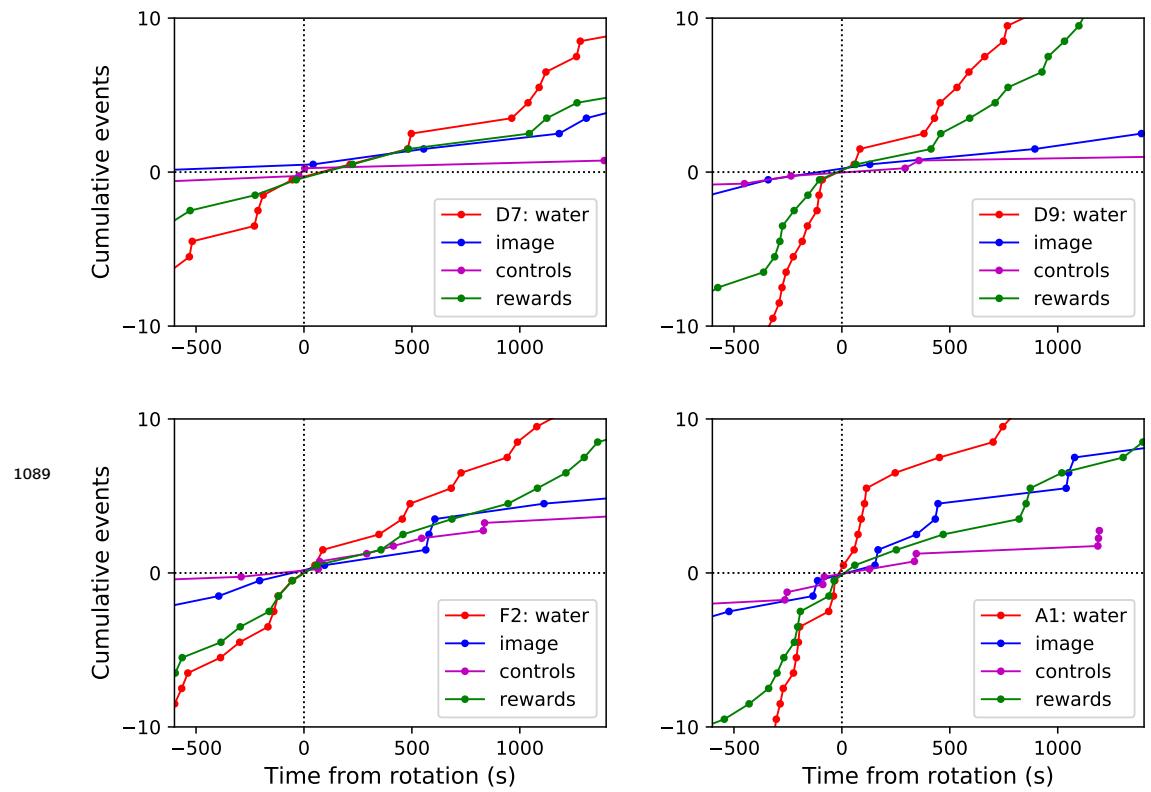
**Figure 1–figure supplement 3. Rates of transition between cage and maze.** (A) The instantaneous probability per unit time  $r_m(t)$  of entering the maze after having spent time  $t$  in the cage. Note this rate is highest immediately upon entering the cage, then declines by a large factor. (B) The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



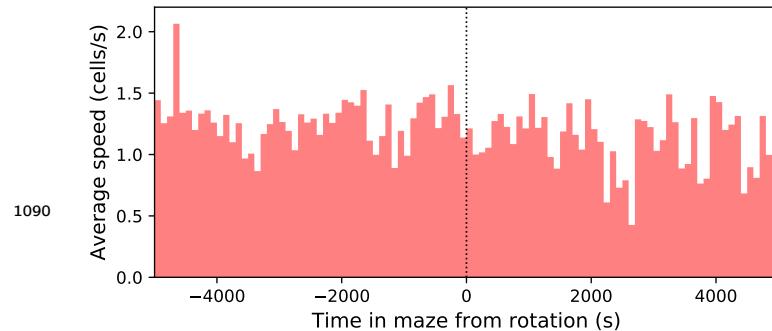
**Figure 2–figure supplement 1. The speed of locomotion in the maze is approximately constant.** Left: Speed plotted as Mean  $\pm$  SD over the population of rewarded and unrewarded animals. Right: expanded axis for early times. To assess the speed of locomotion we divided the maze into square cells 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.



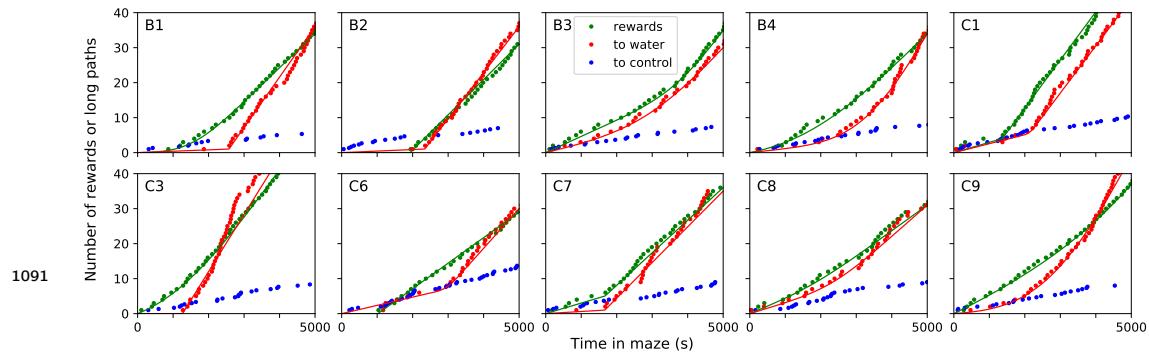
**Figure 3-figure supplement 1. Definition of node trajectories.** A numbering scheme for all 127 nodes of the maze. Green: a direct path from the entrance to the water port (“water run”) with the node sequence ( $s_i$ ) = (0, 2, 6, 13, 28, 57, 116), involving 6 decisions. Magenta: a direct path from 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).



**Figure 4-figure supplement 1. Navigation before and after maze rotation.** Cumulative number of rewards, visits to the water port, the image of the water port, and the control nodes, plotted vs time before and after the maze rotation. Display as in **Figure 4C**, but split for each of 4 animals.



**Figure 4-figure supplement 2. Speed of the mouse vs time in the maze.** Average over 4 animals. Time is plotted relative to the maze rotation.



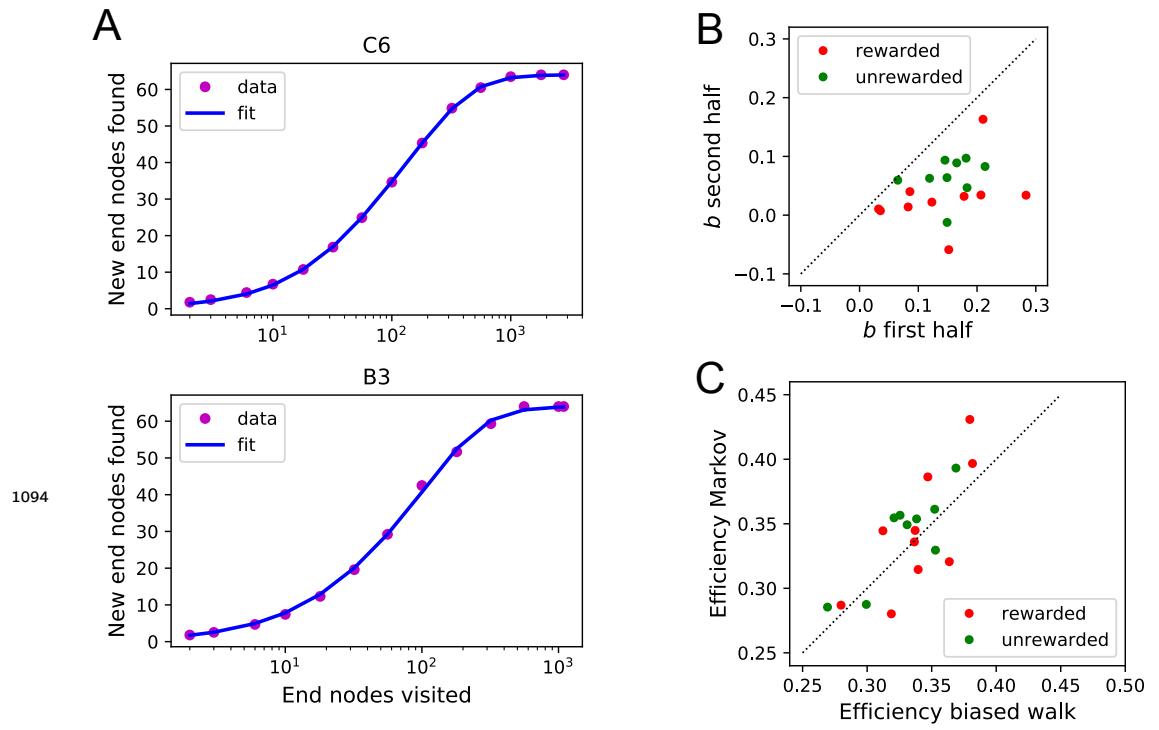
**Figure 5–figure supplement 1. Sudden changes in behavior for all rewarded animals.** For each of the 10 water-deprived animals this shows the cumulative rate of rewards, of long direct paths ( $>6$  steps) to the water port, and of similar paths to 3 control nodes. Display as in **Figure 5**; panels B-D of that figure are included again here. Dots are data, lines are fits using a 4-parameter sigmoid function for the rate of occurrence of the events.

Animal	Time of step (s)	Ratio of rates after/before
B1	$2580 \pm 110$	36.4
B2	$2350 \pm 220$	30.3
C1	$2070 \pm 310$	5.49
C3	$1280 \pm 80$	1640
C7	$1680 \pm 280$	16.9

**Figure 5–figure supplement 2. Statistics of sudden changes in behavior.** Summary of the steps in the rate of long paths to water detected in 5 of the 10 rewarded animals. Mean and standard deviation of the step time are derived from maximum likelihood fits of a step model to the data.

A Fraction of time in modes			B Transition probability between modes: rewarded animals			
Mode	rewarded	unrewarded	from / to:	leave	drink	explore
leave	$0.053 \pm 0.014$	$0.054 \pm 0.013$	leave		$0.51 \pm 0.14$	$0.49 \pm 0.14$
drink	$0.103 \pm 0.026$		drink	$0.10 \pm 0.05$		$0.90 \pm 0.05$
explore	$0.844 \pm 0.032$	$0.946 \pm 0.013$	explore	$0.40 \pm 0.11$	$0.60 \pm 0.11$	

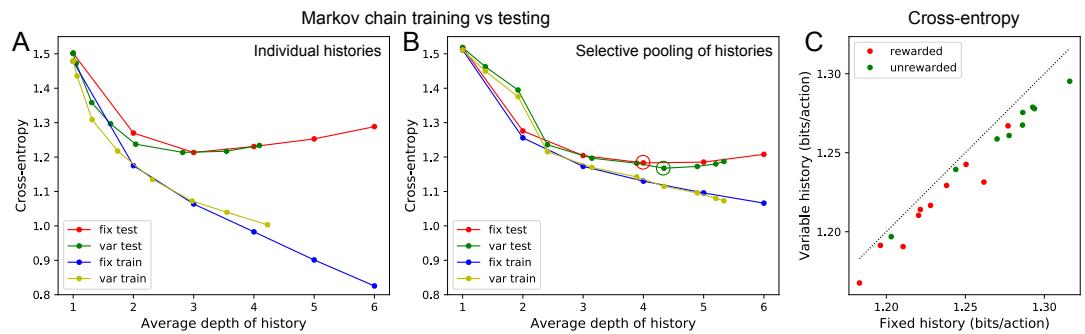
**Figure 7–figure supplement 1. Three modes of behavior.** (A) The fraction of time mice spent in each of the three modes while in the maze. Mean  $\pm$  SD for 10 rewarded and 9 unrewarded animals. (B) Probability of transitioning from the mode on the left to the mode at the top. Transitions from ‘leave’ represent what the animal does at the start of the next bout into the maze.



**Figure 8–figure supplement 1. Functional fits to measure exploration efficiency** (A) Fitting Equation 12 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  $\pm$  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 11D) and the optimal Markov chain (Figure 11C).

	<b>Bias</b>	<b>rewarded</b>	<b>unrewarded</b>
	$P_{SF}$	$0.77 \pm 0.03$	$0.78 \pm 0.02$
	$P_{SA}$	$0.72 \pm 0.02$	$0.71 \pm 0.02$
1095	$P_{BF}$	$0.82 \pm 0.03$	$0.81 \pm 0.03$
	$P_{BS}$	$0.64 \pm 0.02$	$0.63 \pm 0.02$

**Figure 9–figure supplement 1. Statistics of the four turning biases.** Mean and standard deviation of the 4 biases of Figure 9A-B across animals in the rewarded and unrewarded groups.



<sup>1096</sup> **Figure 11–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.