Innate Navigation:
Magnetic Sensation and Maze Learning

Thesis by
Matthew Rosenberg

In Partial Fulfillment of the Requirements for the Degree of
Computation and Neural Systems

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2022
Defended May 20, 2022

ACKNOWLEDGEMENTS

I would like to thank my parents, friends, and colleagues for their enthusiasm, patience, humor, support, instruction, and wisdom. There is no school, lab, or community in which I would have rather spent this formative period of my life.

ABSTRACT

This thesis aims to advance the understanding of the neurobiology of navigation through the investigation of two topics: magnetic sensation and maze navigation. The central question of this work may be framed as follows: how do animals find their way to key resources that are necessary for survival? Three projects are presented to address this.

Chapter II explores a sensory hypothesis that some animals may navigate long distances by directly sensing the earth's magnetic field. Awake zebra finches were stimulated with magnetic fields that varied sinusoidally in time while electrical recordings were collected via multi-channel electrodes. Preliminary negative results are presented, along with a detailed statistical treatment indicating no significant effect of magnetic stimulation on neural activity.

Chapter III presents a novel approach to studying learning and navigation in animal subjects. Mice are allowed free passage between a normal home cage and a complex maze environment, coming and going as they please. Sated animals, with free access to food and water, spend significant portions of a given multi-hour experiment in the maze and display efficient exploration. Water-restricted animals show three additional phenomena: immediate knowledge of the route home, rapid learning of the location of a single water port among 64 similar locations, and a moment of "sudden insight" in which the rate at which long, direct routes to the water source, beginning from many locations, increases discontinuously.

Chapter IV offers a simple, biologically feasible circuit model that recapitulates and explains some of the rapid learning behaviors we observe in mice. This model suggests a mechanism that might allow mice to flexibly store and recall direct routes to different resources that are activated by different internal drives.

The final chapter outlines some potential directions for future inquiry, including potential maze experiments to conduct with wireless electrophysiology and expansion of the range of species tested for magnetic perception. The Appendix briefly describes some follow-up experiments and intriguing preliminary results. Similarities in the navigation deficit displayed by mice that have been experimentally perturbed in several disparate ways is noted briefly. These perturbations include whisker trimming, olfactory neuron ablation, genetic ablation of cortex and hippocampus, and opiate intoxication.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Matthew Rosenberg*, Tony Zhang*, Pietro Perona, and Markus Meister. Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration. *Elife*, 10:e66175, 2021. doi: 10.1093/protein/gzv057.
M.R. participated in conceptualization, data curation, software, formal analysis, validation, investigation, visualization, methodology, and writing of the manuscript (* - authors contributed equally).

[2] Daniel A. Wagenaar, Matthew H. Rosenberg, and Markus Meister. Quantifying stimulus-induced periodic modulation in non-poisson spike trains. *Unpublished technical report*, 2022.
M.R. participated in data collection and curation, software, formal analysis, validation, investigation, visualization, methodology, and writing (review and editing) of the manuscript.

[3] Tony Zhang, Matthew Rosenberg, Pietro Perona, and Markus Meister. Endotaxis: A universal algorithm for mapping, goal-learning, and navigation. *bioRxiv*, 2021. doi: 10.1101/2021.09.24.461751.
M.R. contributed to the conception of the study and revision of the manuscript.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Navigation in the expanse

1.1.1 Sensing magnetic fields

Migratory animals like birds, turtles, and butterflies have been known to travel enormous distances to reach warmer climates or breeding grounds [75, 131, 197]. In many cases, these journeys contain long segments where the local features of the environment do not provide obvious or unambiguous information to the animal about its position relative to its desired destination. In extreme cases like swimming or flying across large bodies of water or otherwise featureless terrain, the local environment seems to present no useful guidance information whatsoever. The fact that animals seem to be able to traverse these regions regardless has prompted the idea that perhaps migrating animals like these possess an internal compass that senses the Earth's magnetic field. I begin with a brief survey of behavioral work that has endeavored, somewhat unsatisfactorily, to demonstrate magnetic sensing abilities in experimental subjects, followed by enumeration of the main hypotheses for how such a compass might work.

Behavioral evidence for an allocentric compass

Broadly speaking, behavioral assays of magnetic sensation have either directly quantified animals navigating or have employed indirect measures, such as spontaneous orienting movements. One classic experiment involved transporting pigeons more than 15 kilometers away from a home location, releasing them with blurred goggles to obscure their vision, and recording their flight paths and ultimate destinations. Though the pigeons generally flew in the correct direction, the reported results leave some room for doubt. Frequently sunny conditions and incomplete visual obstruction by the goggles may have permitted some use of vision for guidance. Furthermore, angular and radial errors in the final locations of the animals relative to the home location, combined with the exclusion of subjects that did not fly sufficiently far from the release point, decrease the certainty of the finding [214].

Indirect measures of magnetic sensation generally attempt to exploit the supposed tendency of migrating animals to spontaneously orient themselves toward a specific compass direction, even if the experimental setup does not permit actual migration.

An example of such a study measured the mean orientation of scratch marks left by European robins on paper at the bottom of their cages during migratory months. The orientation of the scratch marks was compared between conditions with and without the presence of an artificial high frequency magnetic field. The authors show that the applied magnetic field disoriented the birds [243]. Another study employed a similar technique to sea turtles swimming tethered in a tank, reporting that turtles maintained a consistent heading within the tank except when both vision and magnetic fields were disrupted [12].

Proposed mechanisms of magnetic sensation

A different line of research circumvents the difficulty of measuring magnetic effects on behavior by instead investigating possible mechanistic origins to such a sense. These efforts are often focused on the earliest sensory transduction steps including: radical pair chemistry in the eye[109, 165, 201], magnetic particles in the beak [76] or inner ear [274], and electromagnetic induction in electro-sensitive fish and sharks [122, 123]. Others have looked downstream from peripheral sensors and have reported magnetically induced changes in the brains of animals, including humans [64, 258, 274].

We investigate the possibility that zebra finch brain activity might be modulated by magnetic stimulation in Chapter II. Our approach differs from others in that we deliver a stimulus at a specific frequency and search for an elevated Fourier component in the power spectra of recorded neurons' firing rates. We report a preliminary negative result, i.e. a lack of magnetic effect on neural activity, and provide a detailed experimental and analytical approach to continue the search in other species, including both external collaborators in other labs as well as continued efforts within our lab. The goal of the endeavor is to engage the larger research community in an effort to more repeatably and systematically search for neural evidence of magneto sensation. Future work will engage with collaborating labs and survey a broader range of species.

1.1.2 Maps, memories, and manifolds

This section introduces several interrelated bodies of work that inspire and inform the work presented in the chapters below. Direct and systematic engagement with these ideas via experimentation and modeling is left for future work. However, it would be remiss to neglect a cursory review of these topics here given their immense influence on my thinking and on the community at large. A somewhat vague yet highly influential notion of a “cognitive map” is introduced. This is followed by a discussion of experimental and theoretical work devoted to an especially prominent exemplar: path integration.

Cognitive maps

As will be reviewed in greater detail below, prevailing ideas in psychology in the earlier part of the twentieth century regarded animals’ cognitive capacities as relatively rigid, almost reflexive responses to preceding stimuli. Ivan Pavlov had restricted his investigations to learning phenomena in which the stimuli presented to the animals was not contingent on their taking an action [188]. As psychology expanded to the consideration of the role of actions in creating positive or negative outcomes, Edward Thorndike observed that animals that were shown or instructed how to escape from a “puzzle box” did not escape faster than animals that had had no instruction and merely attempted random actions until success. From observations of this kind he proposed the “Law of Effect,” which states that animals essentially try actions at random. Actions that are quickly followed by reward are “reinforced” (the role of punishments was eventually determined to be much less effective in modifying behavior and a “Law of Exercise” was excised from Thorndike’s theory of “Instrumental conditioning”) [246].

Against a strong trend toward reductionism in his contemporaries, Edward Tolman invoked the notion of a “cognitive map,” endowing animals with additional mental components, including “biases” and “hypotheses” [249]. His ideas were inspired by a few observations. For instance, he coined the term “latent learning” to refer to a phenomena in which *unrewarded* pre-exposure to a given maze accelerated subsequent maze learning when seeking rewards, relative to a control group that was not pre-exposed. Additional maze experiments sought evidence for the existence of “insight” or “inference,” based on data in which animals appeared to prefer a novel shortcut prior to any direct experience of that path. He also helped to bring attention to a side-to-side head turning behavior, termed vicarious trial and error (VTE), that is potentially correlated with task engagement or mental deliberation [187, 250]. Tolman’s suggestion that the inner workings of the animal brain might be more akin to a detailed map than a jumble of wires connecting stimuli and responses, as others contended, has remained highly influential [156, 208, 265, 268] despite significant controversy on both experimental and conceptual grounds, as well as inconsistent replication of the core results [22, 121, 222].

Path integration

Although Tolman's definition of the cognitive map was left somewhat ambiguous, the notion of "path integration" is more clearly defined and often considered a notable example of such a map. Here, the map is a spatial coordinate system, usually two dimensional (although see [225, 275]), into which the navigating agent and points of interest are embedded. Path integration refers to the process by which the agent sums the individual spatial vectors comprising its outbound path in order to estimate its current position relative to a reference position, often its starting location. This capacity is often assessed in situations where the agent takes an indirect route to a given destination but is subsequently asked to find a direct path between the two locations. In addition to mentioning informal experiments in blind subjects, Charles Darwin provided an anecdote of Siberian natives crossing a homogeneous expanse of hummocky ice, in which they maintained a consistent global direction of travel, despite their path being comprised of many short segments in which the heading varied considerably. He further speculated that this ability might likely be enabled by the "sense of muscular movement" [57]. Nearly a century later, John Barlow further refined the idea by proposing a mathematical description of the phenomenon and suggesting that proprioceptive vestibular organs provided the relevant sensory input to the neural computation [16]. Path integration has been reported in desert ants [170], hamsters [67, 69, 161], mice [93], and rats [6] (see [68] for further review).

The idea of a cognitive map, particularly the possibility that animals might construct and utilize some type of structured representation of their environment, motivates the behavioral experiments in mazes described in Chapter III. Chapter IV provides a biologically plausible model that builds such a map, specifically the adjacency graph of the environment. In the work presented below, path integration is relevant in that it forms an alternative hypothesis for how mice might learn to solve complex mazes. In other words, the animal might learn the specific structure of the maze environment or, instead, it might employ path integration alone. The likely answer is that it employs a mixture of path integration, irrespective of the environment, and use of a detailed representation of the environment that it gathers through direct experiences in the specific maze. However, the structure of the maze precludes the possibility that the animal *only* uses path integration to plan and execute direct routes from arbitrary starting locations to the reward location.

The Engram—a century of uncertainty

Eight years before his death, Karl Lashley published “In Search of the Engram” which summarized three decades of negative results spanning the majority of his career. He performed lesion experiments in rats and primates hoping to identify essential anatomical regions required for abilities like visual perception, skilled motor action, and memory for mazes. Yet, he ultimately concluded the following from a lifetime of inquiry:

> In experiments extending over the past 30 years I have been trying to trace conditioned reflex paths through the brain or to find the locus of specific memory traces. The results for different types of learning have been inconsistent and often mutually contradictory, in spite of confirmation by repeated tests. I shall summarize today a number of experimental findings. Perhaps they obscure rather than illuminate the nature of the engram, but they may serve at least to illustrate the complexity of the problem and to reveal the superficial nature of many of the physiological theories of memory that have been proposed. ... The engram of a new association, far from consisting of a single bond or neuron connection, is probably a reorganization of a vast system of associations involving the interrelations of hundreds of thousands or millions of neurons. [137].

He performed lesions in many regions with varying extents of damage, yet failed to identify a privileged region that abolished performance of the given task. Instead, he found that the performance of the animal degraded gradually in proportion to the amount of brain tissue removed and, in many cases, animals sufficiently recovered from even drastic lesions of entire regions to allow them to perform the task. From these observations, Lashley proposed two related principles: that the extent of the lesion he created determined the extent of the behavioral deficit he observed and that a given function was distributed throughout a given region such that the surviving portion could compensate for the portion that was lesioned. He named the former “mass action” and the latter “equipotentiality” [135, 136, 138]. Despite a proliferation of subsequent studies that have argued for a high degree of specialization in brain anatomy with respect to function, Lashley’s ideas remain highly relevant and have been replicated numerous times with modern techniques [6, 108, 126, 186].

These findings anticipate the result presented briefly in the appendix. I observe that mice that have been bred or lesioned to lack the vast majority of cortex (and hippocampus in the former case) initially perform very poorly in the maze, yet ultimately recover to nearly the same level of performance as the intact control animals. Despite major exceptions to the principle of mass action, our results and those of our predecessors advocate extreme caution in complacently accepting commonly held notions of indispensability and precise localization of behavioral functions to anatomical regions.

Remarkable exceptions

The results reviewed above and the experiments with acortical mice presented below might leave the reader with the erroneous impression that the brain is a homogenous soup of indistinguishable and redundant components. Yet, there is also a long history of medical observations of humans sustaining lesions of various brain areas. Famous reports include changes in personality following lesion of the frontal cortex [102] and language deficits following lesions to Broca's or Wernicke's cortical regions [24, 263].

In one prominent example, the tragic clinical case of patient H.M. provided a hint that perhaps Lashley's engram might exist to some extent after all. H.M. suffered from intractable epilepsy which was treated by bilateral removal of the hippocampi. This procedure improved H.M.'s seizures but caused a profound memory impairment that completely prevented the acquisition of new episodic memories [217]. That said, H.M. could remember his distant past and retained some capacity to learn new procedural skills, supporting the idea that memory might come in distinct dissociable types [160].

Medical cases like these, as well as the work of anatomists like Korbinian Brodmann [39] that delineated cortical regions based on cytoarchitectural analysis of Nissl stained brain slices, argue for a nuanced view in which brain regions may be somewhat specialized for certain functions after all. Though investigation of the neural underpinnings of the maze navigation behaviors reported in Chapter III is beyond the scope of this dissertation, an overview of the known neurophysiological substrates is warranted. The majority of the prior work reviewed below was obtained from experiments in which neural activity was recorded from rodents ambling around in open arenas, often rectangular boxes. The novel environment I use in the experiments presented in Chapter III quickly elicits interesting behaviors and provides a complementary angle from which to further investigate the neural substrates of navigation. Initial experiments may likely profit from assessing the degree to which known phenomena observed in open arenas, in which the cognitive task is generally trivial, generalize to animals executing skilled movements through a complex and structured environment.

A menagerie of cell types

Inspired in part from knowledge of patient H.M., neuroscience studies in the past few decades have repeatedly reported that selective damage to the hippocampus impairs certain memory tasks [6, 44, 63, 81, 168]. This has prompted greater scrutiny of the hippocampal circuit and closely interconnected medial entorhinal cortex (MEC). Electrical recordings from the hippocampus and medial entorhinal cortex have revealed a plethora of remarkable cell types [156, 210].

So-called "place cells" fire whenever the animal moves through a region within the cell's "place field," a specific region within the environment that elicits firing [182, 275]. Some place cells have been observed to be selective for specific locations near an environmental boundary [181] or to represent the location of a conspecific [56, 185]. Although place cells have been most frequently studied in the hippocampus, they have also been reported in the thalamus [118], dentate gyrus [255], and entorhinal cortex [238].

A couple of remarkable observations have been made about hippocampal place cells. First, these cells appear to be able to localize the animal *within* their place fields through a mechanism called "phase precession." This describes the observation that, during the burst of spikes that place cells emit as the animal moves through the place field, each spike occurs at an earlier point in the ongoing 7-12 Hz theta band of the surrounding local field potential, thereby using this phase difference to encode a more detailed estimate of its own location relative to the center of the place field [47, 119, 120, 147, 183]. Second, during periods of sleep and immobility, place cells reactivate in temporally compressed sequences that "replay" the order in which these place cells were activated when the animal actually moved through these place fields. These "replay" events occur during local field potential periods called "sharp wave ripples" [83, 125, 140, 146]. Remarkably, inhibition of these events impairs aspects of a spatial working memory task [116] yet replay events do not appear to represent the animal's upcoming choice [94]. Instead, it appears that place cells associated with alternate future actions alternate firing at 8 Hz, potentially involving theta processes, prior to selecting the action [127].

A host of exotic functional cell types have been reported in the hippocampus and connected regions. Within the hippocampus, David Tank's lab has described cells that fire specifically for a given region in the frequency space of an artificial tone [9] and cells that encode reward locations [89]. Cells encoding the passage of time have been found in the hippocampus and MEC [105, 132, 148]. Other cells, aptly named “head direction cells,” fire only when an animal faces a certain global heading and have been found in a number of regions including the retrosplenial cortex, postsubiculum, MEC, and thalamus [117, 238, 239].

Perhaps most intriguing of all are the “grid cells” found in the MEC. These cells fire in a periodic fashion along a 2D [100] (and occasionally 3D [276]) triangular lattice. Importantly, when environment barriers are removed or expanded, these cells extrapolate the original firing field lattice with a consistent periodicity in a translationally invariant manner [213]. This finding has prompted the use of a toroidal manifold [45, 88] as a computational description of the firing activities. Grid cells appear to be arranged in a modular fashion with a dorsal-ventral axis along which the spatial frequency of the grid pattern varies [232]. These cells are featured especially prominently in many theories of path integration [208]. Conversely, machine learning models trained to navigate or path integrate appear to display grid-like firing patterns in penultimate output layers [15, 55, 226].

Together, these cells and the circuits among them provide a useful lexicon when articulating models that attempt some measure of biological plausibility. In particular, within the Endotaxis circuit described in Chapter IV, the point cells might map to place cells in that they both fire for specific locations, while the map cells might correspond to the CA3 subregion of the hippocampus in that they both have recurrent connections [7, 113, 158].

1.2 Pavlovian, operant, and reinforcement learning

Navigation in expanses such as the sky, sea, desert, or a typical “open-field” experimental enclosure can largely be accomplished by heading in approximately the right direction alone. Cues that are accessible at all positions in the environment, like the Earth’s magnetic field [269], celestial lights [197], or the visible walls of a laboratory in a typical rodent neuroscience experiment, greatly simplify navigation. Likewise, path integration, though nontrivial given the accumulation of small errors over time in the absence of landmarks to correct the system [179], is often a viable strategy in these relatively simple environments.

Finding one’s way in constrained environments like a labyrinth poses additional challenges. For instance, in the experiments recounted in Chapter III, mice must navigate without obvious global cues, where local regions of the maze look identical to one another and even perfectly path-integrated heading vectors between the animal and destinations will frequently point into dead-ends. Moreover, navigation in mazes like these more closely resembles other forms of decision making, in which one's current state depends upon a precise *sequence* of discrete prior actions rather than the mere average of past actions. For these reasons, I review prior work on learning in general, before returning to navigation per se as a special case.

1.2.1 Behavioral observations and theories of learning

Careful scientific investigation of learning at the behavioral level began with the work of Ivan Pavlov. He described a collection of phenomena, known now as classical or Pavlovian conditioning, in which animals associate neutral stimuli with positive or negative events, transferring intrinsic responses to these outcomes to the stimuli that predict them. Pavlovian theories describe observations relating stimuli to behavioral responses. Specifically, unconditioned stimuli (US), like food or electric shock, evoke unconditioned responses (UR), such as salivation or freezing respectively. Through repeated pairings of initially neutral stimuli like lights and tones, animals learn to associate these conditioned stimuli (CS) with the US, which is revealed by the extent to which the UR is elicited by these CS. These experiments and ideas continue to provide a useful lexicon as well as a rich repository of behavioral observations [188, 199].

Though powerful and immensely influential, this theory largely neglected the role of (voluntary) actions in creating outcomes. Subsequent generations of psychologists extended Pavlov's ideas to include situations in which animals must take specific actions to obtain positive experiences and avoid negative ones. These ideas that emphasize the causal role of actions came to comprise a body of work referred to as “instrumental” or “operant” conditioning. Key figures in this movement include Edward Thorndike (reviewed above) and Burrhus Frederic Skinner.

Behavioral experiments in the mid-twentieth century became increasingly automated. The diligent observations and hand-taken notes of experimenters like Thorndike, Tolman, and Lashley were replaced with machines, often called “Skinner boxes” or “operant chambers,” that were capable of presenting repeatable configurations of stimuli and recording the responses of the animal subjects. These experiments generally present animals with artificial stimuli that are as neutral as possible, i.e. without any prior positive or negative associations, thus characterizing how the temporal arrangement and various combinations of stimuli result in varying degrees of vigor in responding, often measured in the rate at which rodents can be compelled to press a lever [74]. This line of research continues to be active today, where researchers seek to characterize the responses of animals to more elaborate combinations of stimuli [29–31, 73, 106, 198], including Pavlovian effects on instrumental learning [37, 101]. Often, these efforts are accompanied by attempts to assign conceptual elements of the theories to distinct brain regions and genetic cell types, notwithstanding the caveats to localization of function explained above [11, 85, 96, 101, 101, 150, 150, 151, 224, 245].

1.2.2 Reinforcement in models and brains

These studies of "operant" conditioning gradually became codified in models that explained the various changes in conditioned responses that accompany specific combinations and temporal orders of stimuli, based on the difference between the animal's expectation and its actual experience [257, 266]. Andrew Barto and Richard Sutton synthesized work drawn from animal behavior experiments and dynamic programming mathematics [20, 21] to form "reinforcement learning." This subfield of machine learning concerns itself with specifying computational models of agents that interact with an environment via actions, which in turn place the agent in new states, some of which may be associated with reward. Importantly, these rewards are sparse, appearing only periodically when the agent finds them. Thus, the agent must learn to act in a regime in which instructive signals potentially come from the environment sporadically and infrequently [236]. This approach now dominates both certain applications in machine learning, such as beating top human players at games [162, 163, 220, 221, 256], as well as models of animal behavior and neural representation [82, 92, 159, 229, 268].

In the late 1990s, Wolfram Schultz discovered an extraordinary correspondence between reinforcement learning theory, specifically the "temporal difference" learning rule [17, 234] and the responses of dopaminergic neurons in the ventral tegmental area (VTA). Often called the "reward prediction error," Shultz found that dopamine neuron firing rates increased proportionally to the difference between the reward experienced by the animal and the amount it expected. Thus, over time responses to the predictable rewards decreased and instead cells began to fire in response to stimuli that reliably predicted the reward [215, 216].

1.2.3 Extensions and complications

Dopamine cells project extensively throughout the brain making them a natural candidate for a global error signal that is broadcast widely to instruct learning in local circuits [259]. Mammalian brains contain many other neurotransmitters that similarly project widely throughout the brain. The remarkable correspondence of the VTA dopamine signal with the temporal difference rule has led theorists to propose that specific neurotransmitters or brain regions may mediate other terms in the reinforcement learning equations [61, 112].

Yet the data from experimentalists has tended to resist excessively reductive statements. For instance, serotonin neurons were proposed to work in an "opponent" fashion relative to the actions of dopamine, in which tonic and phasic firing rates each carry different signals and the two neurotransmitters have opposite signed responses to reward and punishment [33, 58, 61]. Yet, instead serotonergic neurons in the dorsal raphe nucleus (DRN) seem to display a wide range of responses to reward information, including both those that mimic dopamine neurons as well as others that differ with respect to time course, valence, and repeated presentations of the stimulus [52, 143, 153, 193]. More recent data has further complicated simplistic theories mapping regions and chemicals to terms in equations by demonstrating appreciable heterogeneity within the VTA and DRN, both with respect to the coding properties of neurons as well as the neurotransmitters released by these regions [145, 167]. In some cases the same neuron has been shown to co-release both dopamine and glutamate or GABA, with diverse and sometimes opposite effects in downstream regions [192, 203, 279]

The dominant anatomical theory of reward learning focuses on the dopaminergic VTA projection to the nucleus accumbens (NAc) [177, 270]. In particular, most theories of drug addiction propose that drugs act on presynaptic or postsynaptic sites in this VTA to NAc pathway [240, 241]. Though this idea remains dominant, it is challenged by experiments in which lesion to the NAc leads to only modest changes in the animal's behavior [49, 78, 212]. Moreover, cases where lesions actually increase food consumption of rewards, lead to increased rates of responding on levers, or fail to prevent drug seeking behaviors are especially hard to reconcile with an excessively simplistic view of VTA to NAc signaling, [13, 77, 200].

While reward signaling is arguably inextricably involved in any experimental task that evokes learning, the literature reviewed above is especially relevant to the work presented in Chapter IV and the Appendix. In Chapter IV dopamine signalling is one candidate anatomical substrate that could carry the "goal neuron" signal that guides the Hebbian plasticity that enables the endotaxis computation. In the Appendix, I briefly present data showing severe navigation deficits, resembling those of acortical animals, following opiate drug administration. One explanation of the failure of opiate intoxicated animals to navigate to water is that drug-induced enhancement of signaling in reward pathways prevents the animal from responding to the thirst signals that would normally motivate efficient navigation. Another hypothesis is that altered dopamine activity impairs the learning process more generally. This might occur either from a reduction in the rewarding properties associated with novelty, thereby decreasing efficient exploration of the environment, or by somehow impairing the neural plasticity events that comprise the animal's cognitive map of the maze.

1.3 Reductionistic vs naturalistic experiments

During the mid-twentieth century, behavioral studies became increasingly automated. BF Skinner developed methods to repeatably present stimuli and record elicited behaviors in what became known as "Skinner boxes" or "operant conditioning chambers." The relative ease of inspecting digitized records of behavior collected during experiments over earlier methods of careful observation, timing, and note-taking by experimenters accelerated the pace of behavioral experiments at the cost of drastically increasing reductionist experimental approaches and theories of behavior.

1.3.1 The Unbearable Slowness of Training

The series of decisions inspected by early researchers gradually became replaced by the predominance of "two alternative forced choice" (2AFC) paradigms in which animals typically provide a binary response during a response period that follows presentation of sensory stimuli. Animals are laboriously trained to attend to the stimuli to yield this 1-bit decision. While earlier researchers tended to plot "learning curves" showing the acquisition of a given task [74, 246], the growing predominance of simplified behavioral tasks tended to shift the focus toward steady-state performance of a task, i.e. after the learning process is complete, emphasizing percent of trials answered correctly and concomitant correlations between correct and incorrect trials with underlying brain states. The goal was to control stimuli as precisely as possible both between separate trials within a given subject as well as with respect to the experiences of the subjects as well. In recent decades, new genetic tools in mice prompted the adoption of behavioral tasks developed in monkey experiments to rodents in the hope of generating new discoveries [62]. Interesting findings emerged, but the highly artificial scenarios in which experimental animals were tested led to the unfortunate side effect that training experimental subjects often took weeks or months of experimenter labor [41, 44, 46, 95, 151, 206, 224, 245]. This naturally curtails the throughput of these behavioral assays.

1.3.2 Observing natural behaviors in a laboratory setting

The experiments presented in Chapter III adopt a different approach in which experiments aim to elicit natural behaviors without employing laborious shaping procedures to train animals to do unnatural things. Inspired by early "ethologists" such as Tinbergen who studied predator avoidance and imprinting behaviors in birds [247], "naturalistic" or "ethologically-relevant" studies employ experimental paradigms that elicit rich, complex behaviors naturally without extended training. Such behaviors include courtship and mating [124, 144], aggression [144], prey capture [110, 157], predator avoidance [70, 254, 277], and navigation among others [6, 69]. Many of these are arguably innate, i.e. they do not require learning, but navigation is among a privileged class of behaviors that are simultaneously trivial to elicit experimentally while permitting an opportunity for experimental subjects to demonstrate learning of numerous complex multi-action sequences.

The mouse navigation experiments described in Chapter III provide a compromise between carefully controlled experiments employing arbitrary, artificial stimuli and observations of animals in nature in which complex behaviors unfold, but myriad uncontrolled (and often unmeasured) factors preclude clear interpretations of the data. Learning in maze environments occurs in a tiny fraction of the time it takes to train animals on "easier" tasks and assesses a capacity that is of undeniable importance to the animal: locating water.

1.4 Modeling the neural implementation of rapid learning

In an ideal world, scientific progress results from a close interaction of experimental and theoretical work. Data from experiments constrains the parameters of models, while models generate predictions that experiments can falsify. This process both refines theories as well as guides experimentalists in deciding which experiment should be conducted next.

Yet, within the field of neuroscience, all too often experimental and theoretical efforts exist in separate spheres. Experiments often fail to be motivated by relevant theory or present data without clearly stating how these observations should be interpreted relative to a model of the same phenomena. Likewise, theoretical work often involves model components that are not directly observable or controllable in experiments. In some cases, technological development may someday allow inspection of quantities that are currently unobservable. In other cases, like assessing the memory capacity of a neural network [8, 50, 264], analogous experiments appear to be strictly impossible. For instance, experiments, experimenters, and animal subjects have finite life spans, thus it appears unlikely that experiments will ever be able to assess the *maximum* number (assumed to be massive) of memories that a biological brain can store.

Chapters III and IV attempt a closer relationship between experimental data and theoretical model. The rapid learning of resource locations, based on a small number of experiences, is selected as the phenomenon of interest. A simple circuit model is proposed to explain this learning phenomenon. Unlike many of the current models of learning and navigation, Chapter IV presents a model that is not expressed in the language of reinforcement learning and does not require the backpropagation of error technique.

1.4.1 Constraints imposed on models by experimental data

The navigation data presented in Chapter III demonstrates that mice can learn to navigate to a reward location after only approximately 10 experiences with the reward. Though far from fully precluding traditional reinforcement learning models [194], the speed at which this spatial knowledge is acquired, along with the flexible manner in which mice navigate directly to reward from many locations in the maze, suggest that the animals may be constructing a cognitive map in which the structure of the environment and resource information are more decoupled than in most reinforcement learning accounts [59, 82, 91, 166, 207, 228, 229]. For this reason, the model in Chapter IV is expressed without explicit reference to existing treatments of the problem in reinforcement learning.

1.4.2 Neural plasticity

The model discussed in Chapter IV is constructed from simple, canonical pieces. In brief, the biologically feasible circuit involves a recurrent network of “map” cells that stores the structure of the environment via Hebbian plasticity, using Oja’s rule to avoid weight explosion. Hebbian plasticity refers to an immensely influential scheme, proposed by Donald Hebb in 1949, by which neurons might change the strength of connections between one another. Hebb proposed that presynaptic and postsynaptic neurons that fired in close temporal succession would yield a strengthened connection between the two [103]. Oja’s rule is simply one popular way to prevent the gradual explosion in the strength of connections between neurons. It does so by ceasing to update the connections when these weights and the postsynaptic activity can be used to predict the presynaptic activity [180]. Sensory inputs are passed to the model in a feedforward manner and the output of the map network is in turn sent to the goal neurons via feedforward synapses. Goal cell plasticity is endowed with a third factor [114] (beyond the two implied by generic Hebbian learning), perhaps corresponding to a dopaminergic signal, in order to combine information about resource availability and the spatial structure of the environment. Although experimental work on plasticity continues to be conducted, biological substrates for Hebbian plasticity are already widely acknowledged [26, 32]. Chapter IV elaborates on the motivation, structure, and performance of the Endotaxis circuit sketched here.

Chapter 2

GLOBAL SEARCH FOR GLOBAL SENSATION

Quantifying stimulus-induced periodic modulation in non-Poisson spike trains

Daniel A. Wagenaar, Matthew H. Rosenberg, and Markus Meister

Technical Report, May 16, 2022

2.1 Introduction

We are engaged in a large-scale collaboration to find evidence of magnetoreception in the nervous systems of multiple species. Each lab in the collaboration presents periodic magnetic stimuli while recording from whichever area of the nervous system they study in the course of their other projects. The stimulus in all cases consists of a magnetic field that oscillates, typically at a few Hz, with a strength comparable to or larger than the earth magnetic field. The lab then processes the raw data by their own standard methods, and shares the resulting recordings with the collaboration.

Here we focus on the case where the recordings are spike trains. The central question becomes: Do those spike trains exhibit any modulation induced by the stimulus? Detecting modulation in continuous time series is a well-explored subject. Detecting modulation in a point process (such as a spike train) is somewhat less common. Here, we will describe one method that can detect stimulus-induced modulations, and also yields an upper bound on the magnitude of modulations that would remain undetected. When applied to 5 data sets of recordings from zebra finch brain, the method revealed no significant modulations of firing rate at the magnetic field frequency.

2.2 The data

We analyzed spike trains recorded with silicon probes from a 1.3 year old female zebra finch. The experimental pipeline comprised three phases: surgical preparation, acclimation to restraint, and awake electrophysiology. All procedures followed animal welfare guidelines under a protocol reviewed and approved by the Caltech IACUC.

2.2.1 Craniotomies and headpost implantation for neural recordings

The animal was anesthetized with isoflurane, a small head post was cemented to the skull (C&B metabond), and two small craniotomies were opened—one for insertion of the electrode anterior to the midsagittal bifurcation and another over the cerebellum for the ground/reference wire (~1-2.5 and 1 mm diameter respectively). The craniotomies and exposed skull were then covered with bio-compatible polymer (Kwik-sil). All surgical procedures were conducted under sterile conditions and with standard postoperative pain management.

2.2.2 Acclimation to restraint

The animal was acclimated to restraint and head fixation 5–14 days after the surgical procedure. The animal experienced three sessions of increasing duration (30 minutes, 1 hour, and 2 hours) spread across three days. On each day, the animal was restrained in a plastic (Falcon) tube to prevent injury to the animal or probe damage during the recording session resulting from wing flapping. The head was immobilized by clamping the headpost to a fixed arm.

2.2.3 Awake electrophysiology

Neural recordings were collected from a Neuropixels probe system (Janelia, version 3B1) with a National Instruments USB-6221 DAQ and SpikeGLX software.

Prior to the recording session, the probe was coated with DiI to facilitate reconstruction of the probe path in subsequent histology. The animal was restrained in the plastic tube and head-fixed as during the acclimation phase. The craniotomies were exposed, the ground/reference wire was implanted in the cerebellum and secured with Kwik-sil, and the probe was inserted with a micromanipulator. A drop of mineral oil was placed over the insertion site to keep it moist during the recording. After waiting for the signals to settle, experimental recordings began. Each recording had a duration of approximately 5 minutes in which different types of stimuli were presented to the bird.

A magnetic field was generated using custom solenoids, created by Haixiang Xu, driven by a function generator with sinusoidal voltage at 3 Hz or 5 Hz frequency. Two solenoids were placed on either side of the animal under the beak with the axis oriented toward the head, approximately 45 degrees down from the sagittal plane. The magnetic field amplitude at the location of the bird head exceeded the strength of the Earth's magnetic field.

Seven recordings were collected under continuous sinusoidal magnetic stimulation. One of these was excluded due to an artifact in the magnetic stimulus; another was excluded due to excessive noise across all channels. Two of the five recordings were conducted with 3 Hz stimulation, the other three with 5 Hz stimulation. Seven additional recordings were collected but not analyzed here (steady-state, settling, and metronome stimulus).

2.2.4 Spike sorting

Raw recordings from the Neuropixels probes were referenced to the average across all electrodes. Spike sorting was performed with Kilosort2. Units with excessive noise or abnormal spike waveforms were excluded. Ambiguous cases were included in the analysis to diminish the chances of excluding magneto-sensitive cells during spike sorting.

2.3 Mathematical basics

2.3.1 Fourier series

Any (reasonable) real- or complex-valued function $g(t)$ defined on the finite interval $[0, T]$ can be written as a linear superposition of complex exponentials:

$$g(t) = \sum_{n=-\infty}^{+\infty} \mathrm{e}^{2\pi i n t/T} c_n \,.$$

Here $\mathrm{e}^{2\pi i n t/T}$ is the complex exponential with frequency $f = n/T$, c_n denotes the weight of the contribution from that exponential, and the sum combines contributions from all different frequencies. The numbers c_n are known as the *Fourier coefficients* of $g(t)$ and are given by:

$$c_n = \frac{1}{T} \int_0^T \mathrm{e}^{-2\pi i n t/T} g(t) \mathrm{d}t.$$

2.3.2 Fourier coefficients of a point process

Fourier analysis as described above is usually applied to time series data, i.e., data that are described by a function of (continuous or discretized) time. However, the same basic analysis can be applied to point processes, i.e., data that consist of a sequence of timestamps, such as spike trains. That is because a spike train comprising

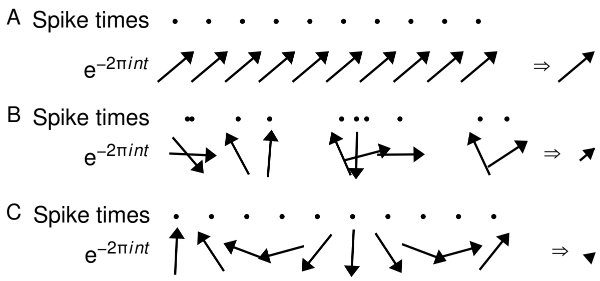


Figure 2.1: **Fourier analysis of point processes.** (A) Fourier analysis of a periodic spike train at the frequency of that spike train. Arrows represent the complex exponential (phasor) connected to each spike time. All the phasors are aligned and produce a large sum vector. (B) Fourier analysis of a train of irregularly timed spikes. Here the phasors point in random directions. (C) Fourier analysis of a periodic spike train at a different frequency. Here the phasors are systematically arranged at different phases, and thus cancel destructively.

N spikes at times t_k (where $k = 1 \dots N$) can be written as a function:

$$g(t) = \sum_{k=1}^{N} \delta(t - t_k),$$

where $\delta(x)$ is the Dirac delta function (whose value is zero every except at $x = 0$ and whose integral is one). Thus, the Fourier coefficients of a spike train $\{t_k\}$ are:

$$c_n = \frac{1}{T} \sum_k \mathrm{e}^{-2\pi i n t_k / T}. \tag{2.1}$$

To gain an intuition of what this means, consider two alternative spike trains comprising 10 spikes in a 10-s recording period: In the first train spikes happen at perfectly regular intervals (Fig. 2.1A); in the second they occur at irregular intervals (Fig. 2.1B). When analyzed at the frequency of the spikes, the former will yield a large (absolute) value for c_n, because the phasors $\mathrm{e}^{-2\pi i n t_k / T}$ align constructively, producing a large sum in the complex plane. The latter yields a much smaller value for c_n, since the phasor $\mathrm{e}^{-2\pi i n t_k / T}$ will be different for each of the spikes, so the vector sum is smaller. Finally, if a periodic spike train is analyzed at a frequency other than the firing rate (or its multiples), the phasors sum destructively and the resulting Fourier coefficient is close to zero (Fig. 2.1C).

2.4 Detecting modulation in spike trains

For simplicity, we choose our recording interval T such that it is an exact multiple of the period of the stimulus, i.e., such that $f_s T$ is integer, where f_s is the frequency of the stimulus. In that scenario, the Fourier coefficient at the frequency of the stimulus is $c_\mathrm{s} := c_{n=f_\mathrm{s}T}$. If this coefficient is significantly elevated in the presence of a stimulus, one may conclude that the spike train is modulated by that stimulus.

The obvious question is: What constitutes significant elevation? It helps to consider some toy models for the statistics of the spike train and evaluate the predicted Fourier coefficients.

2.4.1 Warm-up exercise: assuming Poisson statistics

If the spike train is a homogeneous Poisson process, then spikes happen independently of each other and with constant probability per unit time. This implies that the phasors $\mathrm{e}^{-2\pi i n t_k/T}$ (Fig. 2.1B) have uniformly distributed random phases in the complex plane. Thus the Fourier coefficient c_n of Eqn. 2.1 is $1/T$ times the sum of N random unit arrows in the complex plane, whose phase angles are independently and uniformly distributed. This is true at all frequencies n, so we will drop the index on c for notational convenience:

$$c = \frac{1}{T}\sum_{k=1}^{N} s_k, \tag{2.2}$$

where s_k is a random phasor

$$s_k = \mathrm{e}^{-2\pi i \phi_k},$$

with ϕ_k uniformly distributed in $[0, 2\pi)$. One can show that the absolute value $|c|$ is distributed as

$$P(|c|) = \frac{2|c|T}{R}\mathrm{e}^{-|c|^2 T/R}, \tag{2.3}$$

where $R = N/T$ is the firing rate of the Poisson process. (See the Appendix for a derivation of this result.)

Of course one can easily simulate the Poisson process numerically. For Figure 2.2 we simulated 10,000 examples of spike trains, each spanning 100 s at a given firing rate. We show the distribution of all Fourier amplitudes from these spike trains. As can be seen, it nicely matches the analytical result. This probability distribution represents a prediction from the null hypothesis that the spike train is Poisson at constant rate with no modulation from the magnetic field or anything else. For an experimental spike train, one can now compare the measured Fourier coefficients to this probability distribution and derive a likelihood for the null model.

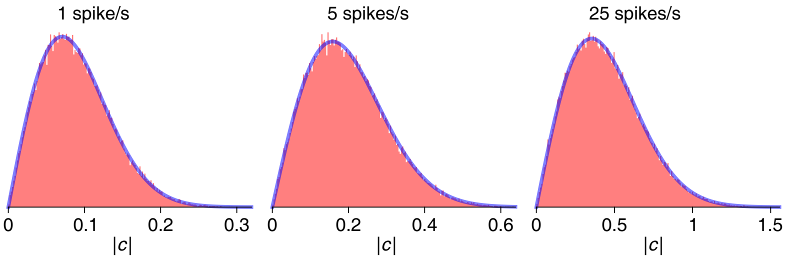


Figure 2.2: **Fourier coefficients from artificial spike trains that obey Poisson statistics.** Constructed from 10,000 instances of 100-s long artificial "recordings." Pink: sampled distribution. Blue: prediction from Eqn 2.3.

2.4.2 The Fourier components of actual spike trains

Unfortunately, real spike trains rarely obey Poisson statistics. One well-known deviation is the refractory period, a brief obligatory silent interval following each spike. Many spike trains are also "bursty," in the sense that spikes tend to come in clusters that are separated by longer intervals. So spikes "repel each other" at short intervals of 1–2 ms but often attract each other at longer intervals. Both are obvious discrepancies from the Poisson model of independent spikes. As a result, a comparison to predictions from Poisson behavior leads to large numbers of false positive detections of modulation (Fig. 2.3).

Our task here is well-defined: to detect modulation at one specific frequency, namely that of the oscillating magnetic field. Figure 2.4 suggests a simple approach: Within some range of the modulation frequency (say 4.0 ± 0.3 Hz) the power spectra look reasonably flat. It appears that the Fourier coefficients in this range are drawn from approximately the same distribution. So we can compare the Fourier coefficient at the stimulus frequency with a probability distribution constructed from all the other coefficients in that limited range.

Figure 2.5 illustrates the results. For a given spike train, the Fourier coefficients within the range of interest (4.0 ± 0.3 Hz) scatter with a distribution that appears independent of frequency (Fig. 2.5A) and that follows a Gaussian shape (Fig. 2.5B). Comparing across cells, the variance of that distribution increases with the firing rate (Fig. 2.5C), as expected because more phasors contribute to the coefficient (Fig. 2.1). However the mean of the distribution, the skewness, and the kurtosis all remain near zero (Fig. 2.5D–F). All this suggests that the Fourier coefficients follow a Gaussian distribution to good approximation.

These observations can be used for a revised test of the null hypothesis: If there is no modulation by the stimulus, then the Fourier coefficient at the stimulus frequency, c_s,

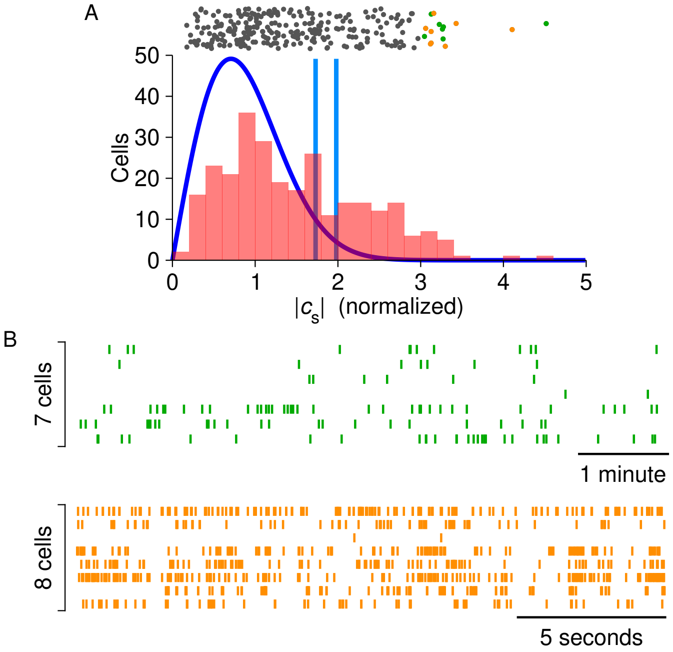


Figure 2.3: **False positive detections resulting from incorrect assumption of Poisson statistics.** (A) Histogram of Fourier coefficients at $f = 4$ Hz for 275 cells recorded for 400 s in the absence of a stimulus at the analysis frequency (*pink*) with theoretical curve assuming Poisson statistics (*blue*). The 95th and 98th percentiles of the theoretical distribution are indicated by light blue lines. Many of the measured coefficients lie beyond the 98th percentile under the Poisson assumption and would be flagged as positive results. *Dots*: Data for individual cells. Colored dots correspond to examples in *B*. (B) Spike trains from cells with the highest Fourier coefficients: 7 cells with relatively low firing rates (*top*) and 8 cells with relatively high firing rates (*bottom*). Note, the pronounced temporal clusters of spikes.

should be drawn from the same Gaussian distribution as the coefficients at nearby frequencies $c_{n \neq f_s T}$. For a useful statistic we define the *normalized response* as the absolute value of the Fourier coefficient at the stimulus frequency normalized by the standard deviation of the Fourier coefficients at non-stimulus frequencies:

$$\hat{c} = \frac{|c_s|}{\sigma_c}. \tag{2.4}$$

where

$$\sigma_c = \sqrt{\left\langle |c_n|^2 \right\rangle_{n \neq f_s T}}. \tag{2.5}$$

Under the null hypothesis, c_s should be distributed just like the other c_n, which implies that $\hat{c}$ should be distributed as

$$P(\hat{c}) = \hat{c}\, \mathrm{e}^{-\frac{1}{2}\hat{c}^2} \tag{2.6}$$

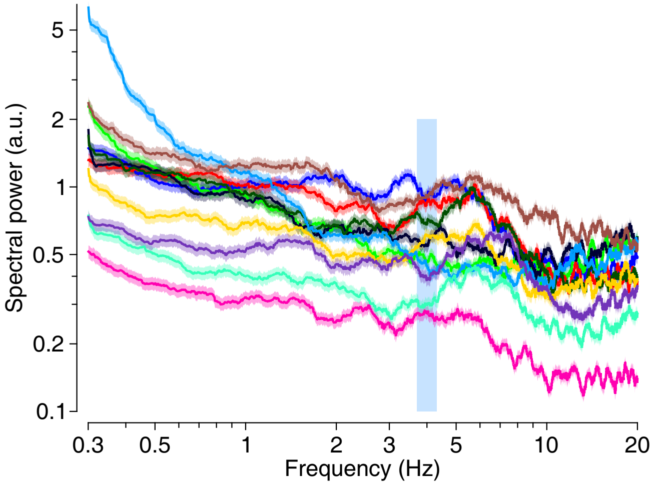


Figure 2.4: **Multitaper power spectra of several spike trains.**

(see Appendix). The actually observed values of $\hat{c}$ can then be compared against this null model, and much as in Figure 2.3, we can test whether any neurons produce unexpectedly large values of $\hat{c}$. Figure 2.6 shows that this is not the case. About 5% of the neurons fall above the 95th percentile of the null distribution, and no unusually large values occur. In other words, these data appear consistent with the null hypothesis of "no modulation."

2.4.3 Confidence limits on modulation

Hypothesis testing for the presence of modulation is not the only use for the null model described in the previous section. The model can also be used to generate confidence limits on undetected modulation.

To explore this, we start from actually recorded spike trains $\{t_k\}$ that have no stimulus-induced modulation, and introduce artificial modulation at a defined frequency f_s. This is done by shifting some of the spikes by half a stimulus period. The probability that a spike at time t_k will get shifted varies sinusoidally in time

$$p(t_k) = \frac{A}{2}\left[1 + \cos(2\pi f_s t_k)\right],$$

where A represents modulation amplitude and f_s is the frequency of the modulation. The resulting spike train has the same mean firing rate, but its rate is modulated at the frequency f_s with relative amplitude (peak/mean − 1) given by A.

We took 275 actual spike trains with firing rates ranging between 0.5 and 91 spikes/s, each of 400-s duration, and created artificially modulated copies of each, at modulation amplitudes ranging from $A = 0.01$ (1%) to 0.5 (50%). For each spike train and

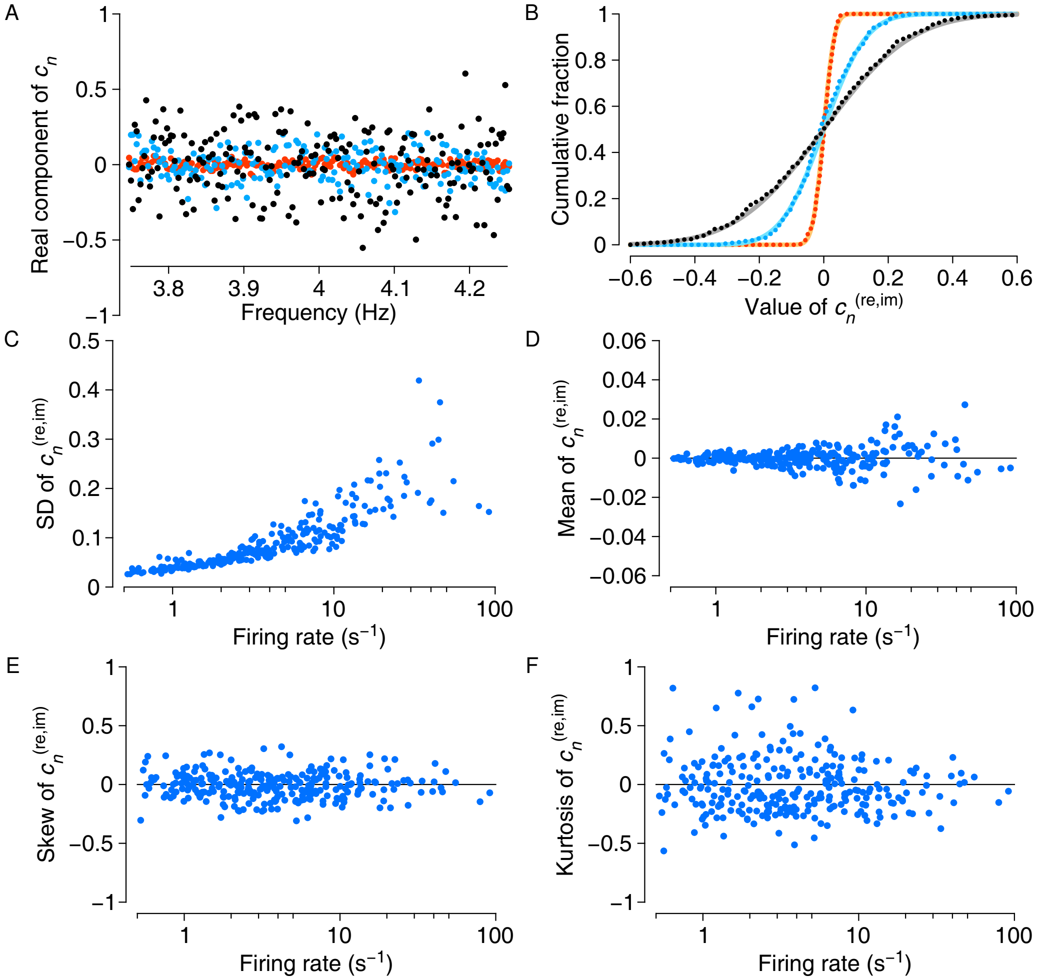


Figure 2.5: **Fourier coefficients follow a Gaussian distribution to good approximation.** (A) Real components of the Fourier coefficients as a function of frequency from three actual spike trains, with firing rates of 0.5 s^{-1} (*red*), 10 s^{-1} (*blue*), and 55 s^{-1} (*black*). Recording time: 400 s. (B) Cumulative distribution of the data shown in *A* (dotted lines) compared with best-fit zero-mean Gaussian. (C) Standard deviation of Fourier components for each of 275 recorded cells plotted against their firing rates. (D)–(F) Means, skew, and kurtosis of Fourier components, ditto.

modulation amplitude, we calculated 1000 instances of artificially modulated spike trains. For each cell, we compared the actual $\hat{c}$-value of the original spike train (Eqn 2.4) with the 1000 artificial $\hat{c}$-values in each group. We found the maximum level of modulation where the actual $\hat{c}$ would be in the lowest 95% of the distribution, and plotted the result in Figure 2.7A. These data represent our 95% confidence limits on the modulation in the actual spike trains.

Unsurprisingly, these confidence limits depend rather steeply on the firing rate of the cell. However, we noted that these confidence limits also varied substantially

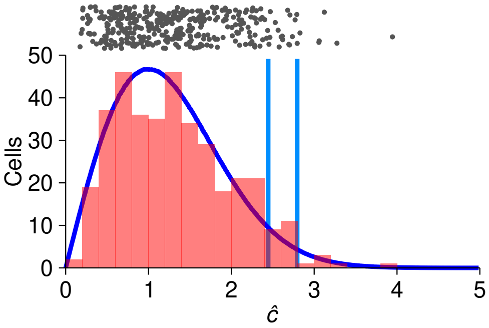


Figure 2.6: **A test for modulation using the improved null model.** Histogram (*pink*) of the normalized responses $\hat{c}$ from all neurons (*dots*) along with the predicted null distribution (*blue*). Vertical lines indicate 95th and 98th percentiles of the null distribution.

among cells with very similar firing rates. To test whether this was an artifact of our sampling procedure, we grouped the 1000 instances of artificially modulated spike trains into 10 groups of 100 (for each cell and at each modulation level), and repeated the above calculation within each group. The population standard deviation among groups is plotted for each cell as error bars in Figure 2.7. The fact that these error bars are small compared to the overall spread in the data indicates that differences in detection limits are due to inherent differences in firing statistics between cells rather than to our sampling procedure.

Another way to think about these results is to ask how small a modulation would be detectable with our method in spike trains with statistics like our recordings. To explore this, we again used the artificially modulated spike trains and applied the test of Figure 2.6 at various confidence levels. We asked what level of modulation would be detected by that method at least 50% of the time, and called that the “minimum detectable modulation” (Fig 2.7B). We may conclude that a cell with a firing rate of 2 spikes/s would need a modulation of 0.25 to be detectable with confidence $p < 10^{-4}$, whereas in a cell firing 50 spikes/s a modulation of only 0.05 would be detectable at the same confidence level. Unsurprisingly, the more relaxed the confidence level, the smaller modulations are detectable, but the effect of going from $p < 10^{-4}$ (appropriate when analyzing up to 500 spike trains in parallel) to $p < 0.01$ (appropriate for up to 5 spike trains) is not as large as one might have imagined.

2.4.4 Results from magnetic stimulation

Having shown that our method produces a null model that accurately describes the observations at frequencies where no stimulus-induced modulation was expected, we now turn to analyzing the data at the actual stimulus frequency. Using two

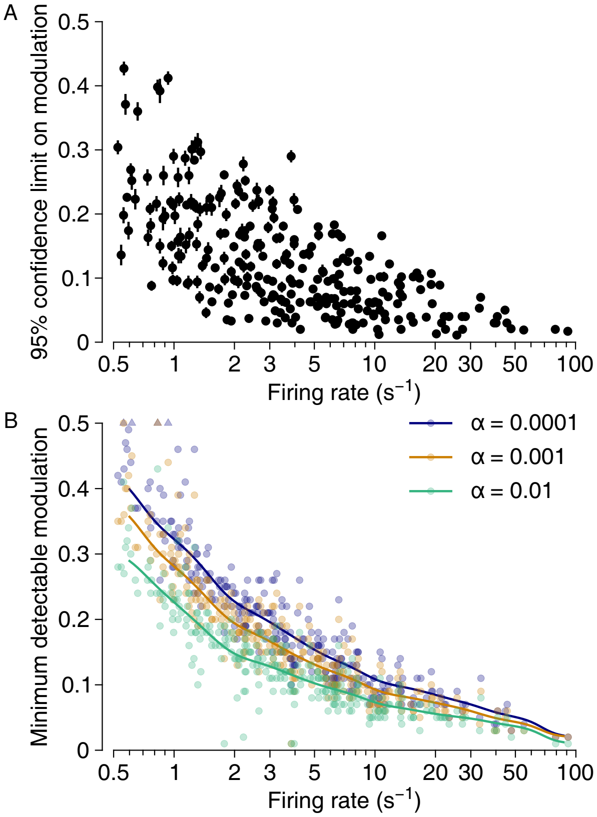


Figure 2.7: **Detectability of stimulus-induced modulations.** (A) Confidence limits on the magnitude of periodic modulation. Each dot represents a spike train from a different neuron. The confidence limit is based on comparison with artificially modulated versions of the same spike train. (B) Minimum detectable modulation for spike trains with different firing rates at several levels α of false-positive detection. Lines are interpolations of the data. Triangles mark outliers: spike trains for which no level of modulation up to 0.5 yielded detectability.

recordings from animals exposed to magnetic fields oscillating at 3 Hz and three recordings at 5 Hz, we repeated the above procedures, analyzing each recording at its stimulation frequency (Fig. 2.8). No cells in any of the experiments exhibited modulation exceeding what would be expected from the null model.

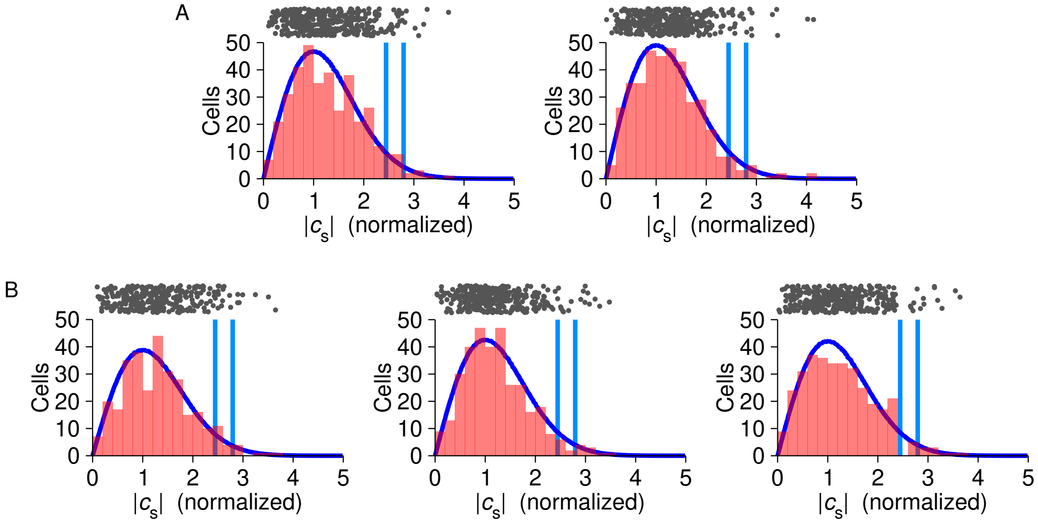


Figure 2.8: **Analysis of recordings during magnetic stimulation.** Display as in Fig 2.6. (A) Two recordings during 3-Hz stimulation. (B) Three recordings during 5-Hz stimulation.

2.5 Appendix: Derivations of probability distributions

2.5.1 The normal distribution in two dimensions

(This derivation, which can be found in many textbooks, is given here for easy reference and to introduce notation.)

Let X be a random variable, normally distributed with mean $\mu = 0$ and variance σ^2. The probability distribution of X is the well-known bell curve:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{1}{2}x^2/\sigma^2}.$$

Let Y be a second such random variable, same distribution, independent of X. Then the joint probability distribution of X and Y is

$$P(x, y) = \frac{1}{2\pi\sigma^2}\,\mathrm{e}^{-\frac{1}{2}(x^2+y^2)/\sigma^2}. \tag{2.7}$$

Let us now introduce $R = \sqrt{X^2 + Y^2}$. What is the distribution of R? We can find out by transforming to polar coordinates ($X = R\cos\Theta$, $Y = R\sin\Theta$):

$$P(r, \theta) = P(x, y)\left|\frac{\partial x\,\partial y}{\partial r\,\partial\theta}\right| = rP(x, y). \tag{2.8}$$

Since the distribution (Eqn. 2.7) is obviously independent of the angle θ, we have

$$P(\theta) = \frac{1}{2\pi}$$

and hence

$$P(r, \theta) = P(r)\,P(\theta) = \frac{1}{2\pi}P(r). \tag{2.9}$$

Combining Eqns. 2.7, 2.8, and 2.9, we conclude:

$$P(r) = \frac{r}{\sigma^2}\,\mathrm{e}^{-\frac{1}{2}r^2/\sigma^2}. \tag{2.10}$$

If the variable of interest is normalized to unit standard deviation, such that $\langle r^2\rangle = \sigma^2 = 1$, then

$$P(r) = r\,\mathrm{e}^{-\frac{1}{2}r^2} \tag{2.11}$$

as used in Eqn 2.6.

2.5.2 Fourier coefficients of Poisson spike trains

As shown in Eqn 2.2, the Fourier coefficients of a Poisson spike train with N spikes all have the same probability distribution: they are essentially the sum of N random unit vectors in the 2-dimensional plane. Say

$$\mathbf{u} = \sum_{i=1}^{N} \mathbf{s}_i,$$

where the $\mathbf{s}_i$ are random unit vectors. Then the covariance matrix of $\mathbf{u}$ is

$$\begin{aligned}\mathbf{C} = \left\langle \mathbf{u}\,\mathbf{u}^\top \right\rangle &= \sum_i \left\langle \mathbf{s}_i\,\mathbf{s}_i^\top \right\rangle + \sum_{i \neq j} \left\langle \mathbf{s}_i\,\mathbf{s}_j^\top \right\rangle \\ &= N \left\langle \mathbf{s}\,\mathbf{s}^\top \right\rangle + \sum_{i \neq j} \left\langle \mathbf{s}_i \right\rangle \left\langle \mathbf{s}_j \right\rangle^\top \\ &= \frac{N}{2}\mathbf{1},\end{aligned}$$

where all the sums are from 1 to N, and $\mathbf{1}$ is the identity matrix. To understand the last step note that for a random unit vector in two dimensions

$$\left\langle \mathbf{s}\,\mathbf{s}^\top \right\rangle = \begin{pmatrix} \left\langle x^2 \right\rangle & \langle xy \rangle \\ \langle yx \rangle & \left\langle y^2 \right\rangle \end{pmatrix}. \tag{2.12}$$

Because of normalization $\left\langle x^2 \right\rangle + \left\langle y^2 \right\rangle = 1$, and because of symmetry $\left\langle x^2 \right\rangle = \left\langle y^2 \right\rangle$ and $\langle xy \rangle = 0$. Therefore $\langle \mathbf{s}\,\mathbf{s}^\top \rangle = \frac{1}{2}\mathbf{1}$.

If N is reasonably large, the Central Limit Theorem applies: Since $\mathbf{u}$ is the sum of many independent random variables, its distribution is a Gaussian with the same covariance matrix as $\mathbf{u}$:

$$\begin{aligned}P(\mathbf{u}) &= \frac{1}{2\pi\sqrt{\det \mathbf{C}}} \mathrm{e}^{-\frac{1}{2}\mathbf{u}^\top \mathbf{C}^{-1}\mathbf{u}} \\ &= \frac{1}{\pi N} \mathrm{e}^{-\frac{|\mathbf{u}|^2}{N}}.\end{aligned}$$

The Fourier coefficient (Eqn. 2.2) is

$$\mathbf{c} = \frac{1}{T}\mathbf{u},$$

so its distribution is the 2-D Gaussian

$$P(\mathbf{c}) = \frac{T^2}{\pi N} \mathrm{e}^{-\frac{T^2 ||\mathbf{c}||^2}{N}}.$$

Using Eqn 2.10, one concludes that the vector length of **c** has distribution

$$P(||\mathbf{c}||) = \frac{2||\mathbf{c}||T^2}{N} \mathrm{e}^{-||\mathbf{c}||^2 T^2/N},$$

as used in Eqn 2.3. (Note that in the main text, we introduced c as a complex number, whereas here we discussed **c** as a 2D vector. The difference is immaterial, since addition and absolute value ($|c|$) for complex numbers are defined in terms of real and imaginary components just as addition and vector length ($||\mathbf{c}||$) are for 2D vectors.)

Chapter 3

MOUSE MAZE NAVIGATION

Mice in a Labyrinth Exhibit
Rapid Learning, Sudden Insight, and Efficient Exploration

Matthew Rosenberg*, Tony Zhang*, Pietro Perona, Markus Meister
* authors contributed equally

eLife, July 1, 2021

3.1 Abstract

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

3.2 Introduction

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

In laboratory studies, one prominent instance of one-shot learning is the Bruce effect [40]. Here the female mouse forms an olfactory memory of her mating partner that allows her to terminate the pregnancy if she encounters another male that threatens infanticide. Another form of rapid learning accessible to laboratory experiments is fear conditioning, where a formerly innocuous stimulus gets associated with a painful experience, leading to subsequent avoidance of the stimulus [34, 71]. These learning systems appear designed for special purposes, they perform very specific associations, and govern binary behavioral decisions. They are likely implemented by specialized brain circuits, and indeed great progress has been made in localizing these operations to the accessory olfactory bulb [38] and the cortical amygdala [139].

In the attempt to identify more generalizable mechanisms of learning and decision making, one route has been to train laboratory animals on abstract tasks with tightly specified sensory inputs that are linked to motor outputs via arbitrary contingency rules. Canonical examples are a monkey reporting motion in a visual stimulus by saccading its eyes [178], and a mouse in a box classifying stimuli by moving its forelimbs or the tongue [46, 98]. The tasks are of low complexity, typically a 1-bit decision based on 1 or 2 bits of input. Remarkably they are learned exceedingly slowly: A mouse typically requires many weeks of shaping and thousands of trials to reach asymptotic performance; a monkey may require many months [48].

What is needed therefore is a rodent behavior that involves complex decision making, with many input variables and many possible choices. Ideally the animals would learn to perform this task without excessive intervention by human shaping, so we may be confident that they employ innate brain mechanisms rather than circuits created by the training. Obviously the behavior should be easy to measure in the laboratory. Finally, it would be satisfying if this behavior showed a glimpse of rapid learning.

Navigation through space is a complex behavior displayed by many animals. It typically involves integrating multiple cues to decide among many possible actions. It relies intimately on rapid learning. For example a pigeon or desert ant leaving its shelter acquires the information needed for the homing path in a single episode. Major questions remain about how the brain stores this information and converts it to a policy for decisions during the homing path. One way to formalize the act of decision-making in the laboratory is to introduce structure in the environment in the form of a maze that defines straight paths and decision points. A maze of tunnels is in fact a natural environment for a burrowing rodent. Early studies of rodent behavior did place the animals into true labyrinths [223], but their use gradually declined in favor of linear tracks or boxes with a single choice point.

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

3.3 Results

3.3.1 Adaptation to the maze

At the start of the experiment, a single mouse was placed in a conventional mouse cage with bedding and food. A short tunnel offered free access to a maze consisting of a warren of corridors (Figure 3.1A-B). The bottom and walls of the maze were constructed of black plastic that is transparent in the infrared. A video camera placed below the maze captured the animal's actions continuously using infrared illumination (Figure 3.1B). The recordings were analyzed offline to track the movements of the mouse, with keypoints on the nose, mid-body, tail base, and the four feet (Figure 3.1D). All observations were made in darkness during the animal's subjective night.

The logical structure of the maze is a binary tree, with 6 levels of branches, leading from the single entrance to 64 endpoints (Figure 3.1C). A total of 63 T-junctions are connected by straight corridors in a design with maximal symmetry (Figure 3.1A, Figure 3.3–Figure 3.16, such that all the nodes at a given level of the tree have the same local geometry. One of the 64 endpoints of the maze is outfitted with a water port. After activation by a brief nose poke, the port delivers a small drop of water, followed by a 90-s time-out period.

After an initial period of exploratory experiments we settled on a frozen protocol that was applied to 20 animals. Ten of these mice had been mildly water-deprived for up to 24 hours; they received food in the home cage and water only from the port hidden in the maze. Another ten mice had free access to food and water in the cage,

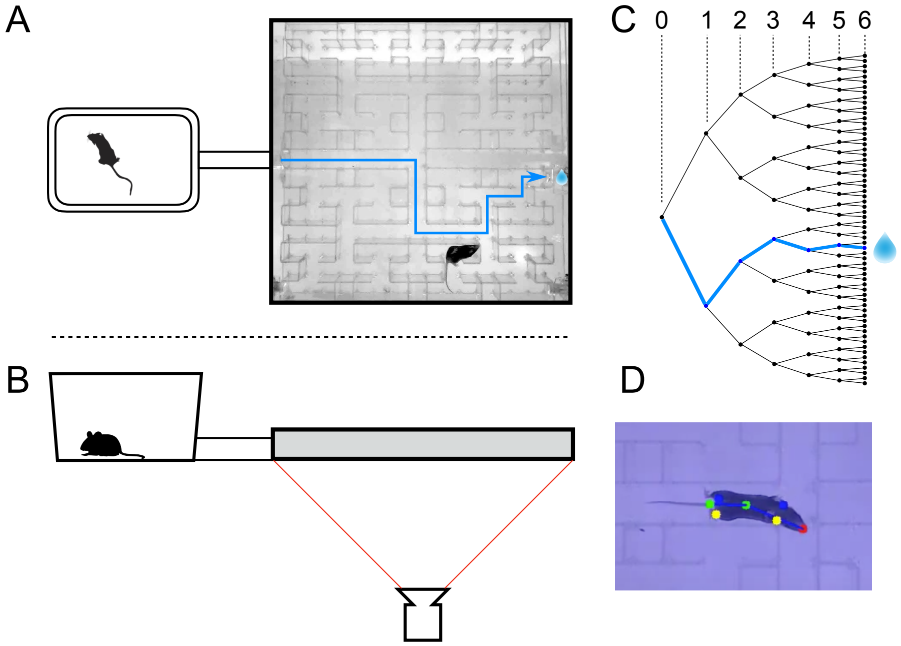


Figure 3.1: **The maze environment.** Top (A) and side (B) views of a home cage, connected via an entry tunnel to an enclosed labyrinth. The animal's actions in the maze are recorded via video from below using infrared illumination. (C) The maze is structured as a binary tree with 63 branch points (in 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 3.12 Occupancy of the maze.
Figure 3.13 Fraction of time in maze by group.
Figure 3.14 Transitions between cage and maze.

and received no water from the port in the maze. Each animal's behavior in the maze was recorded continuously for 7 h during the first night of its experience with the maze, starting the moment the connection tunnel was opened (sample videos here). The investigator played no role during this period, and the animal was free to act as it wished including travel between the cage and the maze.

All of the mice except one passed between the cage and the maze readily and frequently (Figure 3.1–Figure 3.12). The single outlier animal barely entered the maze and never progressed past the first junction; we excluded this mouse's data from subsequent analysis. On average over the entire period of study the animals spent

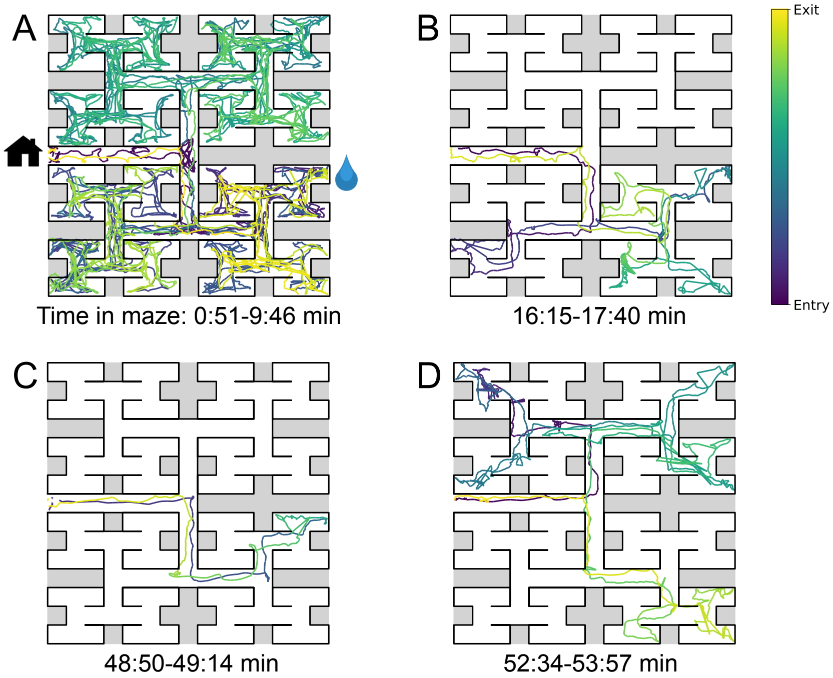


Figure 3.2: **Sample trajectories during adaptation to the maze.** Four sample bouts from one mouse (B3) into the maze at various times during the experiment (time markings at bottom). The trajectory of the animal's nose is shown; time is encoded by the color of the trace. The entrance from the home cage and the water port are indicated in panel A.
Figure 3.15 Speed of locomotion.

46% of the time in the maze (Figure 3.1–Figure 3.13). This fraction was similar whether or not the animal was motivated by water rewards (47% for rewarded vs 44% for unrewarded animals). Over time the animals appeared increasingly comfortable in the maze, taking breaks for grooming and the occasional nap. When the investigator lifted the cage lid at the end of the night some animals were seen to escape into the safety of the maze.

We examined the rate of transitions from the cage to the maze and how it depends on time spent in the cage (Figure 3.1–Figure 3.14A). Surprisingly the rate of entry into the maze is highest immediately after the animal returns to the cage. Then it declines gradually by a factor of 4 over the first minute in the cage and remains steady thereafter. This is a large effect, observed for every individual animal in both the rewarded and unrewarded groups. By contrast, the opposite transition, namely exit from the maze, occurs at an essentially constant rate throughout the visit (Figure 3.1–Figure 3.14B).

The nature of the animal's forays into the maze changed over time. We call each foray from entrance to exit a "bout." After a few hesitant entries into the main corridor, the mouse engaged in one or more long bouts that dove deep into the binary tree to most or all of the leaf nodes (Figure 3.2A). For a water-deprived animal, this typically led to discovery of the reward port. After ~10 bouts, the trajectories became more focused, involving travel to the reward port and some additional exploration (Figure 3.2B). At a later stage still, the animal often executed perfect exploitation bouts that led straight to the reward port and back with no wrong turns (Figure 3.2C). Even at this late stage, however, the animal continued to explore other parts of the maze (Figure 3.2D). Similarly the unrewarded animals explored the maze throughout the night (Figure 3.1–Figure 3.13). While the length and structure of the animal's trajectories changed over time, the speed remained remarkably constant after ~50 s of adaptation (Figure 3.2–Figure 3.18).

Whereas Figure 3.2 illustrates the trajectory of a mouse's nose in full spatio-temporal detail, a convenient reduced representation is the "node sequence." This simply marks the events when the animal enters each of the 127 nodes of the binary tree that describes the maze (see Methods and Figure 3.3–Figure 3.16). Among these nodes, 63 are T-junctions where the animal has 3 choices for the next node, and 64 are end nodes where the animal's only choice is to reverse course. We call the transition from one node to the next a "step." The analysis in the rest of the paper was carried out on the animal's node sequence.

3.3.2 Few-shot learning of a reward location

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

The ten water-deprived animals had no indication that water would be found in the maze. Yet, all 10 discovered the water port in less than 2000 s and fewer than 17 bouts (Figure 3.3A). The port dispensed only a drop of water followed by a 90-s timeout before rearming. During the timeout the animals generally left the port location to explore other parts of the maze or return home, even though they were not obliged to do so. For each of the water-deprived animals, the frequency at which it consumed rewards in the maze increased rapidly as it learned how to find the water port, then settled after a few reward experiences (Figure 3.3A).

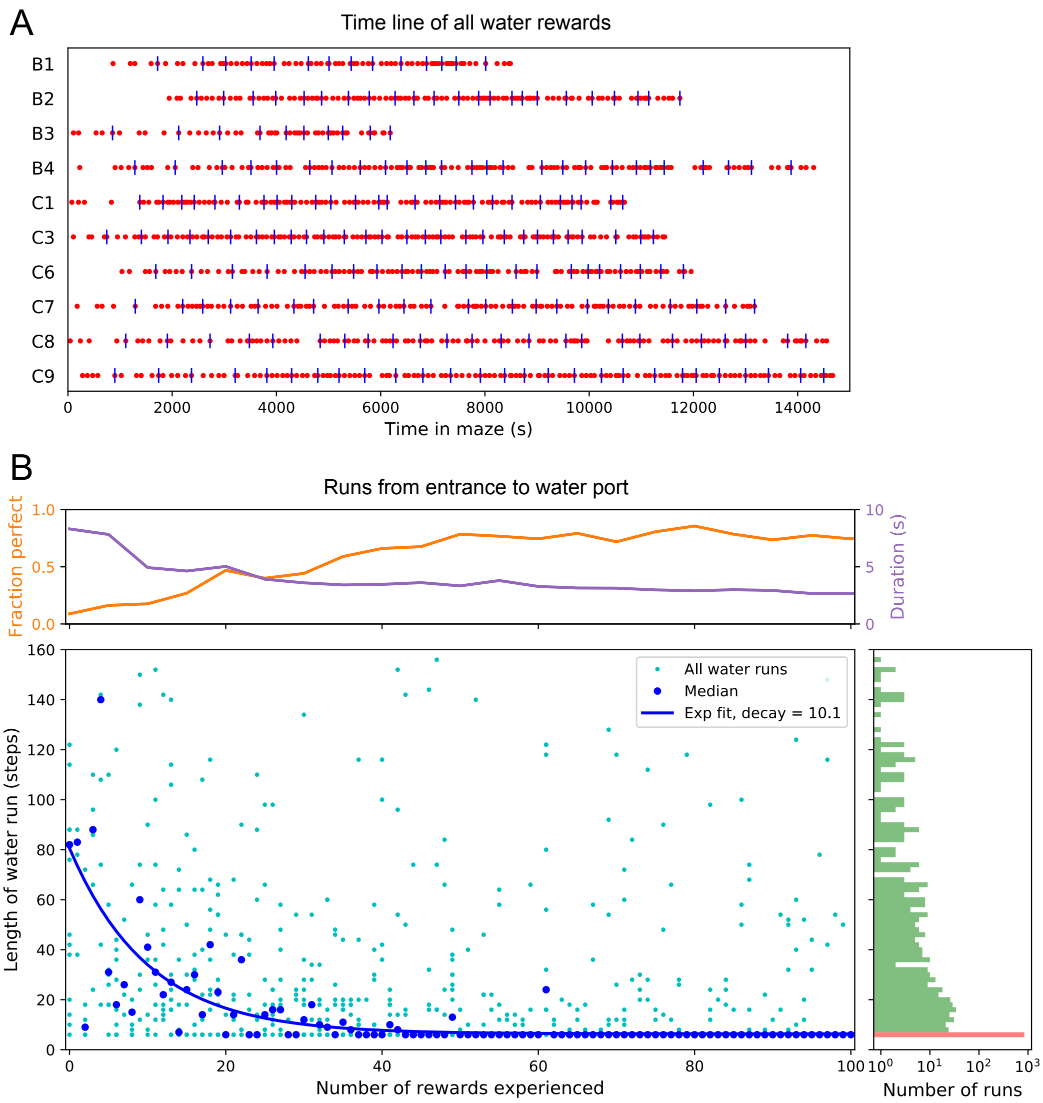


Figure 3.3: **Few-shot learning of path to water.** (A) Time line of all water rewards collected by 10 water-deprived mice (red dots, every fifth reward has a blue tick mark). (B) The length of runs from the entrance to the water port, measured in steps between nodes, and plotted against the number of 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 3.16 Definition of node trajectories.

How many reward experiences are sufficient to teach the animal reliable navigation to the water port? To establish a learning curve one wants to compare performance on the identical task over successive trials. Recall that this experiment has no imposed trial structure. Yet the animals naturally segmented their behavior through discrete visits to the maze. Thus we focused on all the instances when the animal started at the maze entrance and walked to the water port (Figure 3.3B).

On the first few occasions, these paths to water can involve hundreds of steps between nodes and their length scatters over a wide range. However, after a few rewards, the animals began taking the perfect path without detours (6 steps, Figure 3.3–Figure 3.16), and soon that became the norm. Note, the path length plotted here is directly related to the number of “turning errors”: every time the mouse turns away from the shortest path to the water port that adds two steps to the path length (Equation 3.7). The rate of these errors declined over time, by a factor of e after ~10 rewards consumed (Figure 3.3B). Late in the night ~75% of the paths to water were perfect. The animals executed them with increasing speed; eventually these fast “water runs” took as little as 2 s (Figure 3.3B). Many of these visits went unrewarded owing to the 90-s timeout period on the water port.

In summary, after ~10 reward experiences on average the mice learn to navigate efficiently to the water port, which requires making 6 correct decisions, each among 3 options. Note that 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 3.7).

3.3.3 The role of cues attached to the maze

These observations of rapid learning raise the question “How do the animals navigate?” In particular, does the mouse build an internal representation that guides its action at every junction? Or does it place marks in the external environment that signal the route to the water port? In an extreme version of externalized cognition, the mouse leaves behind a trail of urine marks or other secretions as it walks away from the water port, and on a subsequent bout simply sniffs its way up the odor gradient (Figure 3.4A). This would require no internal representation.

The following experiment offers some partial insights. Owing to the design of the labyrinth, one can rotate the entire apparatus by 180 degrees, open one wall and close another, and obtain a maze with the same structure (Figure 3.4A). Alternatively, one can also rotate only the floor. After such a modification, all the physical cues attached to the rotated parts now point in the wrong direction, namely to the end node 180 degrees opposite the water port (the “image location"). If the animal navigated to the

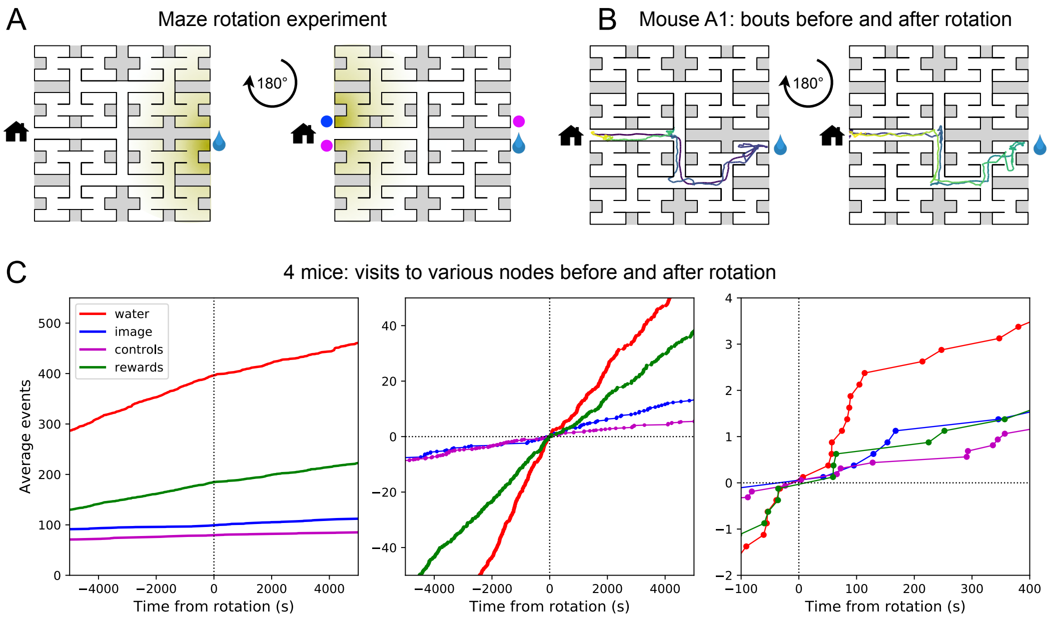


Figure 3.4: **Navigation is robust to rotation of the maze.** (A) Logic of the experiment: The animal may have deposited an odorant in the maze (shading) that is centered on the water port. After 180 degree rotation of the maze, that gradient would lead to the image of the water port (blue dot). We also 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 3.2. (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 3.17 Navigation before and after maze rotation for each animal.
Figure 3.18 Speed before and after maze rotation.

goal following cues previously deposited in the maze, it should end up at that image location.

We performed a maze rotation on four animals after several hours of exposure, when they had acquired the perfect route to water. Immediately after rotation, 3 of the 4 animals went to the correct water port on their first entry into the maze, and before ever visiting the image location (e.g. Figure 3.4B). The fourth mouse visited the image location once and then the correct water port (Figure 3.4–Figure 3.17). The mice continued to collect water rewards efficiently even immediately after the rotation.

Nonetheless, the maze rotation did introduce subtle changes in behavior that lasted for an hour or more (Figure 3.4C). Visits to the image location were at chance levels prior to rotation, then increased by a factor of 1.8. Visits to the water port declined in frequency, although they still exceeded visits to the image location by a factor of 5. The reward rate declined by a factor of 0.7. These effects could be verified for each animal (Figure 3.4–Figure 3.17). The speed of the mice was not disturbed (Figure 3.4–Figure 3.18).

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

3.3.4 Discontinuous learning

While an average across animals shows evidence of rapid learning (Figure 3.3), one wonders whether the knowledge is acquired gradually or discontinuously, through moments of "sudden insight." To explore this, we scrutinized more closely the time line of individual water-deprived animals in their experience with the maze. The discovery of the water port and the subsequent collection of water drops at a regular rate is one clear change in behavior that relies on new knowledge. Indeed, the rate of water rewards can increase rather suddenly (Figure 3.3A), suggesting an instantaneous step in knowledge.

Over time, the animals learned the path to water not only from the entrance of the maze but from many locations scattered throughout the maze. The largest distance between the water port and an end node in the opposite half of the maze involves 12 steps through 11 intersections (Figure 3.5A). Thus we included as another behavioral variable the occurrence of long direct paths to the water port which reflects how directedly the animals navigate within the maze.

Figure 3.5B shows for one animal the cumulative occurrence of water rewards and that of long direct paths to water. The animal discovers the water port early on at 75 s, but at 1380 s the rate of water rewards jumps suddenly by a factor of 5. The

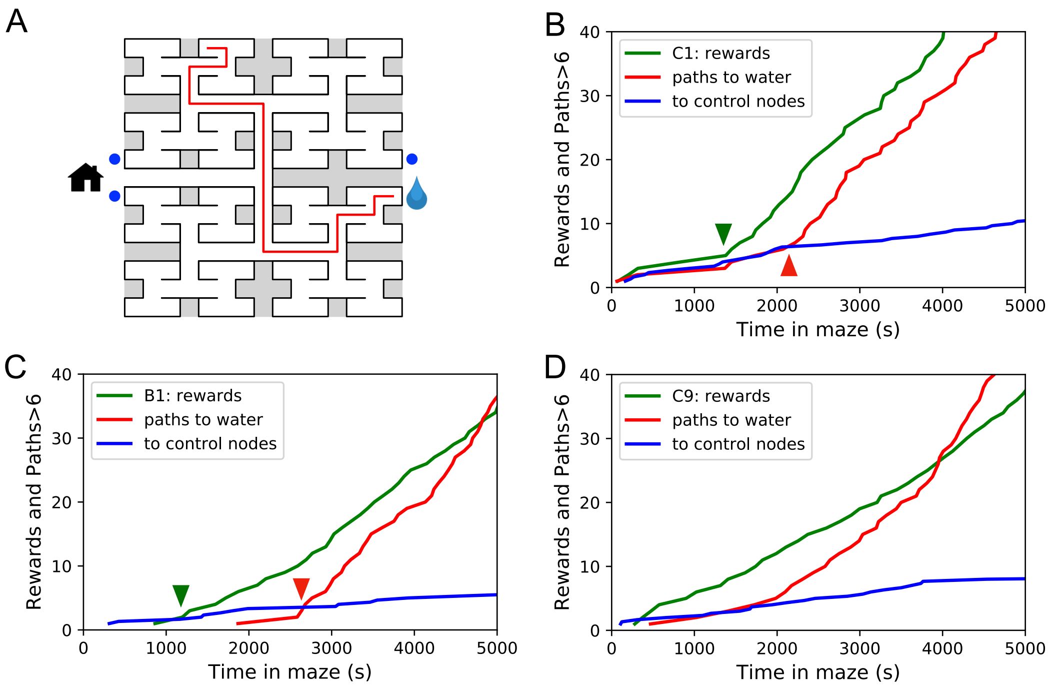


Figure 3.5: **Sudden changes in behavior.** (A) An example of a long uninterrupted path through 11 junctions to the water port (drop icon). Blue circles mark control nodes related by symmetry to the water port to assess the frequency of long paths occurring by chance. (B) For one animal (named C1) the 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.
Source data 1. Statistics of sudden changes in behavior.
Figure 3.19 Long direct paths for all animals.

long paths to water follow a rather different time line. At first they occur randomly, at the same rate as the paths to the unrewarded control nodes. At 2070 s the long paths suddenly increase in frequency by a factor of 5. Given the sudden change in rates of both kinds of events there is little ambiguity about when the two steps happen and they are well separated in time (Figure 3.5B).

The animal behaves as though it gains a new insight at the time of the second step that allows it to travel to the water port directly from elsewhere in the maze. Note that the two behavioral variables are independent: The long paths do not change when the reward rate steps up, and the reward rate does not change when the rate of long paths steps up. Another animal (Figure 3.5C) similarly showed an early step in the reward rate (at 860 s) and a dramatic step in the rate of long paths (at 2580 s). In

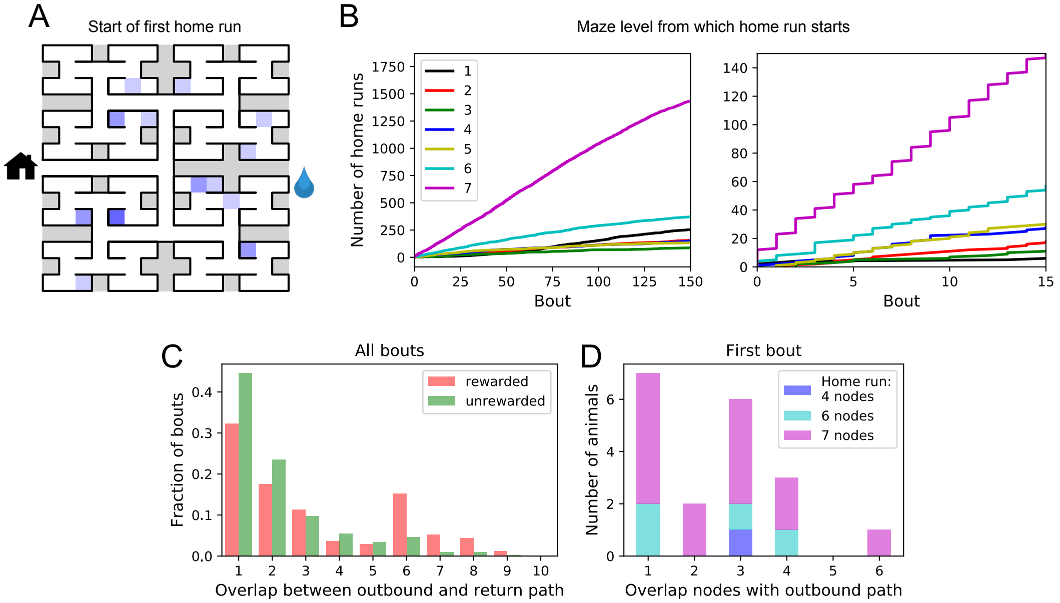


Figure 3.6: **Homing succeeds on first attempt.** (A) Locations in the maze where the 19 animals started their first return to the exit (home run). Some locations were used by 2 or 3 animals (darker color). (B) Left: The cumulative number of home runs from different levels in the maze, summed over all 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*.

this case the emergence of long paths coincided with a modest increase (factor of 2) in the reward rate.

Similar discontinuities in behavior were seen in at least 5 of the 10 water-deprived animals (Figure 3.5B, Figure 3.5–Figure 3.19, Figure 3.5–Figure 3.20), and their timing could be identified to a precision of ~200 s. More gradual performance change was observed for the remaining animals (Figure 3.5 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.

3.3.5 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 [80, 242]. 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 3.3–Figure 3.16). Figure 3.6 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 home runs should appear before long ones in the course of the experiment. The opposite is the case (Figure 3.6 B). In fact, the end nodes (level 7 of the maze) are by far the favorite place from which to return to the exit, and those maximal-length home runs systematically appear before shorter ones. This conclusion was confirmed for each individual animal, whether rewarded or unrewarded.

Clearly, the animals do not practice the home path or build it up gradually. Instead they seem to possess an Ariadne's thread [191] starting with their first excursion into the maze, long before they might have acquired any general knowledge of the maze layout. On the other hand the mouse does not follow the strategy of Theseus, namely to precisely retrace the path that led it into the labyrinth. In that case, the animal's home path should be the reverse of the path into the maze that started the bout. Instead the entry path and the home path tend to have little overlap (Figure 3.6C). Note, the minimum overlap is 1, because all paths into and out of the maze have to pass through the central junction (node 0 in Figure 3.3–Figure 3.16). This is also the most frequent overlap. The peak at overlaps 6-8 for rewarded animals results from the frequent paths to the water port and back, a sequence of at least 7 nodes in each direction. The separation of outbound and return path is seen even on the very first home run (Figure 3.6D). Many home runs from the deepest level (7 nodes) have only the central junction in common with the outbound path (overlap = 1).

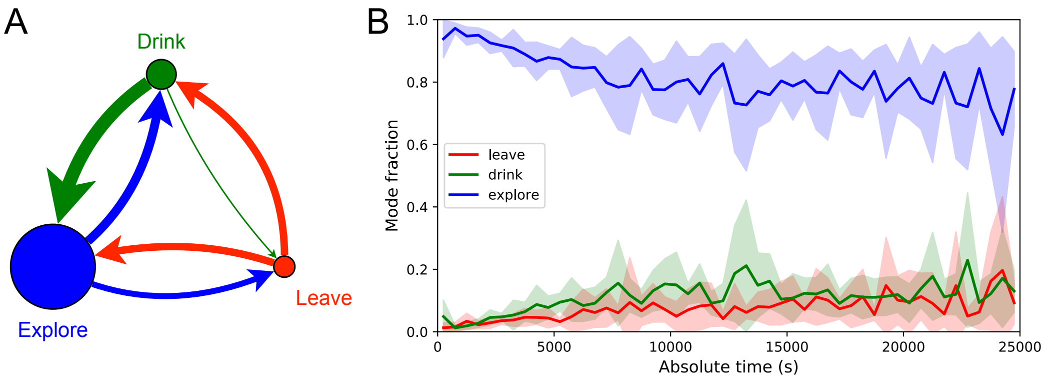


Figure 3.7: **Exploration is a dominant and persistent mode of behavior.** (A) Ethogram for rewarded animals. Area of the circle reflects the fraction of time spent in each behavioral mode averaged over animals and duration of the experiment. Width of the arrow reflects the probability of transitioning 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 ± SD across the 10 rewarded animals.
Source data 1. Three modes of behavior.

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

3.3.6 Structure of behavior in the maze

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

Behavioral states

Once the animal has learned to perform long uninterrupted paths to the water port, one can categorize its behavior within the maze by three states: (1) walking to the water port; (2) 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 3.7 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 3.7–Figure 3.21). The rewarded

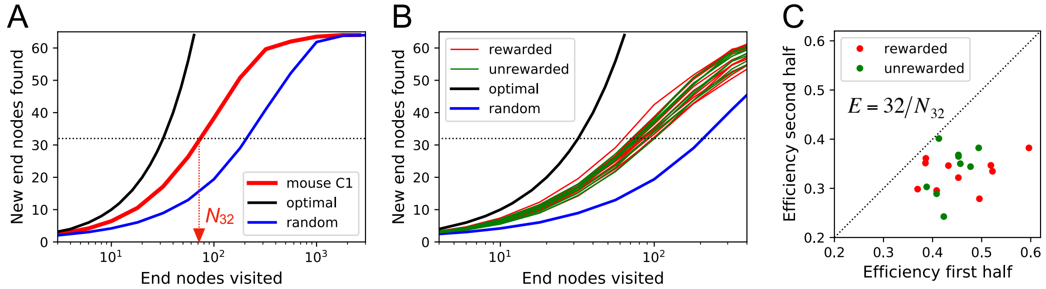


Figure 3.8: **Exploration covers the maze efficiently.** (A) The number of distinct end nodes encountered as a function of the number of end nodes visited for: mouse C1 (red); the optimal explorer agent (black); an unbiased random walk (blue). Arrowhead: the value $N_{32} = 76$ by which mouse C1 discovered 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 ± 0.050 (SD) for rewarded and 0.384 ± 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 ± 0.12 (SD) for rewarded and 0.81 ± 0.13 (SD) for unrewarded mice. **Figure 3.22** Efficiency of exploration.

mice began about half their bouts into the maze with a trip to the water port and the other half by exploring (Figure 3.7A). 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 3.7–Figure 3.21). One suspects that the animals derive some intrinsic reward from the act of patrolling the environment itself.

Efficiency of exploration

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

One can presume that a goal of the exploratory mode is to rapidly survey all parts of the environment for the appearance of new resources or threats. We will measure the efficiency of exploration by how rapidly the animal visits all end nodes of the binary maze, starting at any time during the experiment. The optimal agent with perfect memory and complete knowledge of the maze—including the absence of any loops—could visit the end nodes systematically one after another without repeats, thus encountering all of them after just 64 visits. A less perfect agent, on the other hand, will visit the same node repeatedly before having encountered all of them. Figure 3.8A plots for one exploring mouse the number of distinct end nodes it encountered as a function of the number of end nodes visited. The number of new nodes rises 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

$$E = \frac{32}{N_{32}}. \tag{3.1}$$

This mouse explores with efficiency $E = 32/76 = 0.42$. For comparison, Figure 3.8A 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 3.8B): across animals the efficiency varied by only 11% of the mean (0.387 ± 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 3.8C).

Rules of exploration

What allows the mice to search much more efficiently than a random walking agent? We inspected more closely the decisions that the animals make at each 3-way junction. It emerged that these decisions are governed by strong biases (Figure 3.9). The probability of choosing each arm of a T-junction depends crucially on how the animal entered the junction. The animal can enter a T-junction from 3 places and exit it in 3 directions (Figure 3.9A). By tallying the frequency of all these occurrences across all T-junctions in the maze one finds clear deviations from an unbiased random walk (Figure 3.9B, Figure 3.9–Figure 3.23).

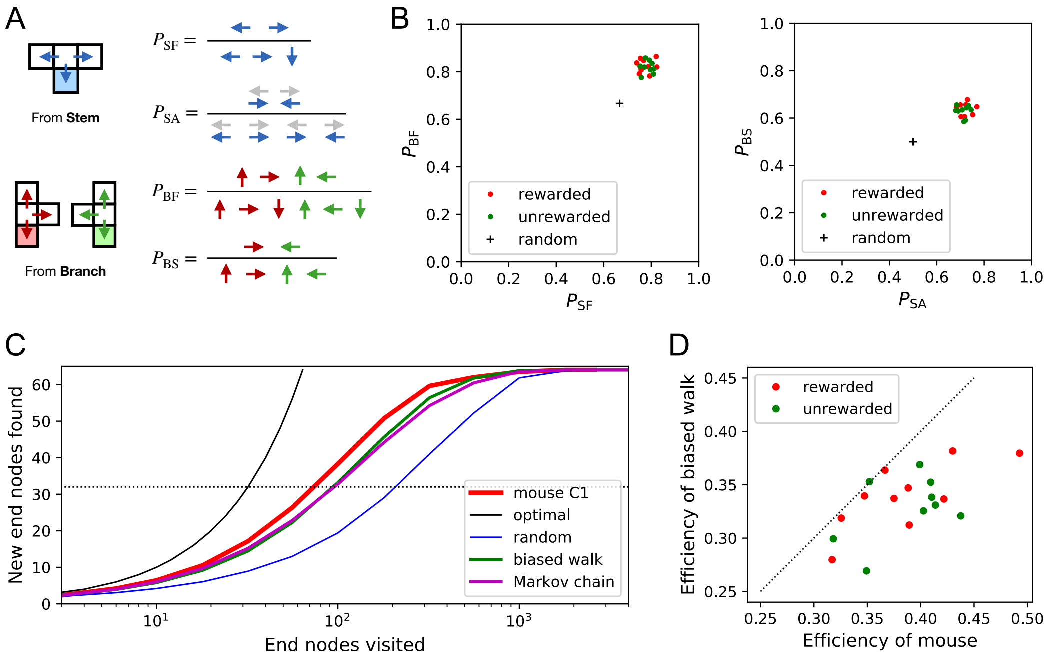


Figure 3.9: **Turning biases favor exploration.** (A) Definition of four turning biases at a T-junction based on the ratios of actions taken. Top: An animal arriving from the stem of the T (shaded) may either reverse or turn left or right. P_{SF} is the probability that it will move forward rather than 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 3.8A, 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 3.11C). (D) Efficiency of exploration (Equation 3.1) in 19 mice compared to the efficiency of the corresponding biased random walk.
Source data 1. Bias statistics.

First, the animals have a strong preference for proceeding through a junction rather than returning to the preceding node (P_{SF} and P_{BF} in Figure 3.9B). Second there is a bias in favor of alternating turns left and right rather than repeating the same direction turn (P_{SA}). Finally, the mice have a mild preference for taking a branch off the straight corridor rather than proceeding straight (P_{BS}). A comparison across animals again revealed a remarkable degree of consistency even in these local rules

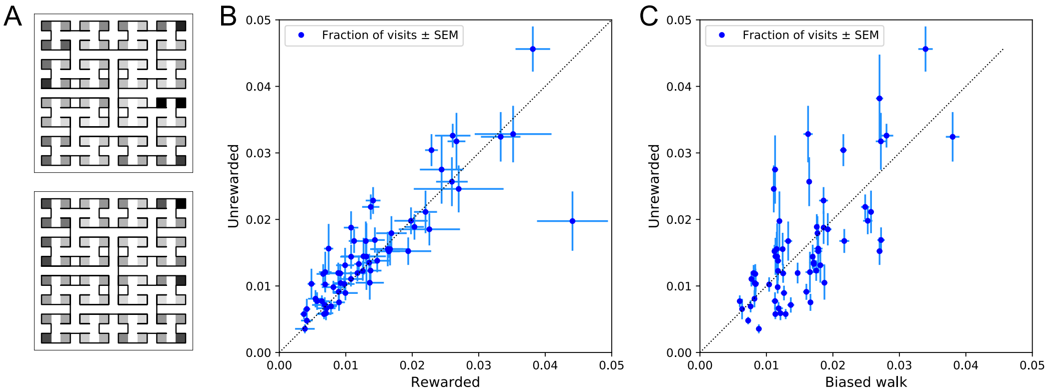


Figure 3.10: **Preference for certain end nodes during exploration.** (A) The number of visits to different end nodes encoded by a gray scale. Top: rewarded, bottom: unrewarded animals. Gray scale spans a factor of 12 (top) or 13 (bottom). (B) The fraction of visits to each end node, comparing the rewarded 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.

of behavior: The turning biases varied by only 3% across the population and even between the rewarded and unrewarded groups (Figure 3.9B, Figure 3.9–Figure 3.23).

Qualitatively, one can see that these turning biases will improve the animal's search strategy. The forward biases P_{SF} and P_{BF} keep the animal from re-entering territory it has covered already. The bias P_{BS} favors taking a branch that leads out of the maze. This allows the animal to rapidly cross multiple levels during an outward path and then enter a different territory. By comparison, the unbiased random walk tends to get stuck in the tips of the tree and revisits the same end nodes many times before escaping. To test this intuition we simulated a biased random agent whose turning probabilities at a T-junction followed the same biases as measured from the animal (Figure 3.9C). These biased agents did in fact search with much higher efficiency than the unbiased random walk. They did not fully explain the behavior of the mice (Figure 3.9D), accounting for ~87% of the animal's efficiency (compared to 60% for the random walk). A more sophisticated model of the animal's behavior—involving many more parameters (Figure 3.11C)—failed to get any closer to the observed efficiency (Figure 3.9C, Figure 3.8–Figure 3.22C). Clearly some components of efficient search in these mice remain to be understood.

Systematic node preferences

A surprising aspect of the animals' explorations is that they visit certain end nodes of the binary tree much more frequently than others (Figure 3.10). This effect is large: more than a factor of 10 difference between the occupancy of the most popular and least popular end nodes (Figure 3.10A-B). This was surprising given our efforts to design the maze symmetrically, such that in principle all end nodes should be equivalent. Furthermore the node preferences were very consistent across animals and even across the rewarded and unrewarded groups. Note that the standard error across animals of each node's occupancy is much smaller than the differences between the nodes (Figure 3.10B).

The nodes on the periphery of the maze are systematically preferred. Comparing the outermost ring of 26 end nodes (excluding the water port and its neighbor) to the innermost 16 end nodes, the outer ones are favored by a large factor of 2.2. This may relate to earlier reports of a "centrifugal tendency" among rats patrolling a maze [253].

Interestingly, the biased random walk using four bias numbers (Figure 3.9, Figure 3.11D) replicates a good amount of the pattern of preferences. For unrewarded animals, where the maze symmetry is not disturbed by the water port, the biased random walk predicts 51% of the observed variance across nodes (Figure 3.10C), and an outer/inner node preference of 1.97, almost matching the observed ratio of 2.20. The more complex Markov-chain model of behavior (Figure 3.11C) performed slightly better, explaining 66% of the variance in port visits and matching the outer/inner node preference of 2.20.

3.3.7 Models of maze behavior

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

At a formal level, the mouse's trajectory through the maze is a string of numbers standing for the nodes the animal visited (Figure 3.11A and Figure 3.3–Figure 3.16). We want to predict the next action of the mouse, namely the step that takes it to the next node. The quality of the model will be assessed by the cross-entropy between

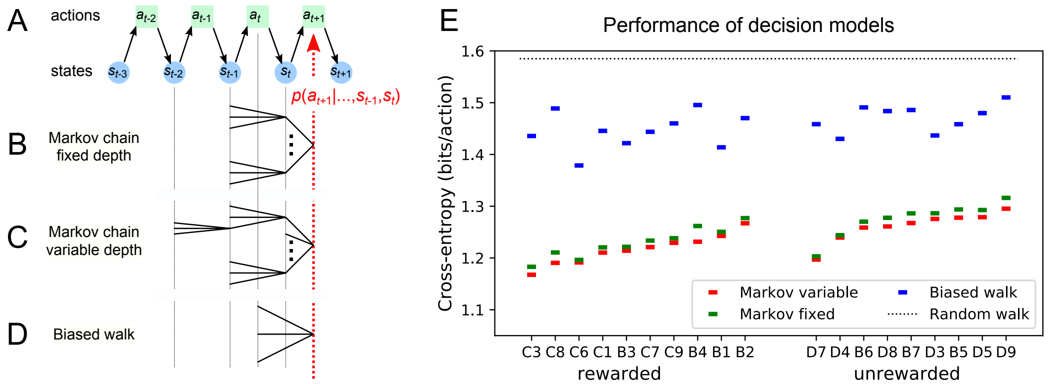


Figure 3.11: **Recent history constrains the mouse's decisions.** (A) The mouse's trajectory through the maze produces a sequence of states s_t = node occupied after step t. From each state, up to 3 possible actions lead to the next state (end nodes allow only one action). We want to predict the animal's 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 × 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 3.9, 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*, and *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 3.24. Markov model fits.

the model's predictions and the mouse's observed actions, measured in bits per action. This is the uncertainty that remains about the mouse's next action given the prediction from the model. The ultimate lower limit is the true source entropy of the mouse, namely that component of its decisions that cannot be explained by the history of its actions.

One family of models we considered are fixed-depth Markov chains (Figure 3.11B). Here the probability of the next action a_{t+1} is specified as a function of the history stretching over the k preceding nodes $(s_{t-k+1}, \ldots, s_t)$. In fitting the model to the mouse's actual node sequence one tallies how often each history leads to each action, and uses those counts to estimate the conditional probabilities $p(a_{t+1}|s_{t-k+1}, \ldots, s_t)$. Given a new node sequence, the model will then use the history strings $(s_{t-k+1}, \ldots, s_t)$

to predict the outcome of the next action. In practice we trained the model on 80% of the animal's trajectory and tested it by evaluating the cross-entropy on the remaining 20%.

Ideally, the depth k of these action trees would be very large, so as to take as much of the prior history into account as possible. However, one soon runs into a problem of over-fitting: Because each T-junction in the maze has 3 neighboring junctions, the number of possible histories grows as 3^k. As k increases, this quickly exceeds the length of the measured node sequence, so that every history appears only zero or one times in the data. At this point one can no longer estimate any probabilities, and cross-validation on a different segment of data fails catastrophically. In practice, we found that this limitation sets in already beyond $k = 2$ (Figure 3.11–Figure 3.24A). To address this issue of data-limitation we developed a variable-depth Markov chain (Figure 3.11C). This model retains longer histories, but only if they occur frequently enough to allow a reliable probability estimate (see Methods, Figure 3.11–Figure 3.24B-C). In addition, we explored different schemes of pooling the counts across certain T-junctions that are related by the symmetry of the maze (see Methods).

With these methods, we focused on the portions of trajectory when the mouse was in 'explore' mode, because the segments in 'drink' and 'leave' mode are fully predictable. Furthermore, we evaluated the models only at nodes corresponding to T-junctions, because the decision from an end node is again fully predictable. Figure 3.11E compares the performance of various models of mouse behavior. The variable-depth Markov chains routinely produced the best fits, although the improvement over fixed-depth models was modest. Across all 19 animals in this study the remaining uncertainty about the animal's action at a T-junction is 1.237 ± 0.035 (SD) bits/action, compared to the prior uncertainty of $\log_2 3 = 1.585$ bits. The rewarded animals have slightly lower entropy than the unrewarded ones (1.216 vs 1.261 bits/action). The Markov chain models that produced the best fits to the behavior used history strings with an average length of ~4.

We also evaluated the predictions obtained from the simple biased random walk model (Figure 3.11D). Recall that this attempts to capture the history-dependence with just 4 bias parameters (Figure 3.9A). As expected this produced considerably higher cross-entropies than the more sophisticated Markov chains (by about 18%, Figure 3.11E). Finally, we used several professional file compression routines to try and compress the mouse's node sequence. In principle, this sets an upper bound on the true source entropy of the mouse, even if the compression algorithm has no understanding of animal behavior. The best such algorithm (bzip2 compression [218]) far under-performed all the other models of mouse behavior, giving 43% higher cross-entropy on average, and thus offered no additional useful bounds.

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

3.4 Discussion

3.4.1 Summary of contributions

We present a new approach to the study of learning and decision-making in mice. We give the animal access to a complex labyrinth and leave it undisturbed for a night while monitoring its movements. The result is a rich data set that reveals new aspects of learning and the structure of exploratory behavior. With these methods we find that mice learn a complex task that requires 6 correct 3-way decisions after only ~10 experiences of success (Figure 3.2, Figure 3.3). Along the way the animal gains task knowledge in discontinuous steps that can be localized to within a few minutes of resolution (Figure 3.5). Underlying the learning process is an exploratory behavior that occupies 90% of the animal's time in the maze and persists long after the task has been mastered, even in complete absence of an extrinsic reward (Figure 3.7). The decisions the animal makes at choice points in the labyrinth are constrained in part by the history of its actions (Figure 3.9, Figure 3.11), in a way that favors efficient searching of the maze (Figure 3.8). This microstructure of behavior is surprisingly consistent across mice, with variation in parameters of only a few percent (Figure 3.9). Our most expressive models to predict the animal's choices still leave a remaining uncertainty of ~1.24 bits per decision (Figure 3.11), a quantitative benchmark by which competing models can be tested. Finally, some of the observations constrain what algorithms the animals might use for learning and navigation (Figure 3.4).

3.4.2 Historical context

Mazes have been a staple of animal psychology for well over 100 years. The early versions were true labyrinths. For example, Small [223] built a model of the maze in Hampton Court gardens scaled to rat size. Subsequent researchers felt less constrained by Victorian landscapes and began to simplify the maze concept. Most commonly the maze offered one standard path from a starting location to a food reward box. A few blind alleys would branch from the standard path, and researchers would tally how many errors the animal committed by briefly turning into a blind [248]. Later on, the design was further reduced to a single T-junction. After all, the elementary act of maze navigation is whether to turn left or right at a junction [249], so why not study that process in isolation? And reducing the concept even further, one can ask the animal to refrain from walking altogether, and instead poke its nose into a hole on the left or the right side of a box [252]. This led to the popular behavior boxes now found in rodent neuroscience laboratories everywhere. Each of these reductions of the "maze" concept enabled a new type of experiment to study learning and decision-making, for example limiting the number of choice points allows one to better sample neural activity at each one. However, the essence of a "confusing network of paths" has been lost along the way, and with it the behavioral richness of the animals navigating those decisions.

Owing in part to the dissemination of user-friendly tools for animal tracking, one sees a renaissance of experiments that embrace complex environments, including mazes with many choice points [4, 155, 175, 202, 211, 272, 278], 3-dimensional environments [97], and infinite mazes [219]. The labyrinth in the present study is considerably more complex than Hampton Court or most of the mazes employed by Tolman and others [43, 172, 248]. In those mazes the blind alleys are all short and unbranched; when an animal strays from the target path it receives feedback quickly and can correct. By contrast, our binary tree maze has 64 equally deep branches, only one of which contains the reward port. If the animal makes a mistake at any level of the tree it can find out only after traveling all the way to the last node.

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

Our experiments had no imposed structure whatsoever; in fact it may be inappropriate to call them experiments. The investigator opened the entry to the maze in the evening and did not return until the morning. A potential advantage of leaving the animals to themselves is that they are more likely to engage in mouse-like behavior, rather than constantly responding to the stress of human interference or the alienation from being a cog in a behavior machine. The result was a rich data set, with the typical animal delivering ~15,000 decisions in a single night, even if one only counts the nodes of the binary tree as decision points. Since the mice made all the choices, the scientific effort lay primarily in adapting methods of data analysis to the nature of mouse trajectories. Somewhat surprisingly, the absence of experimental structure was no obstacle to making precise and reproducible measurements of the animal's behavior.

3.4.3 How fast do animals learn?

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

Many well-studied behaviors have a complexity of 1 bit or less, and often animals can learn these associations after a single experience. For example, in the Bruce effect [40] the female maps two different contexts (smell of mate vs non-mate) onto two kinds of pregnancy outcomes (carry to term vs abort). The mutual information in that contingency table is at most 1 bit, and may be considerably lower, for example if non-mate males are very rare or very frequent. Mice form the correct association after a single instance of mating, although proper memory formation requires several hours of exposure to the mate odor [205].

Similarly fear learning under the common electroshock paradigm establishes a mapping between two contexts (paired with shock vs innocuous) and two actions (freeze vs proceed), again with an upper bound of 1 bit of complexity. Rats and mice will form the association after a single experience lasting only seconds, and alter their behavior over several hours [34, 71]. This is an adaptive warning system to deal with life-threatening events, and rapid learning here has a clear survival value.

Animals are particularly adept at learning a new association between an odor and food. For example bees will extend their proboscis in response to a new odor after just one pairing trial where the odor appeared together with sugar [27]. Similarly rodents will start digging for food in a scented bowl after just a few pairings with that odor [51]. Again, these are 1-bit tasks learned rapidly after one or a few experiences.

By comparison the tasks that a mouse performs in the labyrinth are more complex. For example, the path from the maze entrance to the water port involves 6 junctions, each with 3 options. At a minimum 6 different contexts must be mapped correctly into one of 3 actions each, which involves $6 \cdot \log_2 3 = 9.5$ bits of complexity. The animals begin to execute perfect paths from the entrance to the water port well within the first hour (Figure 3.2C, Figure 3.3B). At a later stage during the night the animal learns to walk direct paths to water from many different locations in the maze (Figure 3.5); by this time it has consumed 10-20 rewards. In the limit, if the animal could turn correctly towards water from each of 63 junctions in the maze, it would have learned $63 \cdot \log_2 3 = 100$ bits. Conservatively, we estimate that the animals have mastered 10-20 bits of complexity based on 10-20 reward experiences within an hour of time spent in the maze. Note, this considers only information about the water port and ignores whatever else the animals are learning about the maze during their incessant exploratory forays. These numbers align well with classic experiments on rats in diverse mazes and problem boxes [172]. Although those tasks come in many varieties, a common theme is that ~10 successful trials are sufficient to learn ~10 decisions [273].

In a different corner of the speed-complexity space are the many 2-alternative-forced-choice (2AFC) tasks in popular use today. These tend to be 1-bit tasks, for example the monkey should flick its eyes to the left when visual motion is to the left [178], or the mouse should turn a steering wheel to the right when a light appears on the left [46]. Yet, the animals take a long time to learn these simple tasks. For example, the mouse with the steering wheel requires about 10,000 experiences before performance saturates. It never gets particularly good, with a typical hit rate only 2/3 of the way from random to perfect. All this training takes 3-6 weeks; in the case of monkeys several months. The rate of learning, measured in task complexity per unit time, is surprisingly low: < 1 bit/month compared to ~10 bits/h observed in the labyrinth. The difference is a factor of 6,000. Similarly when measured in complexity learned per reward experience: The 2AFC mouse may need 5,000 rewards to learn a contingency table with 1 bit complexity, whereas the mouse in the maze needs ~10 rewards to learn 10 bits. Given these enormous differences in learning rate, one wonders whether the ultra-slow mode of learning has any relevance for an animal's natural condition. In the month that the 2AFC mouse requires to finally report the location of a light, its relative in the wild has developed from a baby to having its own babies. Along the way, that wild mouse had to make many decisions, often involving high stakes, without the benefit of 10,000 trials of practice.

3.4.4 Sudden insight

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

From the early days of animal learning experiments there have been warnings against the tendency to average learning curves across subjects [66, 133]. The average of many discontinuous curves will certainly look continuous and incremental, but that reassuring shape may miss the essence of the learning process. A recent reanalysis of many Pavlovian conditioning experiments suggested that discontinuous steps in performance are the rule rather than the exception [86]. Here we found that the same applies to navigation in a complex labyrinth. While the average learning curve presents like a continuous function (Figure 3.3B), the individual records of water rewards show that each animal improves rather quickly but at different times (Figure 3.3A).

Owing to the unstructured nature of the experiment, the mouse may adopt different policies for getting to the water port. In at least half of the animals, we observed a discontinuous change in that policy, namely when the animal started using efficient direct paths within the maze (Figure 3.5, Figure 3.5–Figure 3.20). This second switch happened considerably after the animal started collecting rewards, and did not greatly affect the reward rate. Furthermore, the animals never reverted to the less efficient policy, just as a child rarely unlearns to balance a bicycle.

Presumably this switch in performance reflects some discontinuous change in the animal's internal model of the maze, what Tolman called the "cognitive map" [18, 250]. In the unrewarded animals, we could not detect any discontinuous change in the use of long paths. However, as Tolman argued, those animals may well acquire a sophisticated cognitive map that reveals itself only when presented with a concrete task, like finding water. Future experiments will need to address this. The discontinuous changes in performance pose a challenge to conventional models of reinforcement learning, in which reward events are the primary driver of learning and each event contributes an infinitesimal update to the action policy. It will also be important to model the acquisition of distinct kinds of knowledge that contribute to the same behavior, like the location of the target and efficient routes to approach it.

3.4.5 Exploratory behavior

By all accounts, the animals spent a large fraction of the night exploring the maze (Figure 3.1–Figure 3.13). The water-deprived animals continued their forays into the depths of the maze long after they had found the water port and learned to exploit it regularly. After consuming a water reward, they wandered off into the maze 90% of the time (Figure 3.7B) instead of lazily waiting in front of the port during the timeout period. The sated animals experienced no overt reward from the maze, yet they likewise spent nearly half their time exploring that environment. As has been noted many times, animals—like humans—derive some form of intrinsic reward from exploration [25]. Some have suggested that there exists a homeostatic drive akin to hunger and thirst that elicits the information-seeking activity, and that the drive is in turn sated by the act of exploration [111]. If this were the case, then the drive to explore should be weakest just after an episode of exploration, much as the drive for food-seeking is weaker after a big meal.

Our observations are in conflict with this notion. The animal is most likely to enter the maze within the first minute of its return to the cage (Figure 3.1–Figure 3.14), a strong trend that runs opposite to the prediction from satiation of curiosity. Several possible explanations come to mind: (1) On these very brief visits to the cage, the animal may just want to certify that the exit route to the safe environment still exists, before continuing with exploration of the maze. (2) The temporal contrast between the boredom of the cage and the mystery of the maze is highest right at the moment of exit from the maze, and that may exert pressure to re-enter the maze. Understanding this in more detail will require dedicated experiments. For example, one could deliberately deprive the animals of access to the maze for some hours, and test whether that results in an increased drive to explore, as observed for other homeostatic drives around eating, drinking, and sleeping.

When left to their own devices, mice choose to spend much of their time engaged in exploration. One wonders how that affects their actions when they are strapped into a rigid behavior machine, like a 2AFC choice box. Presumably the drive to explore persists, perhaps more so because the forced environment is so unpleasant. And within the confines of the two alternatives, the only act of exploration the mouse has left is to give the wrong answer. This would manifest as an unexpectedly high error rate on unambiguous stimuli, sometimes called the “lapse rate” [48, 189]. The fact that the lapse rate decreases only gradually over weeks to months of training [46] suggests that it is difficult to crush the animal’s drive to explore.

The animals in our experiments had never been presented with a maze environment, yet they quickly settled into a steady mode of exploration. Once a mouse progressed beyond the first intersection it typically entered deep into the maze to one or more end nodes (Figure 3.6). Within 50 s of the first entry the animals adopted a steady speed of locomotion that they would retain throughout the night (Figure 3.2–Figure 3.18). Within 250 s of first contact with the maze the average animal already spent 50% of its time there (Figure 3.1–Figure 3.13). Contrast this with a recent study of “free exploration” in an exposed arena: Those animals required several hours before they even completed one walk around the perimeter [80]. Here the drive to explore is clearly pitted against fear of the open space, which may not be conducive to observing exploration *per se*.

The persistence of exploration throughout the entire duration of the experiment suggests that the animals are continuously surveying the environment, perhaps expecting new features to arise. These surveys are quite efficient: The animals cover all parts of the maze much faster than expected from a random walk (Figure 3.8). Effectively they avoid re-entering territory they surveyed just recently. It is often assumed that this requires some global memory of places visited in the environment [175, 184]. Such memory would have to persist for a long time: Surveying half of the available end nodes typically required 450 turning decisions. However, we found that a global long-term memory is not needed to explain the efficient search. The animals seem to be governed by a set of local turning biases that require memory only of the most recent decision and no knowledge of location (Figure 3.9). These local biases alone can explain most of the character of exploration without any global understanding or long-term memory. Incidentally, they also explain other seemingly global aspects of the behavior, for example the systematic preference that the mice have for the outer rather than the inner regions of the maze (Figure 3.10). Of course, this argument does not exclude the presence of a long-term memory, which may reveal itself in some other feature of the behavior.

Perhaps the most remarkable aspect of these biases is how similar they are across all 19 mice studied here, regardless of whether the animal experienced water rewards or not (Figure 3.9B, Figure 3.9–Figure 3.23), and independent of the sex of the mouse. The four decision probabilities were identical across individuals to within a standard deviation of <0.03. We cannot think of a trivial reason why this should be so. For example the two biases for forward motion (Figure 3.9B left) are poised halfway between the value for a random walk ($p = 2/3$) and certainty ($p = 1$). At either of those extremes, simple saturation might lead to a reproducible value, but not in the middle of the range. Why do different animals follow the exact same decision rules at an intersection between tunnels? Given that tunnel systems are part of the mouse's natural ecology, it is possible that those rules are innate and determined genetically. Indeed the rules by which mice build tunnels have a strong genetic component [261], so the rules for using tunnels may be written in the genes as well. The high precision with which one can measure those behaviors even in a single night of activity opens the way to efficient comparisons across genotypes, and also across animals with different developmental experience.

Finally, after mice discover the water port and learn to access it from many different points in the maze (Figure 3.5), they are presumably eager to discover other things. In ongoing work, we installed three water ports (visible in the videos accompanying this article) and implemented a rule that activates the three ports in a cyclic sequence. Mice discovered all three ports rapidly and learned to visit them in the correct order. Future experiments will have to raise the bar on what the mice are expected to learn in a night.

3.4.6 Mechanisms of navigation

How do the animals navigate when they perform direct paths to the water port or to the exit? The present study cannot resolve that, but one can gain some clues based on observations so far. Early workers already concluded that rodents in a maze will use whatever sensory cues and tricks are available to accomplish their tasks [173]. Our maze was designed to restrict those options somewhat.

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

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

The role of odors deserves particular attention because the mouse may use them both passively and actively. Does the animal first find the water port by following the smell of water? Probably not. For one, the port only emits a single drop of water when triggered by a nose poke. Second, we observed many instances where the animal is in the final corridor adjacent to the water port yet fails to discover it. The initial discovery seems to occur via touch. The reader can verify this in the videos accompanying this article. Regarding active use of odor markings in the maze, the maze rotation experiment suggests that such cues are not required for navigation, at least once the animals have adopted the shortest path to the water port (Figure 3.4).

Another algorithm that is often invoked for animals moving in an open arena is vector-based navigation [262]. Once the animal discovers a target, it keeps track of that target's heading and distance using a path integrator. When it needs to return to the target it follows the heading vector and updates heading and distance until it arrives. Such a strategy has limited appeal inside a labyrinth because the vectors are constantly blocked by walls. Consider, for example, the "home runs" back to the exit at the end of a bout. Here the target, namely the exit, is known from the start of the bout, because the animal enters through the same hole. At the end of the bout, when the mouse decides to exit from the maze, can it follow the heading vector to the exit? Figure 3.6A shows the 13 locations from which mice returned in a direct path to the exit on their very first foray. None of these locations is compatible with heading-based navigation: In each case an animal following the heading to the exit would get stuck in a different end node first and would have to reverse from there, quite unlike what really happened.

Finally, a partial clue comes from errors the animals make. We found that the rotation image of the water port, an end node diametrically across the entire maze, is one of the most popular destinations for rewarded animals (Figure 3.10A). These errors would be highly unexpected if the animals navigated from the entrance to the water by odor markings, or if they used an absolute representation of heading and distance. On the other hand, if the animal navigates via a remembered sequence of turns, then it will end up at that image node if it makes a single mistake at just the first T-junction.

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

3.5 Methods and materials

3.5.1 Experimental design

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

3.5.2 Maze construction

The maze measured ~24 x 24 x 2 inches (we used materials specified in inches, so dimensions are quoted in those non-SI units where appropriate). The ceiling was made of 0.5 inch clear acrylic. Slots of 1/8 inch width were cut into this plate on a 1.5 inch grid. Pegged walls made of 1/8 inch infrared-transmitting acrylic (opaque in the visible spectrum, ePlastics) were inserted into these slots and secured with a small amount of hot glue. The floor was a sheet of infrared-transmitting acrylic, supported by a thicker sheet of clear acrylic. The resulting corridors (1-1/8 inches wide) formed a 6-level binary tree with T-junctions and progressive shortening of each branch, ranging from ~12 inch to 1.5 inch (Figure 3.1 and Figure 3.2). A single end node contained a 1.5 cm circular opening with a water delivery port (described below). The maze included provision for two additional water ports not used in the present report. Once per week, the maze was submerged in cage cleaning solution. Between different animals, the floor and walls were cleaned with ethanol.

3.5.3 Reward delivery system

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

3.5.4 Cage and connecting passage

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

3.5.5 Animals and treatments

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

3.5.6 Video recording

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

3.5.7 Animal tracking

A version of DeepLabCut [176] modified to support gray-scale processing was used to track the animal's trajectory, using key points at the nose, feet, tail base and mid-body. All subsequent analysis was based on the trajectory of the animal's nose, consisting of positions $x(t)$ and $y(t)$ in every video frame.

3.5.8 Rates of transition between cage and maze

This section relates to Figure 3.1–Figure 3.14. We entertained the hypothesis that the animals become "thirsty for exploration" as they spend more time in the cage. In that case one would predict that the probability of entering the maze in the next second will increase with time spent in the cage. One can compute this probability from the distribution of residency times in the cage, as follows:

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

$$p\left(t\right) = e^{-\int_0^t r(t')dt'} r\left(t\right) \tag{3.2}$$

where $r\left(t\right)$ is the instantaneous rate for entering the maze. So

$$\int_0^t p\left(t'\right) dt' = 1 - e^{-\int_0^t r(t')dt'} \tag{3.3}$$

$$\int_0^t r\left(t'\right) dt' = -\ln\left(1 - \int_0^t p\left(t'\right) dt'\right). \tag{3.4}$$

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

$$r_{\mathrm{m}}\left(t\right) = \text{rate of entry into the maze as a function of time spent in the cage} \tag{3.5}$$

$$r_{\mathrm{c}}\left(t\right) = \text{rate of entry into the cage as a function of time spent in the maze.} \tag{3.6}$$

The rate of entering the maze is highest at short times in the cage (Figure 3.1–Figure 3.14A). It peaks after ~15 s in the cage and then declines gradually by a factor of 4 over the first minute. So the mouse is most likely to enter the maze just after it returns from there. This runs opposite to the expectation from a homeostatic drive for exploration, which should be sated right after the animal returns. We found no evidence for an increase in the rate at late times. These effects were very similar in rewarded and unrewarded groups and in fact the tendency to return early was seen in every animal.

By contrast, the rate of exiting the maze is almost perfectly constant over time (Figure 3.1–Figure 3.14B). In other words the exit from the maze appears like a constant rate Poisson process. There is a slight elevation of the rate at short times among rewarded animals (Figure 3.1–Figure 3.14B top). This may come from the occasional brief water runs they perform. Another strange deviation is an unusual number of very short bouts (duration 2-12 s) among unrewarded animals (Figure 3.1–Figure 3.14B bottom). These are brief excursions in which the animal runs to the central junction, turns around, and runs to the exit. Several animals exhibited these, often several bouts in a row, and at all times of the night.

3.5.9 Reduced trajectories

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

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

In Figure 3.5 and Figure 3.6, we count the occurrence of **direct paths** leading to the water port (a "water run") or to the exit (a "home run"). A direct path is a node sequence without any reversals. Figure 3.3–Figure 3.16 illustrates some examples.

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

$$\text{Errors} = (\text{Length of path} - \text{Length of direct path})/2. \tag{3.7}$$

3.5.10 Maze rotation

The maze rotation experiment (Figure 3.4) was performed on 4 mice, all water-deprived. Two of the animals ('D7' and 'D9') had experienced the maze before, and are part of the 'rewarded' group in other sections of the report. Two additional animals ('F2' and 'A1') had had no prior contact with the maze.

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

For animals 'D7' and 'D9,' we rotated only the floor of the maze, leaving the walls and ceiling in the original configuration. For 'F2' and 'A1,' we rotated the entire maze, moving one wall segment at the central junction and the water port to attain the same shape. Navigation remained intact for all animals. Note that 'A1' performed a perfect path to the water port and back immediately before and after a full maze rotation (Figure 3.4B).

The visits to the 4 locations in the maze (Figure 3.4C, Figure 3.4–Figure 3.17) were limited to direct paths of length at least 2 steps. This avoids counting rapid flickers between two adjacent nodes. In other words, the animal has to move at least 2 steps away from the target node before another visit qualifies.

3.5.11 Statistics of sudden insight

In Figure 3.5, one can distinguish two events: First, the animal finds the water port and begins to collect rewards at a steady rate: this is when the green curve rises up. At a later time, the long direct paths to the water port become much more frequent than to the comparable control nodes: this is when the red and blue curves diverge. For almost all animals, these two events are well separated in time (Figure 3.5–Figure 3.19). In many cases, the rate of long paths seems to change discontinuously: a sudden change in slope of the curve.

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

$$r\left(t\right) = r_{\mathrm{i}} + \frac{r_{\mathrm{f}} - r_{\mathrm{i}}}{2} \operatorname{erf}\left(\frac{t - t_{\mathrm{s}}}{w}\right) \tag{3.8}$$

where

$$\operatorname{erf}\left(x\right) = \frac{2}{\sqrt{\pi}} \int\limits_{0}^{x} e^{-x^2} \mathrm{d}x.$$

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

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

$$L\left[r\left(t\right)\right] = e^{-\int\limits_0^T r(t)dt} \prod_i r\left(t_i\right) \tag{3.9}$$

and the log likelihood is

$$\ln L = \sum_i \ln r\left(t_i\right) - \int\limits_0^T r\left(t\right) dt. \tag{3.10}$$

For each of the 10 rewarded mice, we maximized $\ln L$ over the 4 parameters of the rate model, both for the reward events and the long paths to water. The resulting fits are plotted in Figure 3.5–Figure 3.19.

Focusing on the learning of long paths to water, for 6 of the 10 animals the optimal width parameter w was less than 300 s: B1, B2, C1, C3, C6, C7. These are the same animals one would credit with a sudden kink in the cumulative event count based on visual inspection (Figure 3.5–Figure 3.19).

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

$$r\left(t\right) = \begin{cases} r_{\mathrm{i}}, t < t_{\mathrm{s}} \\ r_f, t > t_{\mathrm{s}} \end{cases} \tag{3.11}$$

and computed the likelihood of the data as a function of the step time t_{s}. We report the mean and standard deviation of the step time over its likelihood in Figure 3.5–Figure 3.20. Animal C6 was dropped from this "sudden step" group, because the uncertainty in the step time was too large (~900 s).

3.5.12 Efficiency of exploration

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

We restricted the animal's node trajectory (s_i) to clips of exploration mode, excluding the direct paths to the water port or the exit. All subsequent steps were applied to these clips, then averaged over clips. Within each clip, we marked the sequence of end nodes (e_i). We slid a window of size n across this sequence and counted the number of distinct nodes d in each window. Then we averaged d over all windows in all clips. Then we repeated that for a wide range of n. The resulting $d(n)$ is plotted in the figures reporting new nodes vs nodes visited (Figure 3.8A,B and Figure 3.9C).

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

$$d(n) \approx 64 \left(1 - \frac{1}{1 + \frac{z + bz^3}{1+b}} \right) \tag{3.12}$$

where

$$z = n \,/ a \ . \tag{3.13}$$

The parameter a is the number of visits n required to survey half of the end nodes, whereas b reflects a relative acceleration in discovering the last few end nodes. This function was found by trial and error and produces absurdly good fits to the data (Figure 3.8–Figure 3.22). The values quoted in the text for efficiency of exploration are $E = 32 \,/ a$ (Equation 3.1).

The value of b was generally small (~0.1) with no difference between rewarded and unrewarded animals. It declined slightly over the night (Figure 3.8–Figure 3.22B), along with the decline in a (Figure 3.8C).

3.5.13 Biased random walk

For the analysis of Figure 3.9, we considered only the parts of the trajectory during 'exploration' mode. Then we parsed every step between two nodes in terms of the type of action it represents. Note that every link between nodes in the maze is either a 'left branch' or a 'right branch,' depending on its relationship to the parent T-junction.

Therefore there are 4 kinds of action:

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

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

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

$$P_{\mathrm{SF}} = \frac{p(0|0) + p(0|1) + p(1|0) + p(1|1)}{p(0|0) + p(0|1) + p(1|0) + p(1|1) + p(2|0) + p(3|1)} \tag{3.14}$$

$$P_{\mathrm{SA}} = \frac{p(0|1) + p(1|0)}{p(0|0) + p(0|1) + p(1|0) + p(1|1)} \tag{3.15}$$

$$P_{\mathrm{BF}} = \frac{p(0|3) + p(1|2) + p(2|2) + p(2|3) + p(3|2) + p(3|3)}{p(0|3) + p(1|2) + p(2|2) + p(2|3) + p(3|2) + p(3|3) + p(0|2) + p(1|3)} \tag{3.16}$$

$$P_{\mathrm{BS}} = \frac{p(2|2) + p(2|3) + p(3|2) + p(3|3)}{p(0|3) + p(1|2) + p(2|2) + p(2|3) + p(3|2) + p(3|3)}. \tag{3.17}$$

For the simulations of random agents (Figure 3.8, Figure 3.9), we used trajectories long enough so the uncertainty in the resulting curves was smaller than the line width.

3.5.14 Models of decisions during exploration

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

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

Fixed depth Markov chain

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

$$n(\mathbf{h}, a) = \text{number of times history } \mathbf{h} \text{ leads to action } a. \tag{3.18}$$

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

$$p\left(a\,|\mathbf{h}\right) = \frac{n\left(\mathbf{h}, a\right) + 1}{\sum\limits_{a'} n\left(\mathbf{h}, a'\right) + 3}. \tag{3.19}$$

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

$$H = \left\langle -\log_2 p\left(a_t\,|\mathbf{h}_t\right)\right\rangle_t. \tag{3.20}$$

Variable depth Markov chain

As one pushes to longer histories, i.e. larger k, the analysis quickly becomes data-limited, because the number of possible histories grows exponentially with k. Soon one finds that the counts for each history-action combination drop to where one can no longer estimate probabilities correctly. In an attempt to offset this problem, we pruned the history tree such that each surviving branch had more than some minimal number of counts in the training data. As expected, this model is less prone to over-fitting and degrades more gently as one extends to longer histories (Figure 3.11–Figure 3.24A). The lowest cross-entropy was obtained with an average history length of ~4.0 but including some paths of up to length 6. Of all the algorithms we tested, this produced the lowest cross-entropies, although the gains relative to the fixed-depth model were modest (Figure 3.11–Figure 3.24C).

Pooling across symmetric nodes in the maze

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

However, one may want to distinguish two types of junctions within a given level: L-nodes are reached by a left branch from their parent junction one level lower in the tree, R-nodes by a right branch. For example, in Figure 3.3–Figure 3.16, node 1 is L-type and node 2 is R-type. When we pooled histories over all the L-nodes at a given level and separately over all the R-nodes the cross-entropy indeed dropped, by about 5% on average. This pooling greatly reduced the amount of over-fitting (Figure 3.11–Figure 3.24B), which allowed the use of longer histories, which in turn improved the predictions on test data. The benefit of distinguishing L- and R-nodes probably relates to the animal's tendency to alternate left and right turns.

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

3.5.15 Data and code availability

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

3.6 Figure supplements

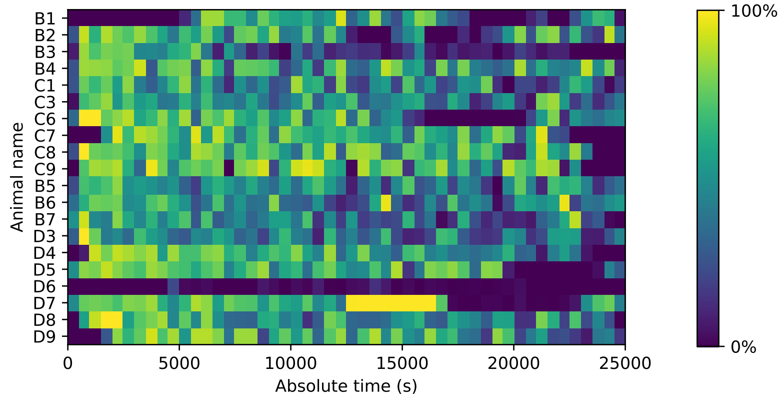


Figure 3.12: **Occupancy of the maze.**
Fraction of time spent in the maze. Mice could move freely between the home cage and the maze. For each animal (vertical), the fraction of time in the maze (color scale) is plotted as a function of time since start of the experiment. Time bins are 500 s. Note that mouse D6 hardly entered the maze; it never progressed beyond the first junction. This animal was excluded from all subsequent analysis steps.

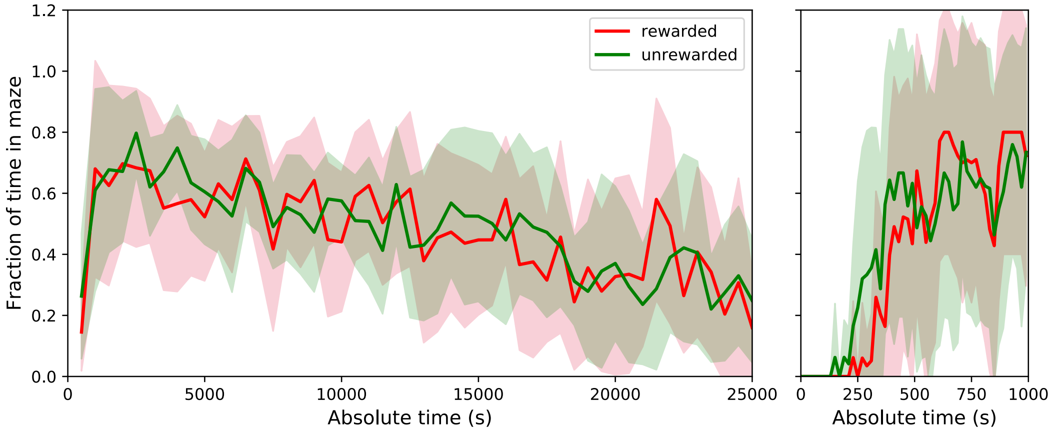


Figure 3.13: **Fraction of time in maze by group.**
Average fraction of time spent in the maze by group. This shows the average fraction of time in the maze as Mean ± SD over the population of 10 rewarded and 9 unrewarded animals. Right: expanded axis for early times. The tunnel to the maze opens at 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 ~70%, then declined gradually over 7 h to ~30%.

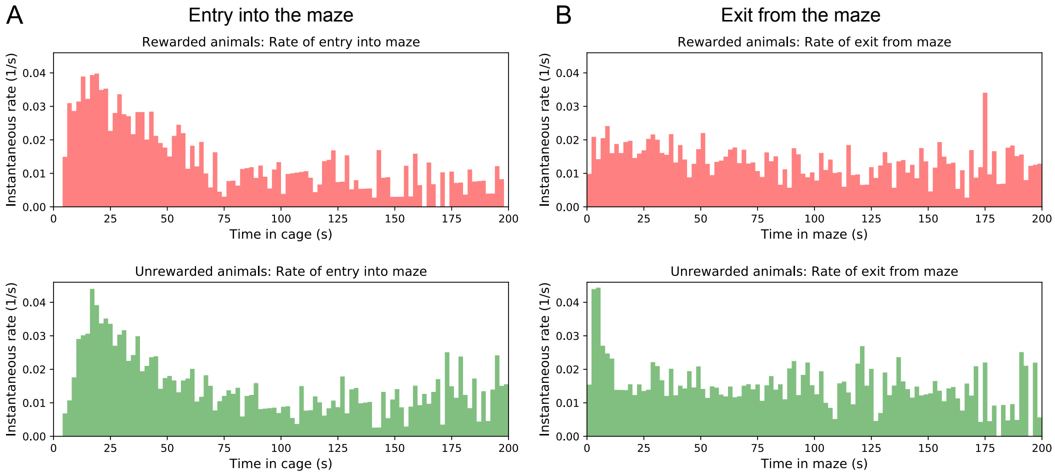


Figure 3.14: **Transitions between cage and maze.**
Rates of transition between cage and maze. (A) The instantaneous probability per unit time $r_{\mathrm{m}}(t)$ of entering the maze after having spent time t in the cage. Note, this rate is highest immediately upon entering the cage, then declines by a large factor. (B) The instantaneous probability per unit time $r_{\mathrm{c}}(t)$ of exiting the maze after having spent time t in the maze.

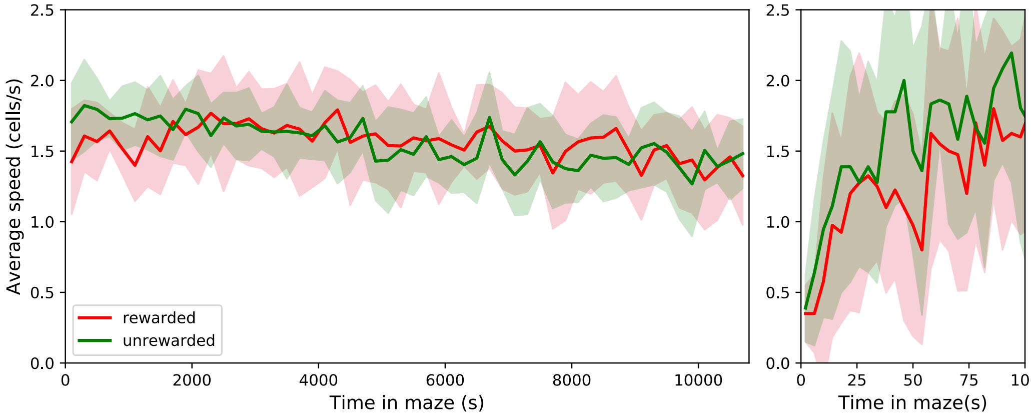


Figure 3.15: **Speed of locomotion.**
The speed of locomotion in the maze is approximately constant. Left: Speed plotted as Mean ± SD over the population of rewarded and unrewarded animals. Right: expanded axis for early times. To assess the speed of locomotion we divided the maze into square cells as 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, ~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.

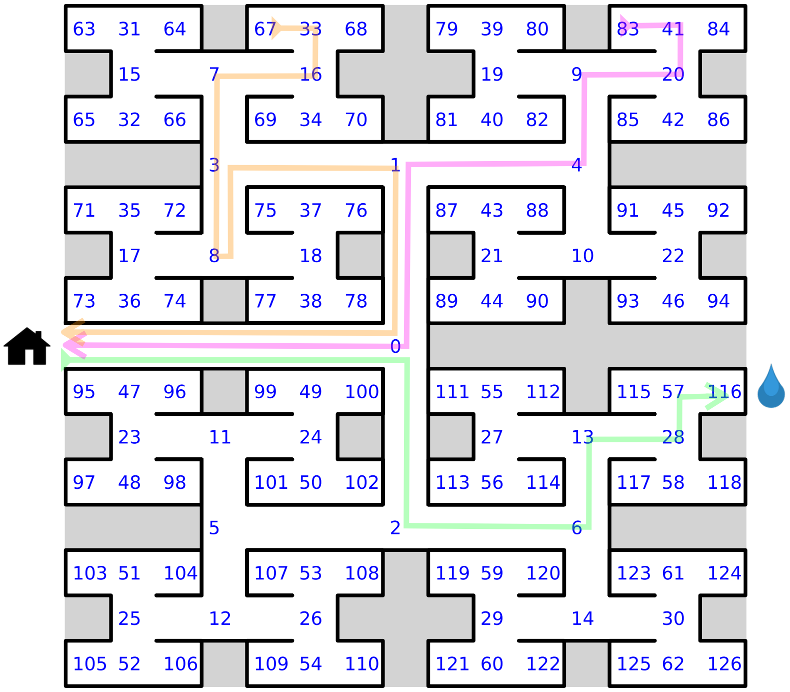


Figure 3.16: **Definition of node trajectories.**
Definition of node trajectories. A numbering scheme for all 127 nodes of the maze. Green: a direct path from the entrance to the water port ("water run") with the node sequence $(s_i) = (0, 2, 6, 13, 28, 57, 116)$, involving 6 decisions. Magenta: a direct 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 $(8, 3, 1, 0)$.

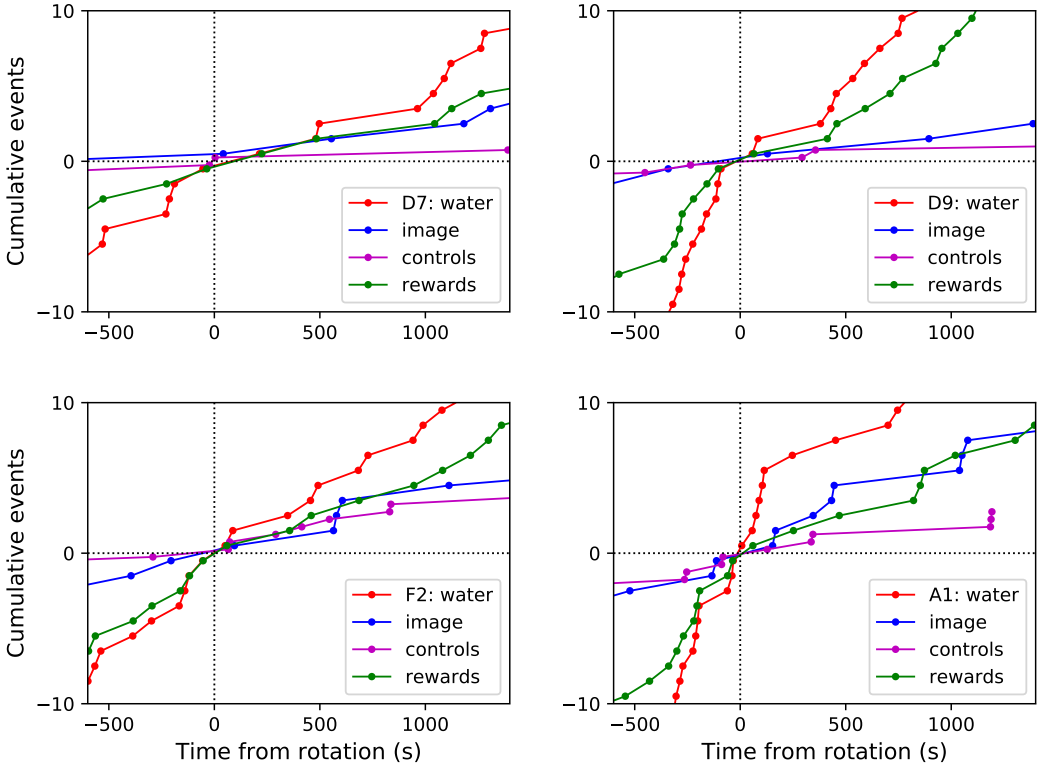


Figure 3.17: **Navigation before and after maze rotation for each animal.**
Navigation before and after maze rotation. Cumulative number of rewards, visits to the water port, the image of the water port, and the control nodes, plotted vs time before and after the maze rotation. Display as in Figure 3.4C, but split for each of 4 animals.

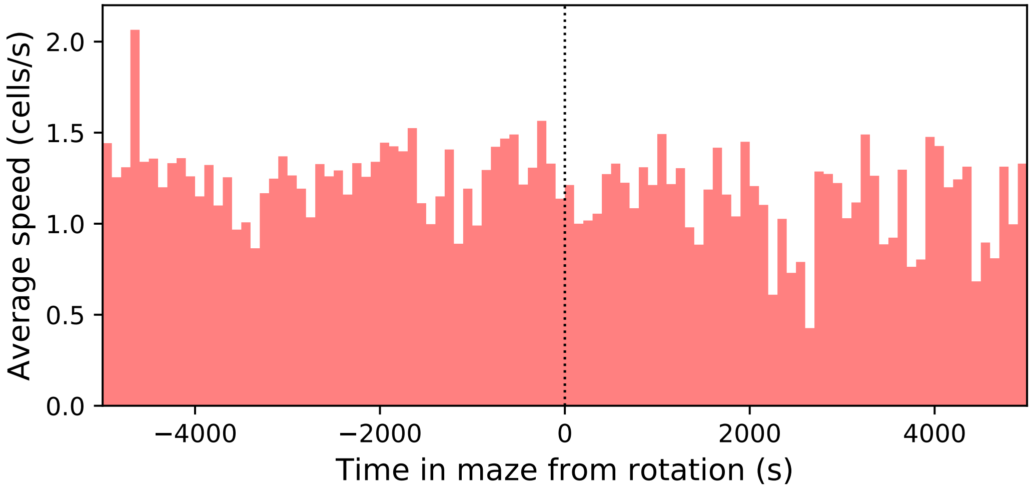


Figure 3.18: **Speed before and after maze rotation.**
Speed of the mouse vs time in the maze. Average over 4 animals. Time is plotted relative to the maze rotation.

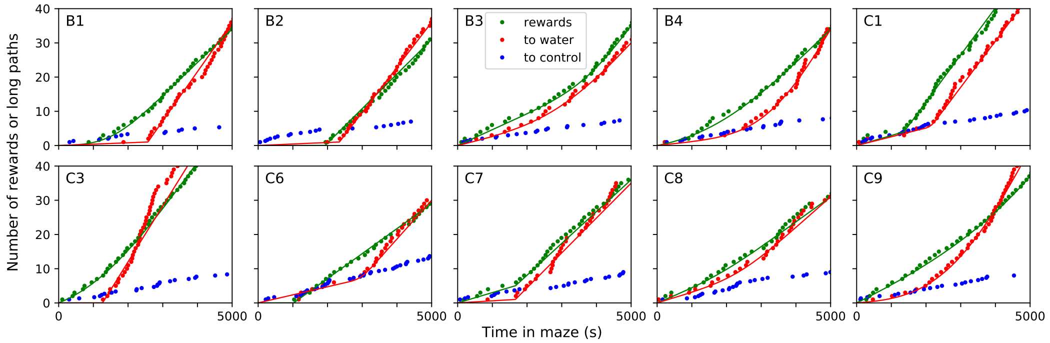
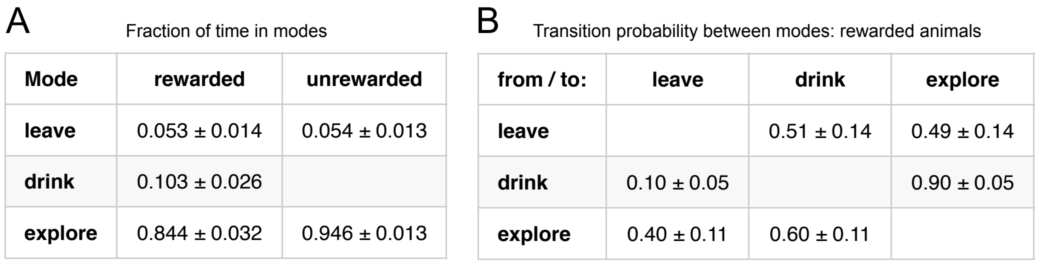


Figure 3.19: **Long direct paths for all animals.**
Sudden changes in behavior for all rewarded animals. For each of the 10 water-deprived animals this shows the cumulative rate of rewards, of long direct paths (>6 steps) to the water port, and of similar paths to 3 control nodes. Display as in Figure 3.5; 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 ± 110	36.4
B2	2350 ± 220	30.3
C1	2070 ± 310	5.49
C3	1280 ± 80	1640
C7	1680 ± 280	16.9

Figure 3.20: **Statistics of sudden changes in behavior.**
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

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

B Transition probability between modes: rewarded animals

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

Figure 3.21: **Three modes of behavior.**
Three modes of behavior. (A) The fraction of time mice spent in each of the three modes while in the maze. Mean ± SD for 10 rewarded and 9 unrewarded animals. (B) Probability of transitioning from the mode on the left to the mode at the top. Transitions from ‘leave’ represent what the animal does at the start of the next bout into the maze.

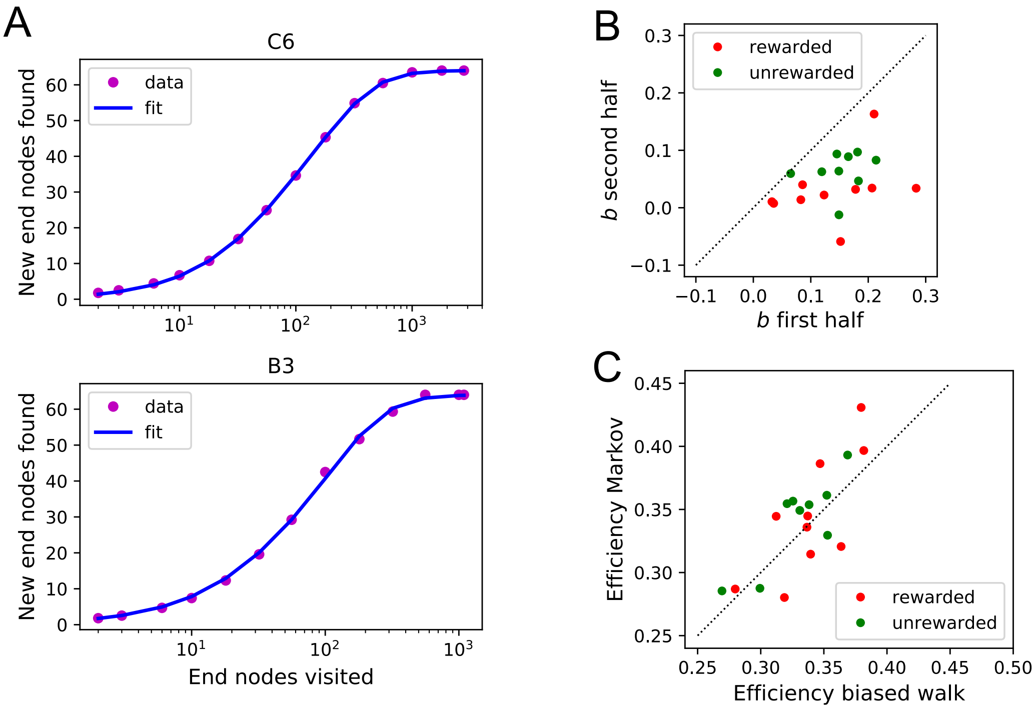


Figure 3.22: **Efficiency of exploration.**
Functional fits to measure exploration efficiency. (A) Fitting Equation 3.12 to the data from the mouse's exploration. Animals with best fit (top) and worst fit (bottom). The relative uncertainty in the two fit parameters a and b was only 0.0038 ± 0.0020 (mean ± SD across animals). (B) The fit parameter b for all animals, comparing the first to the second half of the night. (C) The efficiency E (Equation 3.1) predicted from two models of the mouse's trajectory: The 4-bias random walk (Figure 3.11D) and the optimal Markov chain (Figure 3.11C).

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

Figure 3.23: **Bias statistics.**
Statistics of the four turning biases. Mean and standard deviation of the 4 biases of Figure 3.9A-B across animals in the rewarded and unrewarded groups.

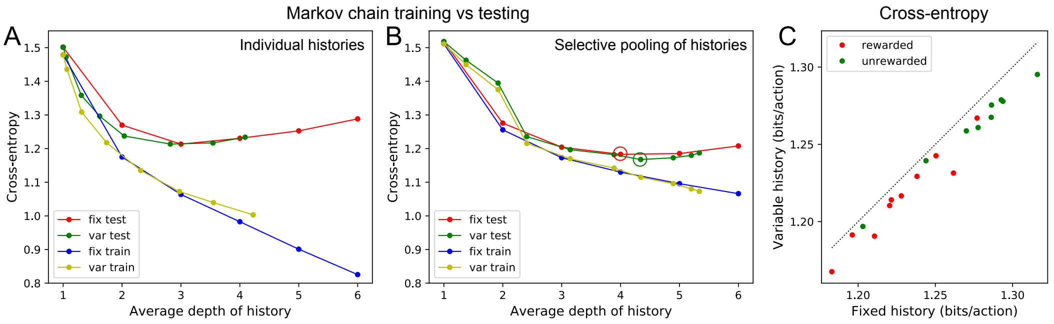


Figure 3.24: **Markov model fits.**
Fitting Markov models of behavior. (A) Results of fitting the node sequence of a single animal (C3) with Markov models having a fixed depth ('fix') or variable depth ('var'). The cross-entropy of the model's prediction is plotted as a function of the average depth 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.

3.7 Acknowledgments

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

Competing interests: The authors declare no competing interests.

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

Chapter 4

A SIMPLE CIRCUIT MODEL OF COGNITIVE MAPPING

Endotaxis: A Universal Algorithm for Mapping, Goal-Learning, and Navigation

Tony Zhang, Matthew Rosenberg, Pietro Perona, and Markus Meister

bioRxiv, September 25, 2021

4.1 Abstract

An animal entering a new environment typically faces three challenges: explore the space for resources, memorize their locations, and navigate towards those targets as needed. Experimental work on exploration, mapping, and navigation has mostly focused on simple environments—such as an open arena, a pond [168], or a desert [174]—and much has been learned about neural signals in diverse brain areas under these conditions [53, 227]. However, many natural environments are highly constrained, such as a system of burrows or of paths through the underbrush. More generally, many cognitive tasks are equally constrained, allowing only a small set of actions at any given stage in the process. Here we propose an algorithm that learns the structure of an arbitrary environment, discovers useful targets during exploration, and navigates back to those targets by the shortest path. It makes use of a behavioral module common to all motile animals, namely the ability to follow an odor to its source [14]. We show how the brain can learn to generate internal "virtual odors" that guide the animal to any location of interest. This *endotaxis* algorithm can be implemented with a simple 3-layer neural circuit using only biologically realistic structures and learning rules. Several neural components of this scheme are found in brains from insects to humans. Nature may have evolved a general mechanism for search and navigation on the ancient backbone of chemotaxis.

4.2 Introduction

Efficient navigation requires knowing the structure of the environment: which locations are connected to which others [250]. One would like to understand how the brain acquires that knowledge, what neural representation it adopts for the resulting map, how it tags significant locations in that map, and how that knowledge gets read out for decision-making during navigation. Here we propose a mechanism that solves all these problems and operates reliably in diverse and complex environments.

One algorithm for finding a valuable resource is common to all animals: chemotaxis. Every motile species has a way to track odors through the environment, either to find the source of the odor or to avoid it [14]. This ability is central to finding food, connecting with a mate, and avoiding predators. It is believed that brains originally evolved to organize the motor response in pursuit of chemical stimuli. Indeed some of the oldest regions of the mammalian brain, including the hippocampus, seem organized around an axis that processes smells [2, 115].

The specifics of chemotaxis, namely the methods for finding an odor and tracking it, vary by species, but the toolkit always includes a random trial-and-error scheme: Try various actions that you have available, then settle on the one that makes the odor stronger [14]. For example, a rodent will weave its head side-to-side, sampling the local odor gradient, then move in the direction where the smell is stronger. Worms and maggots follow the same strategy. Dogs track a ground-borne odor trail by casting across it side-to-side. Flying insects perform similar casting flights. Bacteria randomly change direction every now and then, and continue straight as long as the odor improves [23]. We propose that this universal behavioral module for chemotaxis can be harnessed to solve general problems of search and navigation in a complex environment.

For concreteness, consider a mouse exploring a labyrinth of tunnels (Fig 4.1A). The maze may contain a source of food that emits an odor (Fig 4.1A top). That odor will be strongest at the source and decline with distance along the tunnels of the maze. The mouse can navigate to the food location by simply following the odor gradient uphill. Suppose that the mouse discovers some other interesting locations that do not emit a smell, like a source of water, or the exit from the labyrinth (Fig 4.1A). It would be convenient if the mouse could tag such a location with an odorous material, so it may be found easily on future occasions. Ideally the mouse would carry with it multiple such odor tags, so it can mark different targets each with its specific recognizable odor (Fig 4.1A mid and bottom).

Here we show that such tagging does not need to be physical. Instead we propose a mechanism by which the mouse's brain may compute a "virtual odor" signal that declines with distance from a chosen target. That neural signal can be made available to the chemotaxis module as though it were a real odor, enabling navigation up the gradient towards the target. Because this goal signal is computed in the brain rather than sensed externally, we call this hypothetical process *endotaxis*.

4.3 A circuit to implement endotaxis

In Figure 4.1B, we present a neural circuit model that implements three goals: mapping the connectivity of the environment; tagging of goal locations with a virtual odor; and navigation towards those goals. The model includes four types of neurons: feature cells, point cells, map cells, and goal cells.

Feature cells: These cells fire when the animal encounters an interesting feature that may form a target for future navigation. Each feature cell is selective for a specific kind of resource, for example water or food, by virtue of sensory pathways that respond to those stimuli.

Point cells: This layer of cells represents the animal's location.[1] Each neuron in this population has a small response field within the environment. The neuron fires when the animal enters that response field. We assume that these point cells exist from the outset as soon as the animal enters the environment. Each cell's response field is defined by some conjunction of external and internal sensory signals at that location.

Map cells: This layer of neurons learns the structure of the environment, namely how the various locations are connected in space. The map cells get excitatory input from point cells with low convergence: Each map cell should collect input from only one or a few point cells. These input synapses are static. The map cells also excite each other with all-to-all connections. These recurrent synapses are modifiable according to rules of Hebbian plasticity and, after learning, represent the topology of the environment.

Goal cells: These neurons mark the locations of special resources in the map of the environment. A goal cell for a specific feature receives excitatory input from the corresponding feature cell. It also receives Hebbian excitatory synapses from map cells. Those synapses are strengthened when the presynaptic map cell is active at the same time as the feature cell. In addition, this plasticity may be gated by a reward signal associated with the value of the discovered resource.

Each of the goal cells carries a virtual odor signal for its assigned feature. That signal increases systematically as the animal moves closer to the target feature. A mode

[1] We avoid the term 'place cell' here because (1) that term has a technical meaning in the rodent hippocampus, whereas the arguments here extend to species that do not have a hippocampus; (2) all the cells in this network have a place field, but it is smallest for the point cells.

switch selects one among many possible virtual odors (or real odors) to be routed to the chemotaxis module for odor tracking.[2] The animal then pursues its chemotaxis search strategy to maximize that odor, which leads it to the selected tagged feature.

4.3.1 Why does the circuit work?

The key insight is that the output of the goal cell declines systematically with the distance of the animal from that target. This relationship holds even if the environment is a complex graph with constrained connectivity. Here we explain how this comes about, with mathematical details in the supplement.

As the animal explores a new environment, when it moves from one location to an adjacent one, those two point cells briefly fire together. That leads to a Hebbian strengthening of the excitatory synapses between the two corresponding map cells (Fig 4.2A-B). In this way the recurrent network of map cells learns the connectivity of the graph that describes the environment. To a first approximation, the matrix of synaptic connections among the map cells will converge to the correlation matrix of their inputs [60, 87], which in turn reflects the adjacency matrix of the graph (Eqn 4.21). Now the brain can use this adjacency information to find the shortest path to a target.

After this map learning, the output of the map network is a hump of activity, centered on the current location x of the animal and declining with distance along the various paths in the graph (Fig 4.2C). If the animal moves to a different location y, the map output is another hump of activity, now centered on y (Fig 4.2D). The overlap of the two hump-shaped profiles will be large if nodes x and y are close on the graph, and small if they are distant. Fundamentally the endotaxis network computes that overlap. How is it done?

Suppose the animal visits y and finds water there. Then the profile of map activity $v_i(y)$ gets stored in the synapses G_{gi} onto the goal cell g that responds to water (Fig 4.2D, Eqn 4.25). When the animal subsequently moves to a different location x, the goal cell g receives the current map output $v_i(x)$ filtered through the previously stored synaptic template $v_i(y)$ (Fig 4.2E). This is the desired measure of overlap (Eqn 4.26), and one can show mathematically that it declines exponentially with the shortest graph-distance between x and y (Eqn 4.27).

[2] That mode switch is controlled by the *murinculus*: a tiny mouse inside the mouse that tells the mouse what to do. We do not claim to know how that works.

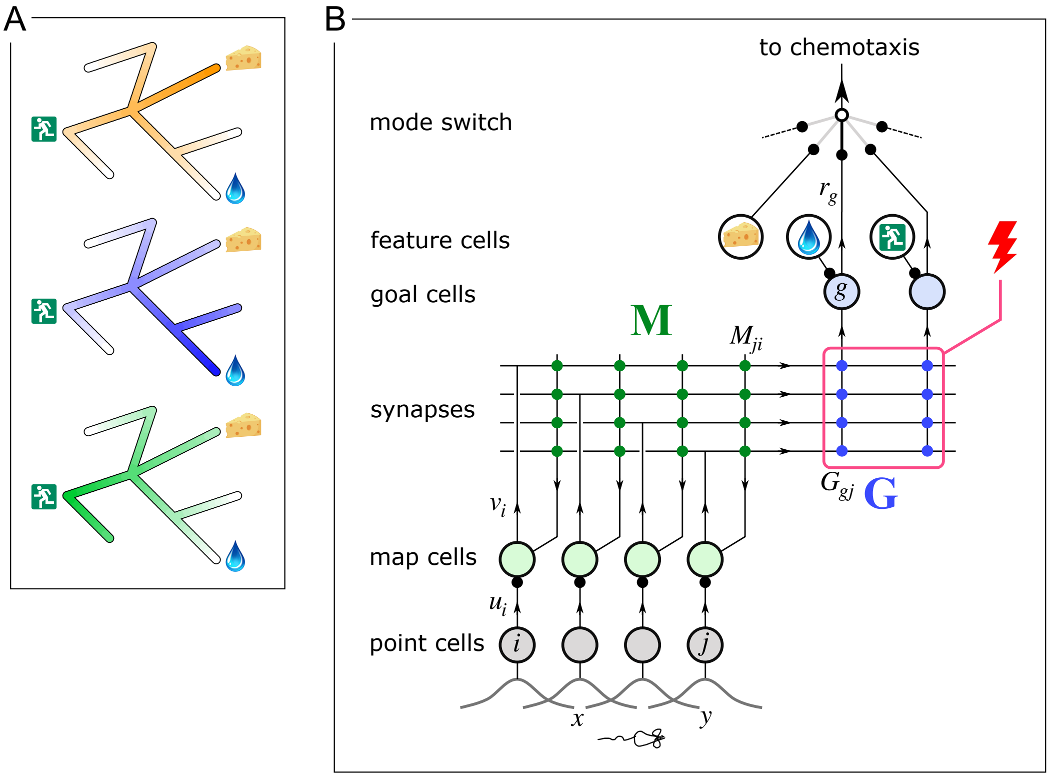


Figure 4.1: **A mechanism for endotaxis.**
(A) A constrained environment of nodes linked by straight corridors, with special locations offering food, water, and the exit. Top: A real odor emitted by the food source decreases with distance (shading). Middle: A virtual odor tagged to the water source. Bottom: A virtual odor tagged to the exit. (B) A neural circuit to implement endotaxis. Open circles: four populations of neurons that represent "feature," "point," "map," and "goal." Arrows: signal flow. Solid circles: synapses. Point cells have small receptive fields localized in the environment and excite map cells. Map cells excite each other (green synapses) and also excite goal cells (blue synapses). Feature cells signal the presence of a resource, e.g. cheese, water, or the exit. Map synapses and goal synapses are modified by Hebbian plasticity. Learning at goal synapses (blue) is gated by a reward signal (flash symbol). A "mode" switch selects among various goal signals depending on the animal's need. They may be virtual odors (water, exit) or real odors (cheese). The resulting signal gets fed to the chemotaxis module for gradient ascent. Mathematical symbols used in the text: u_i is the output of a point cell at location i, v_i is the output of the corresponding map cell, $\mathbf{M}$ is the matrix of synaptic weights among map cells, $\mathbf{G}$ are the synaptic weights from the map cells onto goal cells, and r_g is the output of goal cell g.

4.4 Performance of the endotaxis algorithm

Some important features of endotaxis can already be appreciated at this level of detail. First, the structure of the environment is acquired separately from the location of resources. The graph that connects different points in the environment is learned

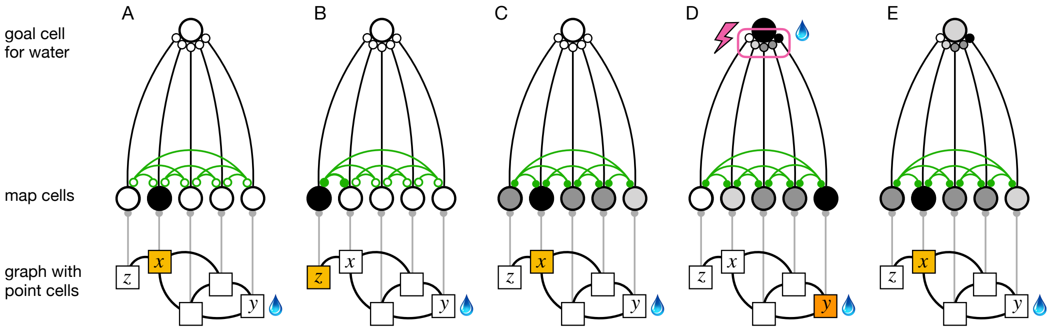


Figure 4.2: **The phases of endotaxis during exploration, goal-tagging, and navigation.**
A portion of the circuit in Figure 4.1 is shown, including a single goal cell that responds to the water feature. Bottom shows a graph of the environment, with the agent's current location shaded in orange. Each node has a point cell that reports the presence of the agent to a corresponding map cell. Map cells are recurrently connected (green) and feed convergent signals onto the goal cell. (A) Initially the recurrent synapses are weak. (B) During exploration the agent moves between adjacent locations on the graph, and that strengthens the connection between their corresponding map cells. (C) After exploration the map synapses reflect the connectivity of the graph. Now the map cells have an extended profile of activity, centered on the agent's current location x and decreasing from there with distance on the graph. (D) When the agent reaches the water source y the goal cell gets activated by the sensation of water. A reward signal (pink) potentiates the synapses from map cells in proportion to their current activity. Thus the state of the map at the water location gets stored in the goal synapses. This event represents tagging of the water location. (E) During navigation, as the agent visits different nodes, the map state gets filtered through the goal synapses to excite the goal cell. This produces a signal in the goal cell that declines with the agent's distance from the target.

by the synapses in the map network. By contrast, the location of special goals within that map is learned by the synapses onto the goal cells. The animal can explore and learn the environment regardless of the presence of threats or resources. Once a resource is found, its location can be tagged immediately within the existing map structure. If the distribution of resources changes, the knowledge of the connectivity map remains unaffected. Second, the endotaxis algorithm is "always on." There is no separation of learning and recall into different phases. Both the map network and the goal network get updated continuously based on the animal's trajectory through the environment, and the goal signals are always available for directed navigation via gradient ascent.

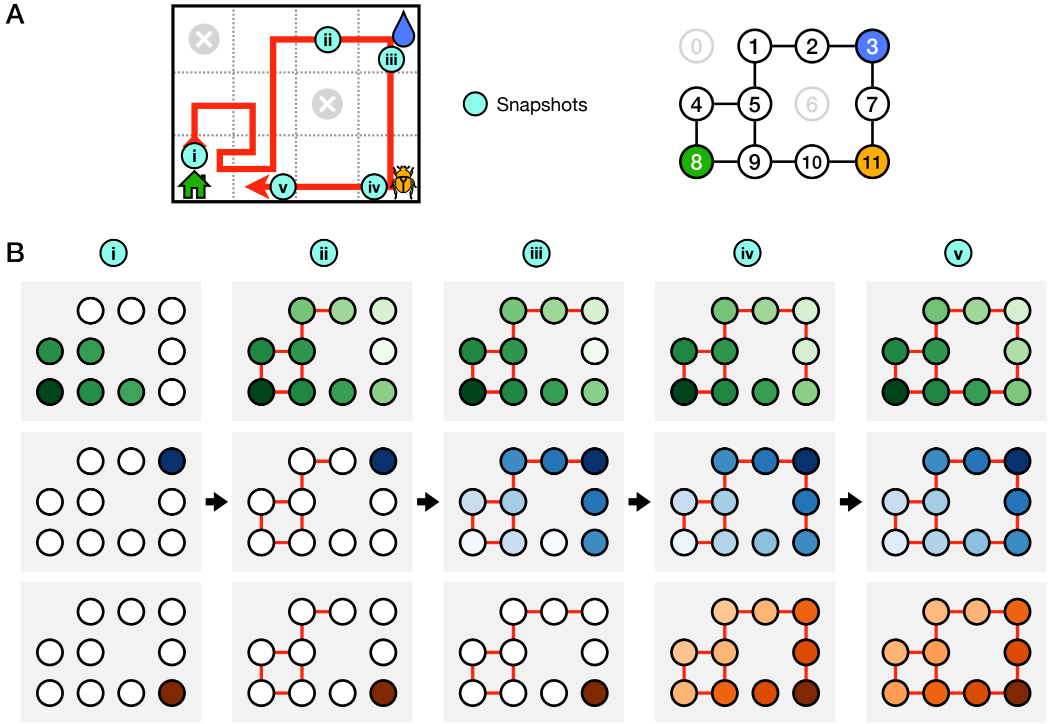


Figure 4.3: **The map and the targets are learned independently.**
(A) Left: an agent explores a simple Gridworld with 3 salient goal locations following the red trajectory. Space is discretized into square tiles, each tile represented by one point cell. Circles with crosses represent obstacles, namely tiles that are not reachable. Right: graph of this environment, where each tile becomes a node, and edges represent traversable connections between tiles. (B) The response fields of three goal neurons for home (top), water (middle), and bug (bottom) at the 5 instants during the learning process (i-v). Red edges connect previously visited nodes. The response (log color scale) is plotted at each location where the agent could be placed. The agent starts random walking from the entrance (i) and gradually discovers the other two goal locations (water at time iii, bug at time iv). Upon discovery of a goal location, the corresponding goal cell's signal is immediately useful in all previously visited locations (iii, iv) as well as nodes that are ≤ 2 steps away. Any new locations visited subsequently and nodes ≤ 2 steps away are also recruited into the goal cell's response field (v).

4.4.1 Simultaneous acquisition of map and targets during exploration

To illustrate these functions, and to explore capabilities that are less obvious from an analytical inspection, we simulated agents navigating by the endotaxis algorithm (Fig 4.1B) through a range of environments (Figs 4.3-4.4) that contained one or more desirable targets. In each case, we allowed the agent to explore by executing a random walk on the graph of the environment. Once the agent had covered the entire graph and discovered the targets, we tested its navigation ability. We placed the agent at an arbitrary node on the graph and let it navigate towards each of the targets following the corresponding virtual odor signal. Then we counted the number of steps the agent required, and compared that to the shortest possible path to the target. During this navigation the agent used a trial-and-error policy: it sampled the virtual odor at each neighboring node and stepped to the node with the highest value.

In simulating the circuit of Figure 4.1B, we assumed that there exist point cells that fire when the agent is at a specific location, owing to a match of their sensory receptive fields with features in the environment. The preferred locations of these point cells define the nodes of the graph that the agent will learn. Map cells transform their synaptic inputs with a nonlinear response function. At the outset, the agent has no knowledge about the topology of the environment or the locations of targets, so both the map synapses and the goal synapses start out with zero synaptic strengths. During exploration, the map synapses get updated based on the simultaneous firing of point cells at neighboring locations. We used a standard formulation of Hebbian learning, called Oja's rule. Similarly the synapses onto goal cells get updated when the agent reaches a target, based on the presynaptic map cell signal and the postsynaptic input from feature cells. Map cells and goal cells were allowed to learn at different rates. To account for the finite resolution of biological neurons, the goal cell output was corrupted by an additive noise signal. For detailed methods of simulation, see Section 4.7.

A simple Gridworld environment (Fig 4.3) serves to observe the dynamics of learning in detail. There are three locations of interest: the entrance to the environment, experienced at the very start of exploration; a water source; and a food item. When the agent first enters the novel space, a feature neuron that responds to the entrance excites a goal cell, which leads to the potentiation of synapses onto that neuron. Effectively that tags the entrance, and from now on that goal cell encodes a virtual "entrance odor" that declines with distance from the entrance. With every step the agent takes, the map network gets updated, and the range of the entrance odor spreads further (Fig 4.3B top). At all times the agent could decide to follow this virtual odor uphill to the entrance.

The water source starts out invisible from anywhere except its special location (Fig 4.3B mid i-ii). However, as soon as the agent reaches the water, the water goal cell gets integrated in the circuit through the potentiation of synapses from map cells. Because the map network is already established along the path that the agent took,

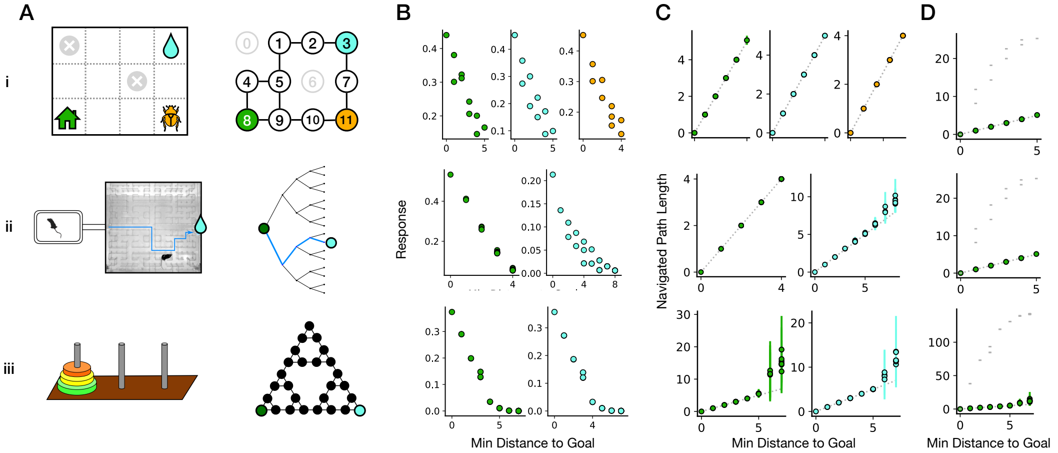


Figure 4.4: **Endotaxis can operate in environments with diverse topologies.** (A) Three tasks and their corresponding graph representations: i) Gridworld of Fig 4.3 with 3 goal nodes (home, water, and food). ii) A binary tree labyrinth similar to that used in mouse navigation experiments [204], with 2 goals (home and water). iii) Tower of Hanoi game, with 2 goals (the configurations of disks that solve the game). (B) The virtual odors after extensive exploration. For each goal neuron the response at every node is plotted against the shortest graph distance from the node to the goal. (C) Navigation by endotaxis: For every starting node in the environment this plots the number of steps to the goal against the shortest distance. (D) Example goal node from *C* comparing against the analytically computed average number of steps from random walker for all starting locations (grey lines).

that immediately creates a virtual “water odor” that spreads through the environment and declines with distance from the water location (Fig 4.3B mid iii).

As the agent explores the environment further, the virtual odors spread accordingly to the new locations visited (Fig 4.3B i-iv). After extensive exploration, the map and goal networks reach a steady state. Now the virtual odors are available at every point in the environment, and they decline monotonically with the shortest-path distance to the respective goal location (Fig 4.3B v). As one might expect, an agent navigating uphill on this virtual odor always reaches the goal location, and does so by the shortest possible path (Fig 4.4B-C i).

We performed a similar simulation for a labyrinth used in a recent study of mouse navigation [204]. The topology of the maze was a binary tree with a single entrance, multiple levels of T-junctions, and many end nodes (Fig 4.4A ii). A single source of water was located at one of the end nodes. In these experiments, mice learned the shortest path to the water source after visiting it ~10 times; they also performed error-free paths back to entrance on the first attempt [204]. In simulating the endotaxis network, we dedicated one goal cell to the entrance, and another to the water. The agent again explored the labyrinth extensively with a random walk, during which it also discovered the water location. Then we challenged the agent to navigate to either of the two targets from an arbitrary location on the graph, by following the respective virtual odor (Fig 4.4B ii). The agent indeed found the shortest route from every node (Fig 4.4C ii). Note incidentally that the goal signal is not perfectly monotonic with distance (Fig 4.4B ii), nonetheless it informs the correct turning decision at every node.

Endotaxis can learn to solve cognitive tasks beyond spatial navigation. For instance, the game "Towers of Hanoi" represents a more complex environment (Fig 4.4A iii). Disks of different sizes are stacked on three pegs, with the constraint that no disk can rest on top a smaller one. The game is solved by rearranging the pile of disks from one peg to another. In any state of the game, there are either 2 or 3 possible actions, and they form an interesting graph with many loops (Fig 4.4A iii). Again the simulated agent explored this graph by random walking. Once it encountered a solution by chance, that state was tagged with a virtual odor. After enough exploration the virtual odor signal was available from every possible game state, and the agent could solve the game rapidly from any arbitrary location. Note that some of these navigation paths were a few steps longer than absolutely necessary (Fig 4.4C iii), revealing the limits of resolution of the goal signal at large distances (Fig 4.4B iii).

4.5 Adaptation to change in the environment

An attractive feature of the endotaxis algorithm is that it separates learning the map from learning the target locations. In many real-world environments, the topology of the map (how are locations connected?) is probably more stable than the targets (which locations are interesting?). Separating the two allows the agent to adjust to changes on both fronts using different rules and time-scales. We illustrate an example of each.

4.5.1 Change in connectivity

Suppose that the connectivity of the environment changes. For example, a shortcut appears between two locations that used to be separated, or a blockage separates two previously adjacent locations (Fig 4.5A i-ii). This alters the correlation in firing among the point cells during the agent's explorations, and over time that will reflect in the synapses of the map network. How will endotaxis adapt to such changes?

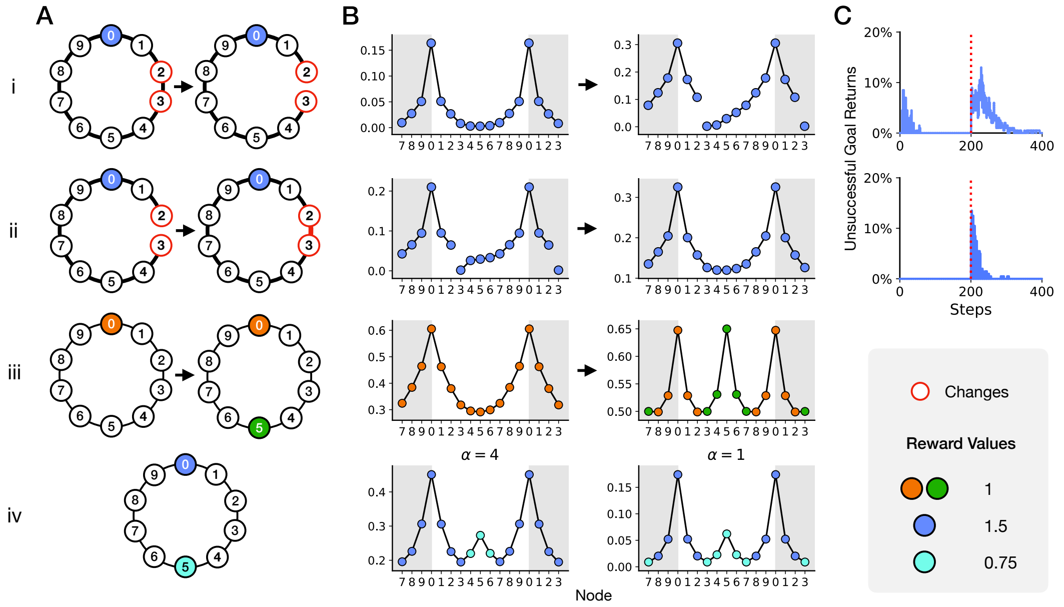


Figure 4.5: **Endotaxis adapts quickly to changes in the environment or the target locations.**
(A) A ring environment modified by sudden appearance of a blockage (i), a shortcut (ii), an additional goal target (iii), or two targets with different reward size (iv). Graphs shown before and after modification. Shaded nodes are target locations. Labels identify nodes on the graph. (B i-iii) Response profile of the goal neuron after sufficient exploration, shown just before modification (left, after 200 random steps) and after adaptation to the change (right, after an additional 200 steps). Color of nodes indicates the target that the agent will reach by following the virtual odor starting from that node. Note, the virtual odor peaks at either one or two targets depending on the environment, with a higher amplitude at the stronger target. (B iv) Varying α in Oja's Rule for map learning adjusts the tradeoff between distance and reward. With a large α the stronger target is favored from more starting nodes. (C) Fraction of errors in endotaxis from all possible starting nodes, as a function of time before and after the modification (dotted line).

To explore these adjustments, we considered navigation on a ring-shaped maze with a single goal location (Fig 4.5A i). Note that the ring is the simplest graph that offers two routes to a target, and we will evaluate whether the algorithm finds the shorter one. A simulated agent explored the ring by stepping among locations in a random walk, and built the map cell network from that experience. After a period of ~100 steps, navigation by endotaxis was perfect, in that the agent chose the shorter route to the goal from every start node (Fig 4.5B-C i). When we broke the ring by removing one link, endotaxis failed from some start nodes because it steered the agent towards the blocked path. However, after ~200 steps of additional exploration navigation returned to perfect performance again (Fig 4.5C i). Over this period the knowledge of the former link was erased from the map network (Fig 4.5B i), because the corresponding map synapses weakened while the link was not used.

When we introduced a new shortcut between previously separated locations (Fig 4.5A ii), a similar change took place. For a brief period, endotaxis was suboptimal, because the agent sometimes took the long route even though a shorter one was available (Fig 4.5C ii). However, that perturbation got incorporated into the map much more quickly than the broken link, after just a few tens of steps of exploration (compare Figs 4.5C i-ii). One can understand the asymmetry as follows: As the agent explores the environment, a newly available link is confirmed with certainty the first time it gets traveled. By contrast, the loss of a link remains uncertain until the agent has not taken that route many times.

4.5.2 Appearance of new targets

Suppose the agent has discovered one location with a water resource. Some time later water also appears at a second location (Fig 4.5A iii). When the agent discovers that, the same water goal cell will get activated and therefore receive a potentiation of synapses active at that second location. Now the input network to that goal cell contains the sum of two templates, corresponding to the map outputs from the two target locations. As before, the current map output gets filtered through these synaptic weights to create the virtual odor. One might worry that this goal signal steers the agent to a location half-way between the two targets. Instead, simulations on the ring showed that the virtual odor peaks at both targets, and endotaxis takes the agent reliably to the nearest one (Fig 4.5B iii).

4.5.3 Choice between multiple targets

Suppose one of the targets offering the same resource is more valuable than the other, for example because it gives a larger reward (Fig 4.5A iv). In the endotaxis model (Fig 4.1B), the larger reward causes higher activity of the feature cell that responds to this resource, and thus stronger potentiation of the synapses onto the associated goal cell (Eqn 4.19). Thus the input template of the goal cell becomes a weighted sum of the map outputs from the two target locations, with greater weight for the location with higher reward. In simulations, the virtual odor still showed two peaks, but the stronger target had a greater region of attraction (Fig 4.5B iv left); for some starting locations the agent chose the longer route in favor of the larger reward, a sensible behavior.

What determines the trade-off between the longer distance and the greater reward? In the endotaxis model (Fig 4.1B) this is set by α_{M}, one of the two parameters of the synaptic learning rule in the map network (Eqns 4.18 ,4.22, 4.27). A small α_{M} raises the cost of any additional step traveled and thus diminishes the importance of reward differences (Fig 4.5B iv right). By contrast, a large α_{M} favors the larger reward regardless of distance traveled.

In summary, endotaxis adapts readily to changes in the environment or in the availability of rewards. Furthermore, it implements a rational choice between multiple targets of the same kind, using a variable weighting of reward versus distance. None of these features required any custom tuning: They all follow directly from the basic formulation in Figure 4.1B.

4.6 Discussion

4.6.1 Summary of claims

We have presented a neural mechanism that can support learning, navigation, and problem solving in complex and changing environments. It is based on chemotaxis, namely the ability to follow an odor signal to its source, which is shared universally by most or all motile animals. The algorithm, called endotaxis, is formulated as a neural network that creates an internal "virtual odor" which the animal can follow to reach any chosen target location (Fig 4.1). When the agent begins to explore the environment, the network learns both the structure of the space, namely how various points are connected, and the location of valuable resources (Fig 4.3). After sufficient exploration, the agent can then navigate back to those target locations from any point in the environment (Fig 4.4). The algorithm is *always on* and it adapts flexibly to changes in the structure of the environment or in the locations of targets (Fig 4.5). Furthermore, even in its simplest form, endotaxis can arbitrate among multiple locations with the same resource, by trading off the promised reward against the distance traveled (Fig 4.5). Beyond spatial navigation, endotaxis can also learn the solution to purely cognitive tasks (Fig 4.4), or any problem defined by search on a graph. The neural network model that implements endotaxis has a close resemblance to known brain circuits. We propose that evolution may have built upon the ancient behavioral module for chemotaxis to enable much more general abilities for search and navigation, even in the absence of odor gradients. In the following sections, we consider how these findings relate to some well-established phenomena and results on animal navigation.

4.6.2 Animal behavior

The millions of animal species no doubt use a wide range of mechanisms to get around their environment, and it is worth specifying which of those problems endotaxis might solve. First, the learning mechanism proposed here applies to complex environments, namely those in which discrete paths form sparse connections between points. For a bird, this is less of a concern, because it can get from every point to any other "as the crow flies." For a rodent and many other terrestrial animals, on the other hand, the paths they may follow are constrained by obstacles and by the need to remain under cover. In those conditions, the brain cannot assume that the distance between points is given by euclidean geometry, or that beacons for a goal will be visible in a straight line from far away, or that a target can be reached by following a known heading. Second, we are focusing on the early experience with a new environment. Endotaxis can get an animal from zero knowledge to a cognitive map that allows reliable navigation towards goals encountered on a previous foray. It explains how an animal can return home from inside a complex environment on the first attempt [204], or navigate to a special location after encountering it just once (Figs 4.3,4.4). But it does not implement more advanced routines of spatial learning, such as stringing a habitual sequence of actions together into one, or internal deliberation to plan entire routes. Clearly, expert animals will make use of algorithms other than the beginner's choice proposed here.

A key characteristic of endotaxis, distinct from other forms of navigation, is the reliance on trial-and-error. The agent does not deliberate to plan the shortest path to the goal. Instead, it finds the shortest path by locally sampling the real-world actions available at its current point, and choosing the one that maximizes the virtual odor signal. In fact, there is strong evidence that animals navigate by real-world trial-and-error, at least in the early phase of learning [195]. Rats and mice often stop at an intersection, bend their body halfway along each direction, then choose one corridor to proceed. Sometimes they walk a few steps down a corridor, then reverse and try another one. These actions—called "vicarious trial and error"—look eerily like sniffing out an odor gradient, but they occur even in absence of any olfactory cues. Lashley [134], in his first scientific paper on visual discrimination in the rat, reported that rats at a decision point often hesitate "with a swaying back and forth between the passages." Similar behaviors occur in arthropods [237] and humans [209] when poised at a decision point. We suggest that the animal does indeed sample a gradient, not of an odor, but of an internally generated virtual odor that reflects the proximity to the goal. The animal uses the same policy of spatial sampling that it would apply to a real odor signal, consistent with the idea that endotaxis is built on the ancient behavioral module for chemotaxis.

Frequently a rodent stopped at a maze junction merely turns its head side-to-side, rather than walking down a corridor to sample the gradient. Within the endotaxis model, this could be explained if some of the point cells in the lowest layer (Fig 4.1B) are selective for head direction or for the view down a specific corridor. During navigation, activation of that “direction cell” systematically precedes activation of point cells further down that corridor. Therefore the direction cell gets integrated into the map network. From then on, when the animal turns in that direction, this action takes a step along the graph of the environment without requiring a walk in ultimately fruitless directions. In this way, the agent can sample the goal gradient while minimizing energy expenditure.

The vicarious trial and error movements are commonplace early on during navigation in a new environment. Later on, the animal performs them more rarely and instead moves smoothly through multiple intersections in a row [195]. This may reflect a transition between different modes of navigation, from the early endotaxis, where every action gets evaluated on its real-world merit, to a mode where many actions are strung together into behavioral motifs. At a late stage of learning, the agent may also develop an internal forward model for the effects of its own actions, which would allow for prospective planning of an entire route. An interesting direction for future research is to seek a neuromorphic circuit model for such action planning; perhaps it can be built naturally on top of the endotaxis circuit.

While rodents engaged in early navigation act as though they are sniffing out a virtual odor, we would dearly like to know whether the experience *feels like* sniffing to them. The prospects for having that conversation in the near future are dim, but in the meantime we can talk to humans about the topic. Human language has an intriguing set of metaphors for decision making under uncertainty: “this doesn’t smell right,” “sniff out a solution,” “that idea stinks,” “smells fishy to me,” “the sweet smell of success.” All these sayings apply in situations where we do not yet understand the rules but are just feeling our way into a problem. Going beyond mere correlation, there is also a causal link: Fishy smells can change people’s decisions on matters entirely unrelated to fish [141]. In the endotaxis model, (Fig 4.1B) this might happen if the mode switch is leaky, allowing real smells to interfere with virtual odors. Perhaps this partial synesthesia between smells and decisions results from the evolutionary repurposing of an ancient behavioral module that was intended for olfactory search.

4.6.3 Brain circuits

The proposed circuitry (Fig 4.1) relates closely to some real existing neural networks: the so-called cerebellum-like circuits. They include the insect mushroom body, the mammalian cerebellum, and a host of related structures in non-mammalian vertebrates [19, 72]. The distinguishing features are: A large population of neurons with selective responses (e.g. Kenyon cells, cerebellar granule cells), massive convergence from that population onto a smaller set of output neurons (e.g. Mushroom body output neurons, Purkinje cells), and synaptic plasticity at the output neurons gated by signals from the animal's experience (e.g. dopaminergic inputs to mushroom body, climbing fiber input to cerebellum). It is thought that this plasticity creates an adaptive filter by which the output neurons learn to predict the behavioral consequences of the animal's actions [19, 271]. This is what the goal cells do in the endotaxis model.

The analogy to the insect mushroom body invites a broader interpretation of what purpose that structure serves. In the conventional picture, the mushroom body helps with odor discrimination and forms memories of discrete odors that are associated with salient experience [104]. Subsequently the animal can seek or avoid those odors. But insects can also use odors as landmarks in the environment. In this more general form of navigation, the odor is not a goal in itself, but serves to mark a route towards some entirely different goal [130, 231]. In ants and bees, the mushroom body receives massive visual input, and the insect uses discrete panoramic views of the landscape as markers for its location [42, 233, 260]. Our analysis shows how the mushroom body circuitry can tie together these discrete points into a cognitive map that supports navigation towards arbitrary goal locations.

In this picture, a Kenyon cell that fires only under a specific pattern of receptor activation becomes selective for a specific location in the environment, and thus would play the role of a map cell in the endotaxis circuit (Fig 4.1).[3] After sufficient exploration of the reward landscape, the mushroom body output neurons come to encode the animal's proximity to a desirable goal, and that signal can guide a trial-and-error mechanism for steering. In fact, mushroom body output neurons are known to guide the turning decisions of the insect [10], perhaps through their projections to the central complex [142], an area critical to the animal's turning behavior. Conceivably

[3] Point cells and Map cells are the same in this picture.

this is where the insect's basic chemotaxis module is implemented, namely the policy for ascending on a goal signal.

Beyond the cerebellum-like circuits, the general ingredients of the endotaxis model—recurrent synapses, Hebbian learning, many-to-one convergence—are found commonly in other brain areas, including the mammalian neocortex and hippocampus. In the rodent hippocampus, an interesting candidate for map cells are the pyramidal cells in area CA3. Many of these neurons exhibit place fields and they are recurrently connected by synapses with Hebbian plasticity. It was suggested early on that random exploration by the agent produces correlations between nearby place cells, and thus the synaptic weights among those neurons might be inversely related to the distance between their place fields [171, 196]. However, simulations showed that the synapses are substantially strengthened only among immediately adjacent place fields [171, 196] (see also our Eqn 4.20), thus limiting the utility for global navigation across the environment. Here we show that a useful global distance function emerges from the *output* of the recurrent network (Eqns 4.23, 4.26, 4.27) rather than its synaptic structure. Further, we offer a biologically realistic circuit (Fig 4.1B) that can read out this distance function for subsequent navigation.

4.6.4 Neural signals

The endotaxis circuit proposes three types of neurons—point cells, map cells, and goal cells—and it is instructive to compare their expected signals to existing recordings from animal brains during navigation behavior. Much of that prior work has focused on the rodent hippocampal formation [169], but we do not presume that endotaxis is localized to that structure. The three cell types in the model all have place fields, in that they fire preferentially in certain regions within the graph of the environment. However, they differ in important respects:

Size and location The place field is smallest for a point cell; somewhat larger for a map cell, owing to recurrent connections in the map network; and larger still for goal cells, owing to additional pooling in the goal network. Such a wide range of place field sizes has indeed been observed in surveys of the rodent hippocampus, spanning at least a factor of 10 in diameter [129, 267]. Some place cells show a graded firing profile that fills the available environment. Furthermore, one finds more place fields near the goal location of a navigation task, even when that location has no overt markers [107]. Both of those characteristics are expected of the goal cells in the endotaxis model.

Dynamics The endotaxis model assumes that point cells exist from the very outset in any environment. Indeed, many place cells in the rodent hippocampus appear within minutes of the animal's entry into an arena [84, 267]. Furthermore, any given environment activates only a small fraction of these neurons. Most of the "potential place cells" remain silent, presumably because their sensory trigger feature does not match any of the locations in the current environment [3, 65]. In the endotaxis model, each of these sets of point cells is tied into a different map network, which would allow the circuit to maintain multiple cognitive maps in memory [171]. Finally a small change in the environment, such as appearance of a local barrier (Fig 4.5), can indeed lead to disappearance and appearance of nearby place cells [5].

Goal cells, on the other hand, are expected to appear suddenly when the animal first arrives at a memorable location. At that moment, the goal cell's input synapses from the map network are activated and the neuron immediately develops a place field. This prediction is reminiscent of a startling experimental observation in recordings from hippocampal area CA1: A neuron can suddenly start firing with a fully formed place field that may be located anywhere in the environment [28]. This event appears to be triggered by a calcium plateau potential in the dendrites of the place cell, which potentiates the excitatory synaptic inputs the cell receives. A surprising aspect of this discovery was the large extent of the resulting place field, which requires the animal several seconds to cover. Subsequent cellular measurements indeed revealed a plasticity mechanism that extends over several seconds [149]. The endotaxis model offers an alternative mechanism underlying the large place field of a goal cell: Through its input synapses, the goal cell taps into the map network, which has already developed long-range connections prior to the agent finding the goal location. In this picture all the synaptic changes may be local in time and space, without requiring an extended time scale for plasticity.

4.6.5 Learning theories

Endotaxis has similarities with *reinforcement learning* (RL) [236]. In both cases, the agent explores a number of locations in the environment. In RL, these are called *states*, and every state has an associated *value* representing how close the agent is to rewards. In endotaxis, this is the role of the virtual odor, represented by the activity of a goal neuron. The value function gets modified through the experience of reward when the agent reaches a valuable resource; in endotaxis, this happens via update of the synapses in the goal network (**G** in Fig 4.1B). In both RL and endotaxis, when the animal wishes to exploit a given resource, it navigates so as to maximize the value function. Over time that value function converges to a form that allows the agent to find the goal directly from every starting state. The exponential decay of the virtual odor with increasing distance from the target (Eqn 4.27) is reminiscent of the exponential decay of the value function in RL, controlled by the discount factor, γ [236].

In endotaxis, much of the learning happens independent of any reinforcement. During exploration, the circuit learns the topology of the environment, specifically by updating the synapses in the map network (**M** in Fig 4.1B). The presence of rewards is not necessary for map learning: Until a resource is found for the first time, the value function remains zero because the **G** synapses have not yet been established (Eqn 4.17). Eventually, when the goal is encountered, **G** is updated in one shot and the value function becomes nonzero throughout the known portion of the environment. Thus the agent learns how to navigate to the goal location from a single reinforcement (Fig 4.3). This is possible because the ground has been prepared, as it were, by learning a map. In animal behavior this phenomenon is called *latent learning*. Early debates in animal psychology pitched latent learning and reinforcement learning as alternative explanations [244]. Instead, in the endotaxis algorithm, neither can function without the other (see Eqn 4.17). In *model-based* reinforcement learning, the agent could learn a forward model of the environment and uses it to update a value function. A key difference is that endotaxis learns the distances between all pairs of states, and can then establish a value function after a single reinforcement, whereas RL typically requires an iterative method to establish the value function [99, 164, 235].

The neural signals in endotaxis bear some similarity to the so-called *successor representation* [54, 59, 230]. This is a proposal for how the brain might encode the current state of the agent, intended to simplify the mathematics of time-difference reinforcement learning. Each neuron stands for a possible state of the agent. The activity of neuron j is proportional to the time-discounted probability that the agent will find itself at state j in the future. Thus, the output of the endotaxis map network (Eqns 4.6, 4.23) qualitatively resembles a successor representation. However there are some important differences: First, the successor representation depends not only on the structure of the environment, but on the optimal policy of the agent, which in turn depends on the distribution of rewards. Thus the successor representation must itself be learned through a reinforcement algorithm. There is agreement in the literature that the successor representation would be more useful if the model of the environment were independent of reward structure [90]; however, it is believed that “it is more difficult to learn” [59]. By contrast, the map matrix in the endotaxis mechanism is built from a policy of random exploration independent of the reward landscape. Second, no plausible biomorphic mechanism for learning the successor representation has been proposed yet, whereas the endotaxis circuit is made entirely from biologically realistic components.

4.6.6 Outlook

In summary, we have proposed a simple model for spatial learning and navigation in an unknown environment. It includes an algorithm, as well as a fully-specified neural circuit implementation. The model makes quantitative and testable predictions that match a diverse set of observations in behavior, anatomy, and physiology, from insects to rodents (Secs 4.6.2-4.6.4). Of course, the same observables may be consistent with other models, and in fact multiple navigation mechanisms may be at work in parallel or during successive stages of learning. Perhaps the most distinguishing features of the endotaxis algorithm are its reliance on trial-and-error sampling and the close relationship to chemotaxis. To explore these specific ingredients, future research could work backwards: First find the neural circuit that controls the random trial-and-error sampling of odors. Then test if that module receives a convergence of goal signals from other circuits that process non-olfactory information. If so, that could lead to the mode switch which routes one or another goal signal to the decision-making module. Finally, upstream of that mode switch lies the soul [190] of the animal that tells the navigation machinery what goal to pursue. Given recent technical developments we believe that such a program of module-tracing is within reach, at least for the insect brain.

4.7 Supplement

The core function of the endotaxis network is to learn the distance between any two points in the environment starting from purely local connectivity. As the agent explores the graph of the environment, the point cells for two adjacent locations briefly fire together. This is the local event that drives synaptic learning in the map population. Eventually the map network learns the global structure of the graph. In particular, for any chosen goal node on the graph, the network computes a virtual odor signal that varies with the agent’s location and declines monotonically with the distance from the goal. Using that distance function, the agent can navigate to the goal node by the shortest path. In this section, we explain how this global distance measure comes about. We start with an analytical result about computing distances on a graph, continue with a linear rate model of the endotaxis network, then proceed to a nonlinear treatment and numerical experiments that supplement results in the text.

4.7.1 A neuromorphic function to compute the shortest distance on a graph

Finding the shortest path between all pairs of nodes on a graph is a central problem of graph theory, known as "all pairs shortest path" (APSP) [1]. Generally an APSP algorithm delivers a matrix containing the distances D_{ij} for all pairs of nodes. That matrix can then be used to construct the actual sequence corresponding to the shortest path iteratively. The Floyd-Warshall algorithm [79] is simple and works even for the more general case of weighted edges between nodes. Unfortunately, we know of no plausible way to implement Floyd-Warshall's three nested loops of comparison statements with neurons.

There is, however, a simple function for APSP that can be solved by a recurrent neural network. Specifically: If a connected, directed graph has adjacency matrix A_{ij},

$$A_{ij} = \begin{cases} 1, & \text{if node } i \text{ can be reached from node } j \text{ in one step} \\ 0, & \text{otherwise, including the } i = j \text{ case} \end{cases}, \tag{4.1}$$

then with a suitably small positive value of γ the shortest path distances are given by

$$D_{ij} = \left\lceil \frac{\log\left[(\mathbf{1} - \gamma\,\mathbf{A})^{-1}\right]_{ij}}{\log \gamma} \right\rceil \tag{4.2}$$

where **1** is the identity matrix, and the half-square brackets mean "round up to the nearest integer."

Proof: The powers of the adjacency matrix represent the effects of taking multiple steps on the graph, namely

$$\left[\mathbf{A}^k\right]_{ij} = N_{ij}^{(k)} = \text{number of distinct paths to get from node } j \text{ to node } i \text{ in } k \text{ steps}$$

where a path is an ordered sequence of edges on the graph. This can be seen by induction as follows. By definition:

$$N_{ij}^{(1)} = A_{ij}.$$

Suppose we know $N_{ij}^{(k)}$ and want to compute $N_{ij}^{(k+1)}$. Every path from j to i of length $k+1$ steps has to reach a neighbor of node i in k steps. Therefore,

$$N_{ij}^{(k+1)} = \sum_l A_{il} N_{lj}^{(k)}. \tag{4.3}$$

The RHS corresponds to multiplication by $\mathbf{A}$, so the solution is

$$N_{ij}^{(k)} = \left[\mathbf{A}^k\right]_{ij}.$$

We are particularly interested in the shortest path from node j to node i. If the shortest distance D_{ij} from j to i is k steps, then there must exist a path of length k but not of any length $< k$. Therefore,

$$D_{ij} = \min_k N_{ij}^{(k)} > 0. \tag{4.4}$$

Now consider the Taylor series

$$\begin{aligned} \mathbf{Y} &= (\mathbf{1} - \gamma\mathbf{A})^{-1} \\ &= \mathbf{1} + \gamma\mathbf{A} + \gamma^2\mathbf{A}^2 + \ldots \end{aligned} \tag{4.5}$$

Then

$$Y_{ij} = \sum_{k=0}^{\infty} N_{ij}^{(k)} \gamma^k = N_{ij}^{(D_{ij})} \gamma^{D_{ij}} + N_{ij}^{(D_{ij}+1)} \gamma^{D_{ij}+1} + \ldots \tag{4.6}$$

We will show that if γ is chosen positive but small enough, then the growth of $N_{ij}^{(k)}$ with increasing k gets eclipsed by the decay of γ^k such that

$$\gamma^{D_{ij}} < Y_{ij} < \gamma^{D_{ij}-1}. \tag{4.7}$$

The left inequality is obvious from Eqn 4.6 because $N_{ij}^{(D_{ij})} \geq 1$ by Eqn 4.4.

To understand the right inequality, note first that $N_{ij}^{(k)}$ is bounded by a geometric series. From Eqn 4.3, it follows that

$$N_{ij}^{(k)} < q^k$$

where q is the largest number of neighbors of any node on the graph. So from Eqn 4.6,

$$Y_{ij} < (q\gamma)^{D_{ij}} + (q\gamma)^{D_{ij}+1} + \cdots = \frac{(q\gamma)^{D_{ij}}}{1 - q\gamma}. \tag{4.8}$$

This expression is $< \gamma^{D_{ij}-1}$ (Eqn 4.7) as long as

$$\gamma < \frac{1}{q + q^{D_{ij}}}. \tag{4.9}$$

In addition, because

$$D_{ij} < n \equiv \text{ number of nodes on the graph}$$

this is satisfied if one chooses γ such that

$$\gamma < \frac{1}{q + q^n}. \tag{4.10}$$

With that condition on γ the inequality (4.7) holds, and taking the logarithm on both sides leads to the desired result:

$$D_{ij} = \left\lceil \frac{\log Y_{ij}}{\log \gamma} \right\rceil.$$

We will see that the endotaxis network, in its linear approximation, computes a goal signal equal to the scalar products of the column-vectors in $\mathbf{Y}$, namely

$$E_{ij} = \text{ “goal signal from node } j \text{ to } i\text{”} = \sum_k Y_{ki} Y_{kj}. \tag{4.11}$$

To understand how that goal signal E_{ij} varies with distance, one can follow arguments parallel to those that led to Eqn 4.6. Using the upper bound by the geometric series (Eqn 4.8) and inserting in Eqn 4.11, one finds again that it is possible to choose a γ small enough to satisfy

$$\gamma^{D_{ij}} < E_{ij} < \gamma^{D_{ij}-1}. \tag{4.12}$$

Under those conditions, the goal signal E_{ij} decays exponentially with the graph distance D_{ij}.

4.7.2 Linear approximation of the endotaxis network

Here we formalize the network model of Figure 4.1B, and develop a linear approximation that allows an analytical treatment and connection with the previous section.

The environment is parcelled into a set of discrete locations that are sparsely connected to each other. The locations and connectors form a graph that is fully specified by the adjacency matrix A_{ij} (Eqn 4.1).

Each node on the graph has a point cell corresponding to that location. The point cell i fires at a rate of 1 when the agent is at location i, and at a lower level w, with $0 < w < 1$, when the agent is at the neighboring nodes. Thus the firing fields of neighboring point cells overlap somewhat; this produces correlations among point cells along the agent's trajectory which will drive synaptic plasticity.

$$\begin{aligned} u_i(x) &= \text{firing rate of point cell } i \text{ with the agent at node } x && (4.13)\\ &= \delta_{ix} + w\, A_{ix} && (4.14) \end{aligned}$$

where δ_{ix} is the Kronecker delta.

We begin by treating neural processing (Fig 4.1B) using a textbook linear rate model [60]. In this version, the output of the map network (Fig 4.1B) is

$$\mathbf{v} = \mathbf{u} + \mathbf{M}\mathbf{v} = (\mathbf{1} - \mathbf{M})^{-1}\mathbf{u} \tag{4.15}$$

where $\mathbf{u}$ is the vector of point cell outputs, $\mathbf{v}$ is the vector of map cell outputs, and $\mathbf{M}$

is the matrix of recurrent synapses among map cells.

A goal cell g receives sensory input s_g from neurons that signal the goal resource available to the agent at the current node:

$$s_g(y) = \text{amount of resource } g \text{ present when the agent is at node } y. \tag{4.16}$$

In addition the goal cell gets input from the map neurons via the network of goal synapses. Thus the vector of goal cell activities with the agent at node x is

$$\mathbf{r}(x) = \mathbf{s}(x) + \mathbf{G} \quad \mathbf{v}(x) = \mathbf{s}(x) + \mathbf{G}(\mathbf{1} - \mathbf{M})^{-1}\mathbf{u}(x). \tag{4.17}$$

The recurrent synapses among map cells undergo Hebbian plasticity. To keep the synaptic strengths bounded, we adopted the standard Oja's Rule [60]:

$$\frac{\mathrm{d}M_{ij}}{\mathrm{d}t} = \beta_\mathrm{M}(\alpha_\mathrm{M} v_i v_j - M_{ij} {v_i}^2) \tag{4.18}$$

where β sets the speed of synaptic plasticity and α its strength. The map network has no self-synapses: $M_{ii} = 0$.

The synapses from map cells to goal cells also undergo Hebbian plasticity, again via Oja's Rule

$$\frac{\mathrm{d}G_{gi}}{\mathrm{d}t} = \beta_\mathrm{G}(\alpha_\mathrm{G} r_g v_i - G_{gi} {r_g}^2). \tag{4.19}$$

These updates occur only when the agent is at the target location, gated by a reward signal associated with the resource located there. Because learning about targets is conceptually different from learning the map of the environment, we allowed $\alpha_\mathrm{G}, \beta_\mathrm{G}$ to differ from $\alpha_\mathrm{M}, \beta_\mathrm{M}$.

During exploration, the agent performs a random walk throughout the entire graph. The resulting activity of point cells drives the map network. It is well known that a Hebbian recurrent network of this type will learn the correlation structure of its inputs [60, 87]. Evaluating Eqn 4.18 after synapses have equilibrated leads to

$$M_{ij} = \alpha \frac{\langle v_i v_j \rangle}{\langle {v_j}^2 \rangle}. \tag{4.20}$$

In the limit of small M_{ij}, i.e. if the inputs from point cells dominate, then $v_i \approx u_i$ and one gets to lowest order

$$M_{ij} \approx \alpha \frac{\langle u_i u_j \rangle}{\langle {u_i}^2 \rangle} = \alpha\, w\, A_{ij} \equiv \gamma\, A_{ij} \tag{4.21}$$

where we assumed

$$\gamma = \alpha\, w << 1. \tag{4.22}$$

In this approximation, the recurrent synapses M_{ij} directly reflect the connections among point cells and thus the adjacency matrix of the graph.

The output of the map network (Eqn 4.15) is

$$\mathbf{v} = (\mathbf{1} - \mathbf{M})^{-1}\mathbf{u} = (\mathbf{1} - \gamma\mathbf{A})^{-1}\mathbf{u}. \tag{4.23}$$

Comparison to Eqn 4.5 shows that the recurrent network of map cells effectively computes the all-pairs distance function Y_{ij}. If the agent is at node x, then the map output $\mathbf{v}(x)$ equals the x-th column vector of the matrix $\mathbf{Y}$ (in the limit of small w and γ):

$$v_i(x) = Y_{ix} \tag{4.24}$$

which declines exponentially with the graph distance D_{ix} (Eqn 4.7). These distance-dependent humps of activity are schematized in Fig 4.2C-E.

The remaining problem is how to use the map output to encode the distance to a specific remembered goal location. Suppose goal g has a rewarding resource only at node y, specifically $s_g(x) = \delta_{xy}$ (Eqn 4.16). When the agent first arrives at location y, the synaptic plasticity rule (Eqn 4.19) updates the goal synapses G_{gi} from zero to a profile proportional to the current map output:

$$G_{gi} \sim v_i(y). \tag{4.25}$$

Subsequent visits will strengthen that profile. From then on, when the agent is at a

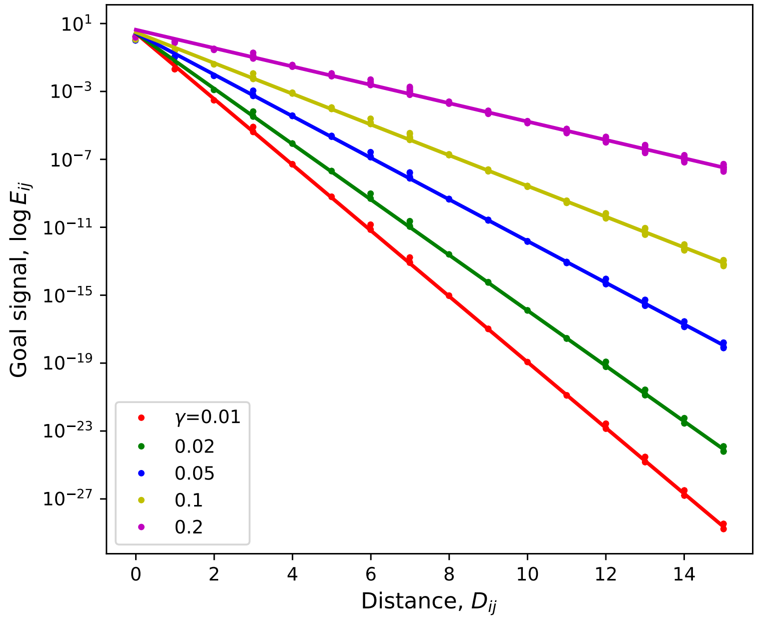


Figure 4.6: **The goal signal and the choice of γ.** The goal signal declines exponentially with graph distance (the tower of Hanoi graph with 4 levels was used for these simulations). Data points indicate the goal signal between all pairs of nodes, computed with different values of γ, and plotted against the distance on the graph between the nodes. Lines are exponential fits to the data.

location $x \neq y$, the virtual odor varies according to Eqn 4.17:

$$\begin{aligned} r_g(x) &= \mathbf{s}(x) + \mathbf{G} \quad \mathbf{v}(x) \\ &\sim 0 + \mathbf{v}(y) \cdot \mathbf{v}(x) \equiv E_{xy}. \end{aligned} \tag{4.26}$$

This corresponds to the goal signal E analyzed above (Eqns 4.11, 4.12, Fig 4.6). Thus the virtual odor computed by the endotaxis network decays exponentially with the agent's distance from the goal:

$$E_{xy} \sim \gamma^{D_{xy}}. \tag{4.27}$$

4.7.3 Limitations of the linear model

The analytical treatment above relied on multiple small-signal approximations, and the results are guaranteed only in the limit of $\gamma << 1$ (Eqn 4.22). To explore the practical range of γ, we performed numerical experiments on a variety of graphs and found that the exponential dependence of the goal signal E_{ij} on distance (Eqn 4.12) holds over a wide range of γ (Fig 4.6).

However, there is a critical value of γ beyond which the goal signal fails catastrophically: Recall that the Taylor expansion (4.5) has a convergence radius of 1. That means all the eigenvalues of $\gamma\mathbf{A}$ must have absolute value < 1, which requires

$$\gamma < \gamma_{\mathrm{c}} \equiv \frac{1}{\text{largest absolute eigenvalue of } \mathbf{A}}. \tag{4.28}$$

Outside of that convergence radius, the expression $(\mathbf{1} - \gamma\mathbf{A})^{-1}$ can no longer be interpreted as counting paths on the graph and therefore loses any connection to graph distance. The largest eigenvalue of the adjacency matrix is related to the typical number of edges per node. For the types of environmental graphs considered here, $\gamma_{\mathrm{c}} = 0.3 - 0.4$.

This upper bound on γ means that the goal signal decays rather steeply with distance. In the most favorable case $\gamma = \gamma_{\mathrm{c}}$, the decay still amounts a factor of ~ 3 per step along the graph, spanning many orders of magnitude over the extent of even modest graphs (Fig 4.6). For an agent to navigate to a target from a distance, it must be able to resolve changes in the goal signal along the entire way. But realistic neurons operate only over a limited dynamic range of the inputs, spanning perhaps 2 orders of magnitude. Thus, for a biologically plausible implementation, one needs to enhance the model with non-linearities.

4.7.4 Endotaxis with non-linear activation functions and noise

Perhaps the simplest enhancement is an expansive input-output function for the map neurons. This can amplify the signal locally within the map network, thus converting the exponential decay with distance to a more linear decay profile. For this purpose we modeled the response of a map cell to its input as

$$v_i = \phi\left(u_i + \sum_j M_{ij} v_j\right) \tag{4.29}$$

where the activation function ϕ is

$$\phi(x) = \max\left(0, a \cdot \left(1 + \frac{\log x_i}{k}\right)\right) \tag{4.30}$$

and a and k are positive constants.

The goal cell response was modified to include a noise signal:

$$r_g = s_g + \sum_j G_{gj} v_j + \epsilon \tag{4.31}$$

where the noise ϵ is a Gaussian variable with standard deviation of 2% of the range of the goal cell signal. This sets a limit on the decisions the agent makes during navigation, because differences in goal signal smaller than ϵ cannot be resolved.

The nonlinear activation of map cells allows for a greater signal spread throughout the map network. At the same time the map synapses should come to reflect the local adjacencies on the graph (Eqn 4.23). These desirables can be reconciled by imposing a simple threshold on the Hebbian learning rule for the map synapses:

$$\frac{\mathrm{d}M_{ij}}{\mathrm{d}t} = \rho(\beta_{\mathrm{M}}(\alpha_{\mathrm{M}} v_i v_j - M_{ij} {v_i}^2)) \tag{4.32}$$

$$\rho(y) = \begin{cases} y, & \text{if } y \geq \sigma \\ 0, & \text{otherwise} \end{cases}. \tag{4.33}$$

The Hebbian learning rule for the goal connections G_{gj} is unchanged from the linear model (Eqn 4.19).

4.7.5 Simulations

Figures 4.3, 4.4, and 4.5 report the results of endotaxis under the nonlinear response model described above. During exploration we gave the agent a trajectory through the graph, either chosen by design (Fig 4.3) or as an unbiased random walk (Figs 4.4, 4.5). After every step of the exploratory walk we computed the cell activities in a forward pass from point cells to goal cells. Then we updated the synaptic weights in the two networks **M** and **G** via the above learning rules. Matrix operations were

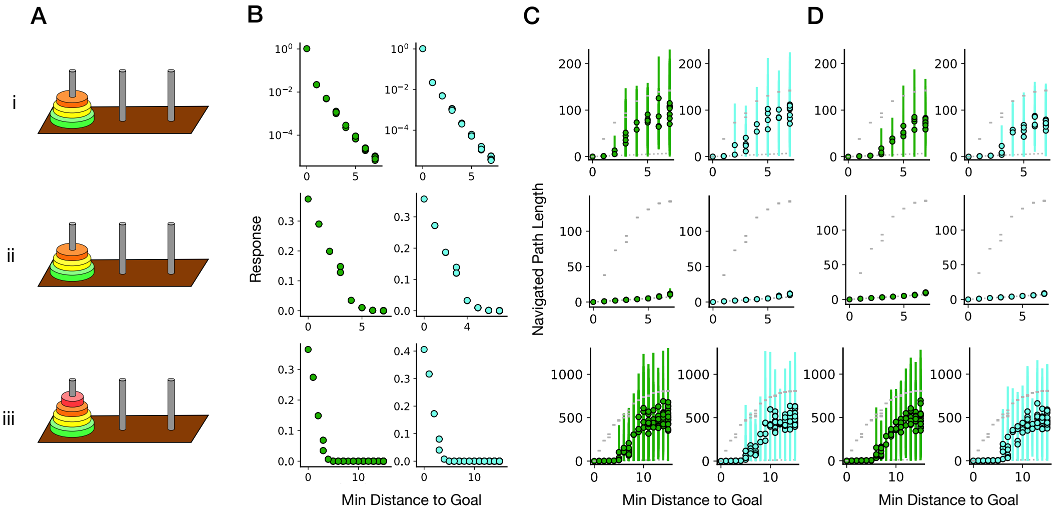


Figure 4.7: **Performance of models with linear vs logrithmic map cell activation.** (i) linear model. (ii-iii) nonlinear model. (A) Task. (B) Goal response. (C) Navigation performance with additive noise sampled from Gaussians with a standard deviation of 2% and (D) 0.5% of the goal cell dynamic range.

implemented in JAX [36], but for the task complexity explored in this paper there was no need for GPU acceleration.

For tests of navigation towards a goal, the agent was placed at an arbitrary starting node on the graph. The agent evaluated the goal signal at each of the neighboring nodes and stepped to the one with the highest value. This iterated until the agent reached the target.

The parameters used during these simulations are summarized in Table 4.1.

Table 4.1: **Learning parameters used for simulations.**

Task	α_M	β_M	α_G	β_G	k	a	r_g	Steps
Gridworld	2.9	1.2	$15\alpha_M$	0.07	5.5	0.04	0.1	$3e3$
Binary Tree	3.3	1.2	$15\alpha_M$	0.07	5.5	0.04	0.1	$7e3$
Hanoi (3 Discs)	2.65	1.2	$15\alpha_M$	0.07	5.5	0.04	0.1	$5e3$
Hanoi (3 Discs, Linear)	0.05	0.02	$0.5\alpha_M$	0.3	-	-	1	$5e3$
Hanoi (4 Discs)	2.291	1.2	$15\alpha_M$	0.07	5.5	0.04	0.1	$3e4$
Blockage	5.7	1.2	$15\alpha_M$	0.07	5.5	0.04	0.1	$[2e2, 2e2]$
Shortcut	5.8	1.8	$15\alpha_M$	0.07	5.5	0.04	0.1	$[2e2, 2e2]$
Dual Reward	5.7	1.2	$15\alpha_M$	0.07	5.2	0.04	$[0.1, 0.1]$	$1e3$
Dual Reward (Biased)	$[4, 1]$	1.7	$15\alpha_M$	0.07	5.5	0.04	$[0.1, 0.03]$	$2e3$

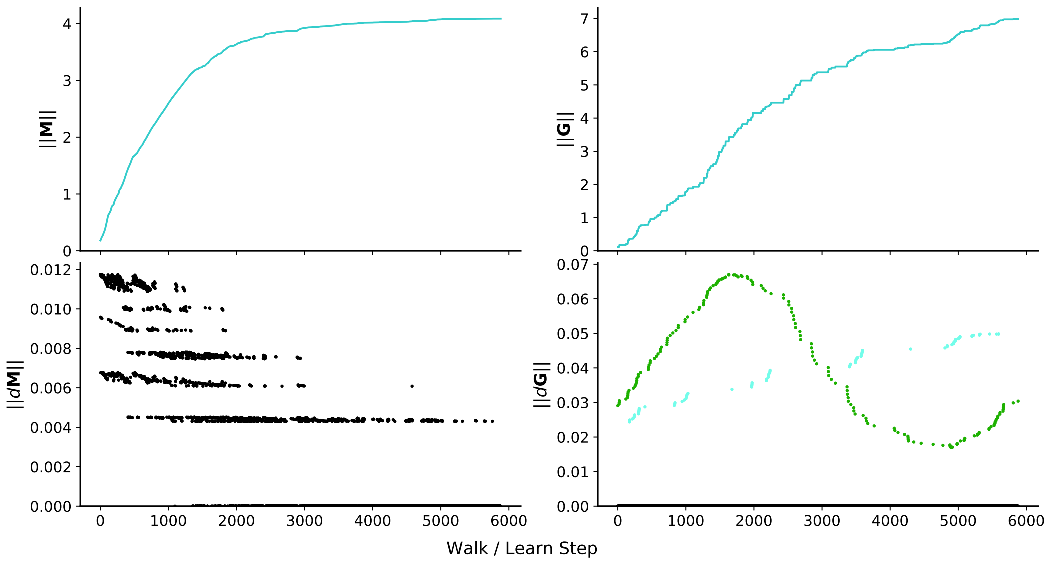


Figure 4.8: **Dynamics of online learning.**
Evolution of the map matrix (||**M**|| and ||d**M**||) and the goal matrix (||**G**|| and ||d**G**||) during exploration of binary tree of Fig 4.4A i. Lower-right figure: colors of scatter points indicate when the agent traverses either goal (green: home, cyan: reward, black: neither). See text for details.

Dynamics of learning

Figure 4.8 illustrates the state of the synaptic networks over the course of online learning, as observed during an exploratory random walk on the binary maze graph (Fig 4.4A-ii). The norm of the map matrix ||**M**|| increases continuously through steady small updates ||d**M**||. By comparison the goal matrix ||**G**|| increases in noticeable steps of ||d**G**|| every time the agent visits a goal location. With sufficiently low α and β, the network learns stably and gradually approaches a steady state. However, as demonstrated in the text, even the first visit to a goal location already produces a goal signal that allows a reliable return to that location.

Change in connectivity

We analyzed how the agent's navigation performance evolved with time following a change in connectivity (Fig 4.5A.i-ii). After each time step during the exploratory random walk we asked if the agent could navigate to the goal by the shortest path. We assumed that the appearance of a block or a shortcut between two adjacent nodes will alter the sensory cues around both locations (2 and 3 in Fig 4.5A.i-ii). Therefore the point cells that used to encode those locations drop silent, and the respective map cells lose their afferent input, while still remaining in the recurrent network.

At the same time two new point cells appear at those locations, because the new cues match their selectivity. Their map cells now receive afferent input from the respective locations, but their recurrent synapses start at zero weight. The agent then continues a random walk around the ring, subject to the new constraints, and the learning algorithm proceeds as usual.

Data and code availability

Data and code to reproduce the reported results are available at `https://github.com/tonyzhang25/Zhang-2021-Endotaxis`. Following acceptance of the manuscript they will be archived in a permanent public repository.

Chapter 5

FUTURE DIRECTIONS

Having presented the majority of my dissertation work, I briefly discuss some possible avenues for further investigation. Of course, these suggestions are not exhaustive and merely serve to provoke thought and dialogue regarding possible avenues to extend the work that has been completed thus far.

5.1 Sensing magnetism

The work contained here serves merely as an initial foray into what will hopefully be a long and productive inquiry into the neural mechanisms of the magnetic sense thought to underlie the astounding ability of animals to migrate long distances. The work presented above demonstrates a minimal approach to experimentation and analysis that endeavors to assess whether a given animal is responsive to magnetic stimulation. The approach assumes that neurons *somewhere* increase their firing rate as the magnetic field around the animal changes. Presenting a sinusoidal magnetic stimulus enables the use of powerful Fourier frequency methods in analysing neural responses (or lack thereof) in brains.

Importantly, the details of the brain's representation of the signal are not assumed; many types of possible encoding strategies should be detectable with our method. This neutrality or agnosticism to coding details makes the approach especially amenable to collaborative efforts between labs. By supplying partner labs with small, readily manipulable solenoids for magnetic stimulation, we hope to minimize setup time and disruption to our collaborators' usual experimental activities. This, combined with our standardized analysis pipeline, facilitates communication and reproducibility of individual claims of measurable magnetic sensation. Efforts are currently underway in the lab to expand both the scope of collaborating labs as well as the range of species tested within our lab to include pigeons, butterflies, and salamanders.

Improved understanding of magnetic perception in animals holds the promise that it might one day be reverse-engineered, similar to the development of optogenetics [35], to enable non-invasive manipulation of neural activity via changes in an ambient magnetic field. This would present enormous opportunities for advancement both within basic neuroscience research as well as potentially in translational medicine.

5.2 Maze navigation

The maze experiments presented in Chapter III reinvigorate a style of research that was much more prevalent a century ago. Machine vision has advanced considerably in recent years [152]. This enables robust and detailed tracking of animals, even in complex environments such as ours. I believe this research is timely, given the prevailing emphasis on behavioral tasks that take an exorbitant amount of time to train. Both on the grounds of data throughput, as well as ethological relevance to the animal, our free-foraging approach to studying navigation presents an exciting opportunity to generate truly novel observations of experimental subjects. To my knowledge, no task elicits as much learning, measured in the memory for sequential decisions, in as little time as the maze navigation paradigm.

Ongoing work, conducted by Jieyu Zheng and Rogério Guimarães, seeks to determine the upper limits of mouse spatial learning. No limit has been found as of yet despite expansion of the maze to require more than twenty branch points between the home cage and water port. Multiple animals, presented with multiple distinct mazes of this degree of complexity, demonstrated the ability to navigate perfectly to the water and back without errors. Other preliminary experiments have found that mice can quickly learn multiple reward locations and adapt to changes—like shortcuts or blockages—in maze connectivity. Ultimately, we aim to further characterize the underlying representations or heuristics the mice use to navigate these complex environments by looking for repeatable correlations between variations in maze configuration and variability in rates of learning.

Concurrently with these behavioral extensions, I am developing a wireless electrophysiology approach to enable recordings of neural activity during navigation behaviors. Daniel Pollak and Jiang Wu are simultaneously exploring more traditional approaches in which data leaves the brain and connects to a computer via a cable. These latter approaches are more easy to implement and allow recording of more neurons but come at the cost of impeding free action through a maze with a closed ceiling.

Myriad fascinating questions might be addressed with a combination of neural recording and complex maze navigation. For instance, what neural events precede moments of sudden insight? Is the rate of learning correlated with replay or theta events? Is there a neural signature that accompanies rapid changes in behavior, such as switching from exploration to exploiting known direct routes between resources? Theoretical work that extends the scope of the Endotaxis model to include other aspects of navigation may generate predictions and hypotheses that can be tested in maze experiments in which neural activity is recorded.

Finally, preliminary data presented in the Appendix suggest that mice can recover from a range of experimental sensory and neural perturbations. How might the brain maintain an appropriate homeostatic state or recover such a state after perturbation? Answers to this question may suggest new ways to restore homeostatic brain function in humans suffering from psychiatric illness.

APPENDIX

The maze experiments described in Chapter III leave open many questions regarding the sensory systems employed by the animal during navigation. We showed that a 180 degree rotation of the maze did not markedly impair the ability of the animal to navigate efficiently from the maze entrance to the water port. However, this tests the ability of the mouse to navigate to the water *after* learning the location of the reward. Perhaps olfaction is necessary for initial learning of the environment despite not being essential for recall of this information late in the experiment. To test this, Jiang Wu performed a series of experiments in which olfactory neurons were ablated with zinc sulfate [154]. Anosmia was confirmed via a digging assay in which hungry mice searched in cage bedding for a hidden food pellet. Animals showed some mild impairment early in the maze experiment but eventually were able to acquire water at a steady rate. Furthermore, mice learned direct routes to the water port from the entrance and some animals showed sudden insight, assessed via a sharp increase in the rate of long paths to the water from multiple locations throughout the maze. These olfactory ablation experiments bolster the idea that the sense of smell is not critical for the mouse to be able to navigate effectively.

What about the other sensory modalities? Is one especially essential? The experiments described in Chapter III occurred exclusively in the dark thereby demonstrating the dispensability of vision for maze navigation in mice. Additional experiments performed by Jiang Wu tested mice navigation after whisker trimming. Mice still managed to reach the water and develop direct routes. One caveat is that we did not anesthetize the animal's face, so the animal was presumably still able to sense the presence of walls, albeit only upon much closer inspection.

Setting peripheral sensory systems aside, one might reasonably ask which brain regions mediate the maze navigation behaviors of Chapter III. In collaboration with Zeynep Turan [251], I collected data from animals lacking hippocampus and the majority of cortex. These lesions were performed either via genetic [128] or surgical means. These lesioned animals showed drastic impairments in efficient exploration early in the session, often repeating movements in small regions of the maze an enormous number of times. Yet, remarkably, these same animals eventually succeeded in learning to take short, often direct, paths to the water location.

Lastly, as part of a collaboration with Anand Muthusamy and Henry Lester, supported by the Chen Innovator Grant, mice were administered high doses of the opiate drug Fentanyl before being tested in the same paradigm as the other binary maze experiments described here. This data is highly preliminary, but there is at least an intriguing similarity between the early trajectories of acortical animals and those of animals under opiate intoxication in that these animals excessively repeat actions and movements, failing to sufficiently cover the environment. Intoxicated animals may be no longer thirsty, unable to generate sufficient choice variability to explore the maze, unable to remember prior actions, or may lose interest in spatial novelty, among myriad possibilities. Further experiments are necessary to support one of these interpretations over others.

Bibliography

[1] Exact and approximate distances in graphs—a survey. In *Algorithms—ESA 2001*, pages 33–48, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

[2] F. Aboitiz and J. F. Montiel. Olfaction, navigation, and the origin of isocortex. *Frontiers in Neuroscience*, 9, 2015.

[3] C. B. Alme, C. Miao, K. Jezek, A. Treves, E. I. Moser, and M.-B. Moser. Place cells in the hippocampus: Eleven maps for eleven rooms. *Proceedings of the National Academy of Sciences*, 111(52):18428–18435, Dec. 2014.

[4] A. Alonso, J. van der Meij, D. Tse, and L. Genzel. Naïve to expert: Considering the role of previous knowledge in memory. *Brain and Neuroscience Advances*, 4:1–17, Jan. 2020.

[5] A. Alvernhe, E. Save, and B. Poucet. Local remapping of place cell firing in the Tolman detour task. *The European Journal of Neuroscience*, 33(9):1696–1705, May 2011.

[6] S. Alyan and B. McNaughton. Hippocampectomized rats are capable of homing by path integration. *Behavioral neuroscience*, 113(1):19, 1999.

[7] D. Amaral and M. Witter. The three-dimensional organization of the hippocampal formation: A review of anatomical data. *Neuroscience*, 31(3):571–591, 1989.

[8] D. J. Amit, H. Gutfreund, and H. Sompolinsky. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters*, 55(14):1530–1533, 1985.

[9] D. Aronov, R. Nevers, and D. W. Tank. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719–722, 2017.

[10] Y. Aso, D. Sitaraman, T. Ichinose, K. R. Kaun, K. Vogt, G. Belliart-Guerin, P. Y. Placais, A. A. Robie, N. Yamagata, C. Schnaitmann, W. J. Rowell, R. M. Johnston, T. T. Ngo, N. Chen, W. Korff, M. N. Nitabach, U. Heberlein, T. Preat, K. M. Branson, H. Tanimoto, and G. M. Rubin. Mushroom body output neurons encode valence and guide memory-based action selection in Drosophila. *Elife*, 3:e04580, 2014.

[11] H. E. Atallah, A. D. McCool, M. W. Howe, and A. M. Graybiel. Neurons in the ventral striatum exhibit cell-type-specific representations of outcome during learning. *Neuron*, 82(5):1145–1156, 2013.

[12] L. Avens and K. J. Lohmann. Use of multiple orientation cues by juvenile loggerhead sea turtles caretta caretta. *Journal of Experimental Biology*, 206(23):4317–4325, 2003.

[13] A. Badiani, D. Belin, D. Epstein, D. Calu, and Y. Shaham. Opiate versus psychostimulant addiction: the differences do matter. *Nature Reviews Neuroscience*, 12(11):685–700, 2011.

[14] K. L. Baker, M. Dickinson, T. M. Findley, D. H. Gire, M. Louis, M. P. Suver, J. V. Verhagen, K. I. Nagel, and M. C. Smear. Algorithms for olfactory search across species. *The Journal of Neuroscience*, 38(44):9383–9389, Oct. 2018.

[15] A. Banino, C. Barry, B. Uria, C. Blundell, T. Lillicrap, P. Mirowski, A. Pritzel, M. J. Chadwick, T. Degris, J. Modayil, et al. Vector-based navigation using grid-like representations in artificial agents. *Nature*, 557(7705):429–433, 2018.

[16] J. S. Barlow. Inertial navigation as a basis for animal navigation. *Journal of Theoretical Biology*, 6(1):76–117, 1964.

[17] A. G. Barto, R. S. Sutton, and C. W. Anderson. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(5):834–846, 1983.

[18] T. E. J. Behrens, T. H. Muller, J. C. R. Whittington, S. Mark, A. B. Baram, K. L. Stachenfeld, and Z. Kurth-Nelson. What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, 100(2):490–509, Oct. 2018.

[19] C. C. Bell, V. Han, and N. B. Sawtell. Cerebellum-like structures and their implications for cerebellar function. *Annual Review of Neuroscience*, 31:1–24, 2008.

[20] R. Bellman. A Markovian Decision Process. *Indiana University Mathematics Journal*, 6(4):679–684, 1957.

[21] R. Bellman. Dynamic programming. *Science*, 1966.

[22] A. T. Bennett. Do animals have cognitive maps? *The journal of experimental biology*, 199(1):219–224, 1996.

[23] H. C. Berg. A physicist looks at bacterial chemotaxis. *Cold Spring Harb Symp Quant Biol*, 53 Pt 1:1–9, 1988.

[24] E. A. Berker, A. H. Berker, and A. Smith. Translation of broca's 1865 report: Localization of speech in the third left frontal convolution. *Archives of neurology*, 43(10):1065–1072, 1986.

[25] D. E. Berlyne. *Conflict, Arousal, and Curiosity*. McGraw-Hill Book Company, New York, NY, US, 1960.

[26] G.-q. Bi and M.-m. Poo. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of neuroscience*, 18(24):10464–10472, 1998.

[27] M. E. Bitterman, R. Menzel, A. Fietz, and S. Schäfer. Classical conditioning of proboscis extension in honeybees (Apis mellifera). *Journal of Comparative Psychology*, 97(2):107–119, 1983.

[28] K. C. Bittner, A. D. Milstein, C. Grienberger, S. Romani, and J. C. Magee. Behavioral time scale synaptic plasticity underlies CA1 place fields. *Science*, 357(6355):1033–1036, 2017.

[29] A. P. Blaisdell, J. C. Denniston, and R. R. Miller. Temporal encoding as a determinant of overshadowing. *Journal of Experimental Psychology: Animal Behavior Processes*, 24(1):72–83, 1998.

[30] A. P. Blaisdell, L. M. Gunther, and R. R. Miller. Recovery from blocking achieved by extinguishing the blocking CS. *Animal Learning & Behavior*, 27(1):63–76, 1999.

[31] A. P. Blaisdell, R. R. Miller, A. S. Bristol, and L. M. Gunther. Overshadowing and latent inhibition counteract each other: Support for the comparator hypothesis. *Journal of Experimental Psychology: Animal Behavior Processes*, 24(3):335–351, 1998.

[32] T. V. Bliss and T. Lømo. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of physiology*, 232(2):331–356, 1973.

[33] Y.-L. Boureau and P. Dayan. Opponency revisited: Competition and co-operation between dopamine and serotonin. *Neuropsychopharmacology*, 36(1):74–97, 2011.

[34] R. Bourtchuladze, B. Frenguelli, J. Blendy, D. Cioffi, G. Schutz, and A. J. Silva. Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. *Cell*, 79(1):59–68, Oct. 1994.

[35] E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, and K. Deisseroth. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neuroscience*, 8(9):1263–1268, 2005.

[36] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang. JAX: composable transformations of Python+NumPy programs. http://github.com/google/jax, 2018.

[37] L. A. Bradfield, A. Dezfouli, M. van Holstein, B. Chieng, and B. W. Balleine. Medial orbitofrontal cortex mediates outcome retrieval in partially observable task situations. *Neuron*, 88(6):1268–1280, 2015.

[38] P. A. Brennan and E. B. Keverne. Neural mechanisms of mammalian olfactory learning. *Progress in Neurobiology*, 51(4):457–481, Mar. 1997.

[39] K. Brodmann. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth, 1909.

[40] H. M. Bruce. An exteroceptive block to pregnancy in the mouse. *Nature*, 184:105, July 1959.

[41] B. Bruinsma, H. Terra, S. de Kloet, A. Luchicchi, A. Timmerman, E. Remmelink, M. Loos, T. Pattij, and H. D. Mansvelder. An automated home-cage-based 5-choice serial reaction time task for rapid assessment of attention and impulsivity in rats. *Psychopharmacology*, 236(7):2015–2026, 2019.

[42] C. Buehlmann, B. Wozniak, R. Goulard, B. Webb, P. Graham, and J. E. Niven. Mushroom bodies are required for learned visual navigation, but not for innate visual behavior, in ants. *Current biology: CB*, 30(17):3438–3443.e2, Sept. 2020.

[43] J. Buel. The linear maze. I. "Choice-point expectancy," "correctness," and the goal gradient. *Journal of Comparative Psychology*, 17(2):185–199, 1934.

[44] M. Bunsey and H. Eichenbaum. Conservation of hippocampal memory function in rats and humans. *Nature*, 379(6562):255–257, 1996.

[45] Y. Burak and I. R. Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLoS computational biology*, 5(2):e1000291, 2009.

[46] C. P. Burgess, A. Lak, N. A. Steinmetz, P. Zatka-Haas, C. B. Reddy, E. A. Jacobs, J. F. Linden, J. J. Paton, A. Ranson, S. Schröder, S. Soares, M. J. Wells, L. E. Wool, K. D. Harris, and M. Carandini. High-yield methods for accurate two-alternative visual psychophysics in head-fixed mice. *Cell Reports*, 20(10):2513–2524, 2017.

[47] N. Burgess, M. Recce, and J. O'Keefe. A model of hippocampal function. *Neural networks*, 7(6-7):1065–1081, 1994.

[48] M. Carandini and A. K. Churchland. Probing perceptual decisions in rodents. *Nature Neuroscience*, 16:824–31, July 2013.

[49] R. N. Cardinal, J. A. Parkinson, G. Lachenal, K. M. Halkerston, N. Rudarakanchana, J. Hall, C. H. Morrison, S. R. Howes, T. W. Robbins, and B. J. Everitt. Effects of selective excitotoxic lesions of the nucleus accumbens core, anterior cingulate cortex, and central nucleus of the amygdala on autoshaping performance in rats. *Behavioral neuroscience*, 116(4):553, 2002.

[50] R. Chaudhuri and I. Fiete. Computational principles of memory. *Nature Neuroscience*, 19(3):394–403, 2016.

[51] T. A. Cleland, V. A. Narla, and K. Boudadi. Multiple learning parameters differentially regulate olfactory generalization. *Behavioral Neuroscience*, 123(1):26–35, Feb. 2009.

[52] J. Y. Cohen, M. W. Amoroso, and N. Uchida. Serotonergic neurons signal reward and punishment on multiple timescales. *eLife*, 4:e06346, 2015.

[53] T. S. Collett and M. Collett. Memory use in insect visual navigation. *Nature Reviews Neuroscience*, 3(7):542–552, July 2002.

[54] D. S. Corneil and W. Gerstner. Attractor network dynamics enable preplay and rapid path planning in maze–like environments. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[55] C. J. Cueva and X.-X. Wei. Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. *arXiv preprint arXiv:1803.07770*, 2018.

[56] T. Danjo, T. Toyoizumi, and S. Fujisawa. Spatial representations of self and other in the hippocampus. *Science*, 359(6372):213–218, 2018.

[57] C. Darwin. Origin of certain instincts. *Nature*, 7:417–418, 1873.

[58] N. D. Daw, S. Kakade, and P. Dayan. Opponent interactions between serotonin and dopamine. *Neural Networks*, 15(4-6):603–616, 2002.

[59] P. Dayan. Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, 5(4):613–624, July 1993.

[60] P. Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Computational Neuroscience. MIT Press, Cambridge, Mass., 2001.

[61] K. Doya. Modulators of decision making. *Nature Neuroscience*, 11(4):410–6, 2008.

[62] C. A. Duan, J. C. Erlich, and C. D. Brody. Requirement of prefrontal and midbrain regions for rapid executive control of behavior in the rat. *Neuron*, 86(6):1491–1503, 2015.

[63] P. A. Dudchenko, E. R. Wood, and H. Eichenbaum. Neurotoxic hippocampal lesions have no effect on odor span and little effect on odor recognition memory but produce significant impairments on spatial span, recognition, and alternation. *Journal of Neuroscience*, 20(8):2964–2977, 2000.

[64] D. Elbers, M. Bulte, F. Bairlein, H. Mouritsen, and D. Heyers. Magnetic activation in the brain of the migratory northern wheatear (oenanthe oenanthe). *Journal of Comparative Physiology A*, 203(8):591–600, 2017.

[65] J. Epsztein, M. Brecht, and A. K. Lee. Intracellular determinants of hippocampal CA1 place and silent cell activity in a novel environment. *Neuron*, 70:109–20, Apr. 2011.

[66] W. Estes. The problem of inference from curves based on group data. *Psychological Bulletin*, 53(2):134–140, 1956.

[67] A. Etienne. The control of short-distance homing in the golden hamster. In *Cognitive processes and spatial orientation in animal and man*, pages 233–251. Springer, 1987.

[68] A. S. Etienne and K. J. Jeffery. Path integration in mammals. *Hippocampus*, 14(2):180–192, 2004.

[69] A. S. Etienne, R. Maurer, F. Saucy, and E. Teroni. Short-distance homing in the golden hamster after a passive outward journey. *Animal Behaviour*, 34(3):696–715, 1986.

[70] D. Evans, A. Stempel, R. Vale, S. Ruehle, Y. Lefler, and T. Branco. A synaptic threshold mechanism for computing escape decisions. *Nature*, 558(7711):590–594, 2018.

[71] M. Fanselow and R. Bolles. Naloxone and shock-elicited freezing in the rat. *Journal of comparative and physiological psychology*, 93:736–44, Sept. 1979.

[72] S. M. Farris. Are mushroom bodies cerebellum-like structures? *Arthropod Struct Dev*, 40:368–79, July 2011.

[73] C. D. Fast and A. P. Blaisdell. Rats are sensitive to ambiguity. *Psychonomic Bulletin & Review*, 18(6):1230–1237, 2011.

[74] C. B. Ferster and B. F. Skinner. Schedules of reinforcement. *Appleton-Century-Crofts*, 1957.

[75] R. C. Fijn, D. Hiemstra, R. A. Phillips, and J. v. d. Winden. Arctic terns sterna paradisaea from the Netherlands migrate record distances across three oceans to Wilkes Land, East Antarctica. *Ardea*, 101(1):3–12, 2013.

[76] G. Fleissner, E. Holtkamp-Rötzler, M. Hanzlik, M. Winklhofer, G. Fleissner, N. Petersen, and W. Wiltschko. Ultrastructural analysis of a putative magnetoreceptor in the beak of homing pigeons. *Journal of Comparative Neurology*, 458(4):350–360, 2003.

[77] S. B. Floresco. The nucleus accumbens: An interface between cognition, emotion, and action. *Annual review of psychology*, 66:25–52, 2015.

[78] S. B. Floresco, S. Ghods-Sharifi, C. Vexelman, and O. Magyar. Dissociable roles for the nucleus accumbens core and shell in regulating set shifting. *Journal of Neuroscience*, 26(9):2449–2457, 2006.

[79] R. W. Floyd. Algorithm 97: Shortest path. *Communications of the ACM*, 5(6):345, June 1962.

[80] E. Fonio, Y. Benjamini, and I. Golani. Freedom of movement and the stability of its unfolding in free exploration of mice. *Proceedings of the National Academy of Sciences of the United States of America*, 106(50):21335–21340, Dec. 2009.

[81] N. J. Fortin, K. L. Agster, and H. B. Eichenbaum. Critical role of the hippocampus in memory for sequences of events. *Nature neuroscience*, 5(5):458–462, 2002.

[82] D. Foster, R. Morris, and P. Dayan. A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus*, 10(1):1–16, 2000.

[83] D. J. Foster and M. A. Wilson. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084):680–683, 2006.

[84] L. M. Frank, G. B. Stanley, and E. N. Brown. Hippocampal plasticity across multiple days of exposure to novel environments. *Journal of Neuroscience*, 24(35):7681–7689, Sept. 2004.

[85] A. Friedman, D. Homma, L. G. Gibb, K.-i. Amemori, S. J. Rubin, A. S. Hood, M. H. Riad, and A. M. Graybiel. A corticostriatal path targeting striosomes controls decision-making under conflict. *Cell*, 161(6):1320–1333, 2015.

[86] C. R. Gallistel, S. Fairhurst, and P. Balsam. The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36):13124–13131, Sept. 2004.

[87] M. Galtier, O. Faugeras, and P. Bressloff. Hebbian learning of recurrent connections: A geometrical perspective. *Neural computation*, 24:2346–83, May 2012.

[88] R. J. Gardner, E. Hermansen, M. Pachitariu, Y. Burak, N. A. Baas, B. A. Dunn, M.-B. Moser, and E. I. Moser. Toroidal topology of population activity in grid cells. *Nature*, pages 1–6, 2022.

[89] J. L. Gauthier and D. W. Tank. A dedicated population for reward coding in the hippocampus. *Neuron*, 99(1):179–193, 2018.

[90] J. P. Geerts, F. Chersi, K. L. Stachenfeld, and N. Burgess. A general model of hippocampal and dorsal striatal learning and decision making. *Proceedings of the National Academy of Sciences*, 117(49):31427–31437, Dec. 2020.

[91] S. J. Gershman. The successor representation: Its computational logic and neural substrates. *The Journal of Neuroscience*, 38(33):7193–7200, 2018.

[92] S. J. Gershman and N. D. Daw. Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology*, 68(1):1–28, 2015.

[93] M. Gil, M. Ancau, M. I. Schlesiger, A. Neitz, K. Allen, R. J. De Marco, and H. Monyer. Impaired path integration in mice with disrupted grid cell firing. *Nature neuroscience*, 21(1):81–91, 2018.

[94] A. K. Gillespie, D. A. A. Maya, E. L. Denovellis, D. F. Liu, D. B. Kastner, M. E. Coulter, D. K. Roumis, U. T. Eden, and L. M. Frank. Hippocampal replay reflects specific past experiences rather than a plan for subsequent choice. *Neuron*, 109(19):3149–3163.e6, 2021.

[95] K. Goto, R. Kurashima, and S. Watanabe. Delayed matching-to-position performance in C57BL/6N mice. *Behavioural processes*, 84(2):591–597, 2010.

[96] A. M. Graybiel. Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, 31(1):359–387, 2008.

[97] M.-C. Grobéty and F. Schenk. Spatial learning in a three-dimensional maze. *Animal Behaviour*, 43(6):1011–1020, June 1992.

[98] Z. V. Guo, N. Li, D. Huber, E. Ophir, D. Gutnisky, J. T. Ting, G. Feng, and K. Svoboda. Flow of cortical activity underlying a tactile decision in mice. *Neuron*, 81(1):179–194, Jan. 2014.

[99] D. Ha and J. Schmidhuber. Recurrent world models facilitate policy evolution. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[100] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005.

[101] B. Halbout, A. T. Marshall, A. Azimi, M. Liljeholm, S. V. Mahler, K. M. Wassum, and S. B. Ostlund. Mesolimbic dopamine projections mediate cue-motivated reward seeking but not reward retrieval in rats. *eLife*, 8:e43551, 2019.

[102] J. M. Harlow. Recovery from the passage of an iron bar through the head. *History of Psychiatry*, 4(14):274–281, 1993.

[103] D. O. Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 2005.

[104] M. Heisenberg. Mushroom body memoir: From maps to models. *Nature Reviews Neuroscience*, 4(4):266–275, Apr. 2003.

[105] J. G. Heys and D. A. Dombeck. Evidence for a subcircuit in medial entorhinal cortex representing elapsed time during immobility. *Nature neuroscience*, 21(11):1574–1582, 2018.

[106] P. C. Holland. Temporal determinants of occasion setting in feature-positive discriminations. *Animal Learning & Behavior*, 14(2):111–120, 1986.

[107] S. A. Hollup, S. Molden, J. G. Donnett, M.-B. Moser, and E. I. Moser. Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *Journal of Neuroscience*, 21(5):1635–1644, Mar. 2001.

[108] Y. K. Hong, C. O. Lacefield, C. C. Rodgers, and R. M. Bruno. Sensation, movement and learning in the absence of barrel cortex. *Nature*, 561(7724):542–546, 2018.

[109] P. J. Hore and H. Mouritsen. The radical-pair mechanism of magnetoreception. *Annual review of biophysics*, 45:299–344, 2016.

[110] J. L. Hoy, I. Yavorska, M. Wehr, and C. M. Niell. Vision drives accurate approach behavior during prey capture in laboratory mice. *Current Biology*, 26(22):3046–3052, 2016.

[111] R. N. Hughes. Intrinsic exploration in animals: Motives and measurement. *Behavioural Processes*, 41(3):213–226, Dec. 1997.

[112] K. Iigaya, M. S. Fonseca, M. Murakami, Z. F. Mainen, and P. Dayan. An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. *Nature Communications*, 9(1):2477, 2018.

[113] N. Ishizuka, J. Weber, and D. G. Amaral. Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *Journal of Comparative Neurology*, 295(4):580–623, 1990.

[114] E. M. Izhikevich. Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, 17(10):2443–2452, 2007.

[115] L. F. Jacobs. From chemotaxis to the cognitive map: The function of olfaction. *Proceedings of the National Academy of Sciences*, 109:10693–10700, June 2012.

[116] S. P. Jadhav, C. Kemere, P. W. German, and L. M. Frank. Awake hippocampal sharp-wave ripples support spatial memory. *Science*, 336(6087):1454–1458, 2012.

[117] M. M. Jankowski, M. N. Islam, N. F. Wright, S. D. Vann, J. T. Erichsen, J. P. Aggleton, and S. M. O'Mara. Nucleus reuniens of the thalamus contains head direction cells. *Elife*, 3:e03075, 2014.

[118] M. M. Jankowski, J. Passecker, M. N. Islam, S. Vann, J. T. Erichsen, J. P. Aggleton, and S. M. O'Mara. Evidence for spatially-responsive neurons in the rostral thalamus. *Frontiers in Behavioral Neuroscience*, 9:256, 2015.

[119] A. Jeewajee, C. Barry, V. Douchamps, D. Manson, C. Lever, and N. Burgess. Theta phase precession of grid and place cell firing in open environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1635):20120532, 2014.

[120] O. Jensen and J. E. Lisman. Position reconstruction from an ensemble of hippocampal place cells: Contribution of theta phase coding. *Journal of neurophysiology*, 83(5):2602–2609, 2000.

[121] R. Jensen. Behaviorism, latent learning, and cognitive maps: Needed revisions in introductory psychology textbooks. *The Behavior Analyst*, 29(2):187–209, 2006.

[122] A. J. Kalmijn. The electric sense of sharks and rays. *Journal of Experimental Biology*, 55(2):371–383, 1971.

[123] A. J. Kalmijn. Electric and magnetic field detection in elasmobranch fishes. *Science*, 218(4575):916–918, 1982.

[124] T. Karigo, A. Kennedy, B. Yang, M. Liu, D. Tai, I. A. Wahle, and D. J. Anderson. Distinct hypothalamic control of same- and opposite-sex mounting behaviour in mice. *Nature*, 589(7841):258–263, 2021.

[125] M. P. Karlsson and L. M. Frank. Awake replay of remote experiences in the hippocampus. *Nature neuroscience*, 12(7):913–918, 2009.

[126] R. Kawai, T. Markman, R. Poddar, R. Ko, A. L. Fantana, A. K. Dhawale, A. R. Kampff, and B. P. Ölveczky. Motor cortex is required for learning but not for executing a motor skill. *Neuron*, 86(3):800–812, 2015.

[127] K. Kay, J. E. Chung, M. Sosa, J. S. Schor, M. P. Karlsson, M. C. Larkin, D. F. Liu, and L. M. Frank. Constant sub-second cycling between representations of possible futures in the hippocampus. *Cell*, 180(3):552–567, 2020.

[128] S. Kim, M. K. Lehtinen, A. Sessa, M. W. Zappaterra, S.-H. Cho, D. Gonzalez, B. Boggan, C. A. Austin, J. Wijnholds, M. J. Gambello, J. Malicki, A. S. LaMantia, V. Broccoli, and C. A. Walsh. The apical complex couples cell fate and cell survival to cerebral cortical development. *Neuron*, 66(1):69–84, 2010.

[129] K. B. Kjelstrup, T. Solstad, V. H. Brun, T. Hafting, S. Leutgeb, M. P. Witter, E. I. Moser, and M.-B. Moser. Finite scale of spatial representation in the hippocampus. *Science (New York, N.Y.)*, 321(5885):140–143, July 2008.

[130] M. Knaden and P. Graham. The sensory ecology of ant navigation: From natural environments to neural mechanisms. In *Annual Review of Entomology, Vol 61*, volume 61, pages 63–76. 2016.

[131] B. Kranstauber, R. Weinzierl, M. Wikelski, and K. Safi. Global aerial flyways allow efficient travelling. *Ecology Letters*, 18(12):1338–1345, 2015.

[132] B. J. Kraus, M. P. Brandon, R. J. Robinson II, M. A. Connerney, M. E. Hasselmo, and H. Eichenbaum. During running in place, grid cells integrate elapsed time and distance run. *Neuron*, 88(3):578–589, 2015.

[133] I. Krechevsky. "Hypotheses" in rats. *Psychological Review*, 39(6):516–532, 1932.

[134] K. S. Lashley. Visual discrimination of size and form in the albino rat. *Journal of Animal Behavior*, 2(5):310–331, 1912.

[135] K. S. Lashley. Mass action in cerebral function. *Science*, 73(1888):245–254, 1931.

[136] K. S. Lashley. Integrative functions of the cerebral cortex. *Physiological Reviews*, 13(1):1–42, 1933.

[137] K. S. Lashley. 1. in search of the engram. In *Brain physiology and psychology*, pages 1–32. University of California Press, 2020.

[138] K. S. Lashley and J. Ball. Spinal conduction and kinesthetic sensitivity in the maze habit. *Journal of Comparative Psychology*, 9(1):71, 1929.

[139] J. E. LeDoux. Emotion circuits in the brain. *Annual Review of Neuroscience*, 23:155–184, 2000.

[140] A. K. Lee and M. A. Wilson. Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, 36(6):1183–1194, 2002.

[141] S. W. S. Lee and N. Schwarz. Bidirectionality, mediation, and moderation of metaphorical effects: The embodiment of social suspicion and fishy smells. *Journal of Personality and Social Psychology*, 103(5):737–749, Nov. 2012.

[142] F. Li, J. W. Lindsey, E. C. Marin, N. Otto, M. Dreher, G. Dempsey, I. Stark, A. S. Bates, M. W. Pleijzier, P. Schlegel, A. Nern, S.-y. Takemura, N. Eckstein, T. Yang, A. Francis, A. Braun, R. Parekh, M. Costa, L. K. Scheffer, Y. Aso, G. S. Jefferis, L. F. Abbott, A. Litwin-Kumar, S. Waddell, and G. M. Rubin. The connectome of the adult drosophila mushroom body provides insights into function. *eLife*, 9:e62576, Dec. 2020.

[143] Y. Li, W. Zhong, D. Wang, Q. Feng, Z. Liu, J. Zhou, C. Jia, F. Hu, J. Zeng, Q. Guo, et al. Serotonin neurons in the dorsal raphe nucleus encode reward signals. *Nature communications*, 7(1):1–15, 2016.

[144] D. Lin, M. P. Boyle, P. Dollar, H. Lee, P. Perona, E. S. Lein, and D. J. Anderson. Functional identification of an aggression locus in the mouse hypothalamus. *Nature*, 470(7333):221–226, 2011.

[145] Z. Liu, J. Zhou, Y. Li, F. Hu, Y. Lu, M. Ma, Q. Feng, J.-e. Zhang, D. Wang, J. Zeng, et al. Dorsal raphe neurons signal reward through 5-HT and glutamate. *Neuron*, 81(6):1360–1374, 2014.

[146] K. Louie and M. A. Wilson. Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, 29(1):145–156, 2001.

[147] E. V. Lubenov and A. G. Siapas. Hippocampal theta oscillations are travelling waves. *Nature*, 459(7246):534–539, 2009.

[148] C. J. MacDonald, K. Q. Lepage, U. T. Eden, and H. Eichenbaum. Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, 71(4):737–749, 2011.

[149] J. C. Magee and C. Grienberger. Synaptic Plasticity Forms and Functions. *Annual Review of Neuroscience*, 43(1):95–117, 2020.

[150] M. Malvaez, C. Shieh, M. D. Murphy, V. Y. Greenfield, and K. M. Wassum. Distinct cortical–amygdala projections drive reward value encoding and retrieval. *Nature Neuroscience*, 22(5):762–769, 2019.

[151] N. Martiros, A. A. Burgess, and A. M. Graybiel. Inversely active striatal projection neurons and interneurons selectively delimit useful behavioral sequences. *Current Biology*, 2017.

[152] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9):1281–1289, 2018.

[153] S. Matias, E. Lottem, G. P. Dugué, and Z. F. Mainen. Activity patterns of serotonin neurons underlying cognitive flexibility. *Elife*, 6:e20552, 2017.

[154] K. McBride, B. Slotnick, and F. L. Margolis. Does intranasal application of zinc sulfate produce anosmia in the mouse? An olfactometric and anatomical study. *Chemical Senses*, 28(8):659–670, 2003.

[155] C. G. McNamara, Á. Tejero-Cantero, S. Trouche, N. Campo-Urriza, and D. Dupret. Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. *Nature Neuroscience*, 17(12):1658–1660, Dec. 2014.

[156] B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M.-B. Moser. Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, 7(8):663–678, 2006.

[157] A. M. Michaiel, E. T. Abe, and C. M. Niell. Dynamics of gaze control during prey capture in freely moving mice. *eLife*, 9:e57458, 2020.

[158] R. Miles and R. K. S. Wong. Single neurones can initiate synchronized population discharge in the hippocampus. *Nature*, 306(5941):371–373, 1983.

[159] K. J. Miller, M. M. Botvinick, and C. D. Brody. Dorsal hippocampus contributes to model-based planning. *Nature Neuroscience*, 20(9):nn.4613, 2017.

[160] B. Milner, S. Corkin, and H.-L. Teuber. Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of hm. *Neuropsychologia*, 6(3):215–234, 1968.

[161] H. Mittelstaedt and M.-L. Mittelstaedt. Homing by path integration. In *Avian navigation*, pages 290–297. Springer, 1982.

[162] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller. Playing atari with deep reinforcement learning. *arXiv*, 2013.

[163] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015.

[164] T. M. Moerland, J. Broekens, and C. M. Jonker. Model-based reinforcement learning: A survey. *arXiv preprint arXiv:2006.16712*, 2020.

[165] A. Möller, S. Sagasser, W. Wiltschko, and B. Schierwater. Retinal cryptochrome in a migratory passerine bird: a possible transducer for the avian magnetic compass. *Naturwissenschaften*, 91(12):585–588, 2004.

[166] I. Momennejad, E. M. Russek, J. H. Cheong, M. M. Botvinick, N. D. Daw, and S. J. Gershman. The successor representation in human reinforcement learning. *Nature Human Behaviour*, 1(9):680–692, 2017.

[167] M. Morales and E. B. Margolis. Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. *Nature Reviews Neuroscience*, 18(2):73–85, 2017.

[168] R. G. M. Morris, P. Garrud, J. N. P. Rawlins, and J. O'Keefe. Place navigation impaired in rats with hippocampal lesions. *Nature*, 297(5868):681–683, June 1982.

[169] M.-B. Moser, D. C. Rowland, and E. I. Moser. Place cells, grid cells, and memory. *Cold Spring Harbor Perspectives in Biology*, 7(2):a021808, Feb. 2015.

[170] M. Müller and R. Wehner. Path integration in desert ants, Cataglyphis fortis. *Proceedings of the National Academy of Sciences*, 85(14):5287–5290, July 1988.

[171] R. U. Muller, M. Stead, and J. Pach. The hippocampus as a cognitive graph. *The Journal of General Physiology*, 107(6):663–694, June 1996.

[172] N. L. Munn. The learning process. In *Handbook of Psychological Research on the Rat; an Introduction to Animal Psychology*, pages 226–288. Houghton Mifflin, Oxford, England, 1950.

[173] N. L. Munn. The role of sensory processes in maze behavior. In *Handbook of Psychological Research on the Rat; an Introduction to Animal Psychology*, pages 181–225. Houghton Mifflin, Oxford, England, 1950.

[174] M. Müller and R. Wehner. Path integration in desert ants, cataglyphis fortis. *Proceedings of the National Academy of Sciences*, 85(14):5287–5290, 1988.

[175] M. Nagy, A. Horicsányi, E. Kubinyi, I. D. Couzin, G. Vásárhelyi, A. Flack, and T. Vicsek. Synergistic benefits of group search in rats. *Current Biology*, Sept. 2020.

[176] T. Nath, A. Mathis, A. C. Chen, A. Patel, M. Bethge, and M. W. Mathis. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, 14(7):2152–2176, July 2019.

[177] E. J. Nestler. Is there a common molecular pathway for addiction? *Nature Neuroscience*, 8(11):1445–1449, 2005.

[178] W. T. Newsome and E. B. Pare. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, 8(6):2201–2211, June 1988.

[179] S. A. Ocko, K. Hardcastle, L. M. Giocomo, and S. Ganguli. Emergent elasticity in the neural code for space. *Proceedings of the National Academy of Sciences*, 115(50):E11798–E11806, 2018.

[180] E. Oja. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3):267–273, 1982.

[181] J. O'Keefe and N. Burgess. Geometric determinants of the place fields of hippocampal neurons. *Nature*, 381(6581):425–428, 1996.

[182] J. O'Keefe and J. Dostrovsky. The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 1971.

[183] J. O'Keefe and M. L. Recce. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, 3(3):317–330, 1993.

[184] D. Olton. Mazes, maps, and memory. *American Psychologist*, 34(7):583–596, 1979.

[185] D. B. Omer, S. R. Maimon, L. Las, and N. Ulanovsky. Social place-cells in the bat hippocampus. *Science*, 359(6372):218–224, 2018.

[186] T. M. Otchy, S. B. Wolff, J. Y. Rhee, C. Pehlevan, R. Kawai, A. Kempf, S. M. Gobes, and B. P. Ölveczky. Acute off-target effects of neural circuit manipulations. *Nature*, 528(7582):358–363, 2015.

[187] A. E. Papale, M. C. Zielinski, L. M. Frank, S. P. Jadhav, and A. D. Redish. Interplay between hippocampal sharp-wave-ripple events and vicarious trial and error behaviors in decision making. *Neuron*, 92(5):975–982, 2016.

[188] I. P. Pavlov. Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Translated GV Anrep. London: Oxford University Press*, 1927.

[189] S. Pisupati, L. Chartarifsky-Lynn, A. Khanal, and A. K. Churchland. Lapses in perceptual decisions reflect exploration. *eLife*, 10:e55490, Jan. 2021.

[190] Plato. The apology. ca 400 BCE.

[191] Pseudo-Apollodorus. Epitome. In *Library and Epitome*, page Ch 1 Sec 9. I-II Century AD.

[192] J. Qi, S. Zhang, H.-L. L. Wang, D. J. Barker, J. Miranda-Barrientos, and M. Morales. VTA glutamatergic inputs to nucleus accumbens drive aversion by acting on GABAergic interneurons. *Nature neuroscience*, 19(5):725–33, 2016.

[193] S. P. Ranade and Z. F. Mainen. Transient firing of dorsal raphe neurons encodes diverse and specific sensory, motor, and reward events. *Journal of neurophysiology*, 102(5):3026–3037, 2009.

[194] G. Reddy. Reinforcement waves as a mechanism for discontinuous learning. *bioRxiv*, 2022.

[195] A. D. Redish. Vicarious trial and error. *Nature reviews. Neuroscience*, 17(3):147–159, Mar. 2016.

[196] A. D. Redish and D. S. Touretzky. The role of the hippocampus in solving the Morris water maze. *Neural Computation*, 10(1):73–111, Jan. 1998.

[197] S. M. Reppert, P. A. Guerra, and C. Merlin. Neurobiology of monarch butterfly migration. *Annual review of entomology*, 61, 2016.

[198] R. A. Rescorla. Reduction in the effectiveness of reinforcement after prior excitatory conditioning. *Learning and Motivation*, 1(4):372–381, 1970.

[199] R. A. Rescorla. Behavioral studies of pavlovian conditioning. *Annual Review of Neuroscience*, 11(1):329–352, 1988.

[200] S. M. Reynolds and K. C. Berridge. Positive and negative motivation in nucleus accumbens shell: Bivalent rostrocaudal gradients for GABA-elicited eating, taste “liking”/“disliking” reactions, place preference/avoidance, and fear. *The Journal of Neuroscience*, 22(16):7308–7320, 2002.

[201] T. Ritz, M. Ahmad, H. Mouritsen, R. Wiltschko, and W. Wiltschko. Photoreceptor-based magnetoreception: optimal design of receptor molecules, cells, and neuronal processing. *Journal of the Royal Society Interface*, 7(suppl_2):S135–S146, 2010.

[202] L. Rondi-Reig, G. H. Petit, C. Tobin, S. Tonegawa, J. Mariani, and A. Berthoz. Impaired sequential egocentric and allocentric memories in forebrain-specific-NMDA receptor knock-out mice during a new task dissociating strategies of navigation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 26(15):4071–4081, Apr. 2006.

[203] D. H. Root, D. J. Estrin, and M. Morales. Aversion or salience signaling by ventral tegmental area glutamate neurons. *iScience*, 2:51–62, 2018.

[204] M. Rosenberg, T. Zhang, P. Perona, and M. Meister. Mice in a labyrinth exhibit rapid learning, sudden insight, and efficient exploration. *eLife*, 10:e66175, July 2021.

[205] A. E. Rosser and E. B. Keverne. The importance of central noradrenergic neurones in the formation of an olfactory memory in the prevention of pregnancy block. *Neuroscience*, 15(4):1141–1147, Aug. 1985.

[206] P. H. Rudebeck, M. E. Walton, A. N. Smyth, D. M. Bannerman, and M. F. Rushworth. Separate neural pathways process different decision costs. *Nature neuroscience*, 9(9):1161–1168, 2006.

[207] E. M. Russek, I. Momennejad, M. M. Botvinick, S. J. Gershman, and N. D. Daw. Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLOS Computational Biology*, 13(9):e1005768, 2017.

[208] A. Samsonovich and B. L. McNaughton. Path integration and cognitive mapping in a continuous attractor neural network model. *The Journal of Neuroscience*, 17(15):5900–5920, 1997.

[209] D. Santos-Pata and P. F. M. J. Verschure. Human vicarious trial and error is predictive of spatial navigation performance. *Frontiers in Behavioral Neuroscience*, 12:237, Oct. 2018.

[210] F. Sargolini, M. Fyhn, T. Hafting, B. L. McNaughton, M. P. Witter, M.-B. Moser, and E. I. Moser. Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*, 312(5774):758–762, 2006.

[211] N. Sato, C. Fujishita, and A. Yamagishi. To take or not to take the shortcut: Flexible spatial behaviour of rats based on cognitive map in a lattice maze. *Behavioural Processes*, 151:39–43, June 2018.

[212] B. T. Saunders and T. E. Robinson. The role of dopamine in the accumbens core in the expression of pavlovian-conditioned responses. *European Journal of Neuroscience*, 36(4):2521–2532, 2012.

[213] F. Savelli and J. J. Knierim. Origin and role of path integration in the cognitive representations of the hippocampus: computational insights into open questions. *Journal of Experimental Biology*, 222:jeb188912, 2019.

[214] K. Schmidt-Koenig and C. Walcott. Tracks of pigeons homing with frosted lenses. *Animal Behaviour*, 26:480–486, 1978.

[215] W. Schultz. Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80(1):1–27, 1998.

[216] W. Schultz, P. Dayan, and P. R. Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997.

[217] W. B. Scoville and B. Milner. Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(1):11, 1957.

[218] J. Seward. Bzip2, July 2019.

[219] T. Shokaku, T. Moriyama, H. Murakami, S. Shinohara, N. Manome, and K. Morioka. Development of an automatic turntable-type multiple T-maze device and observation of pill bug behavior. *Review of Scientific Instruments*, 91(10):104104, Oct. 2020.

[220] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. v. d. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[221] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. v. d. Driessche, T. Graepel, and D. Hassabis. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354, 2017.

[222] B. F. Skinner. Are theories of learning necessary? *Psychological review*, 57(4):193, 1950.

[223] W. S. Small. Experimental study of the mental processes of the rat. II. *The American Journal of Psychology*, 12(2):206–239, 1901.

[224] K. S. Smith and A. M. Graybiel. A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, 79(2):361–374, 2013.

[225] K. Soman, S. Chakravarthy, and M. M. Yartsev. A hierarchical anti-hebbian network model for the formation of spatial cells in three-dimensional space. *Nature communications*, 9(1):1–15, 2018.

[226] B. Sorscher, G. C. Mel, S. A. Ocko, L. Giocomo, and S. Ganguli. A unified theory for the computational and mechanistic origins of grid cells. *bioRxiv*, 2020.

[227] M. Sosa and L. M. Giocomo. Navigating for reward. *Nature Reviews Neuroscience*, pages 1–16, July 2021.

[228] K. L. Stachenfeld, M. Botvinick, and S. J. Gershman. Design principles of the hippocampal cognitive map. *Advances in neural information processing systems*, 27, 2014.

[229] K. L. Stachenfeld, M. M. Botvinick, and S. J. Gershman. The hippocampus as a predictive map. *Nature neuroscience*, 20(11):1643–1653, 2017.

[230] K. L. Stachenfeld, M. M. Botvinick, and S. J. Gershman. The hippocampus as a predictive map. *Nature Neuroscience*, 20(11):1643–1653, Nov. 2017.

[231] K. Steck, B. S. Hansson, and M. Knaden. Smells like home: Desert ants, cataglyphis fortis, use olfactory landmarks to pinpoint the nest. *Frontiers in Zoology*, 6(1):5, Feb. 2009.

[232] H. Stensola, T. Stensola, T. Solstad, K. Frøland, M.-B. Moser, and E. I. Moser. The entorhinal grid map is discretized. *Nature*, 492(7427):72–78, 2012.

[233] X. Sun, S. Yue, and M. Mangan. A decentralised neural model explaining optimal integration of navigational strategies in insects. *eLife*, 9:e54026, June 2020.

[234] Sutton and Barto. Time-derivative models of pavlovian reinforcement. *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, 1990.

[235] R. S. Sutton. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Machine learning proceedings 1990*, pages 216–224. Elsevier, 1990.

[236] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Nov. 2018.

[237] M. Tarsitano. Route selection by a jumping spider (portia labiata) during the locomotory phase of a detour. *Animal Behaviour*, 72(6):1437–1442, Dec. 2006.

[238] J. S. Taube. The head direction signal: origins and sensory-motor integration. *Annu. Rev. Neurosci.*, 30:181–207, 2007.

[239] J. S. Taube, R. U. Muller, and J. B. Ranck. Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *Journal of Neuroscience*, 10(2):420–435, 1990.

[240] J. R. Taylor and T. W. Robbins. Enhanced behavioural control by conditioned reinforcers following microinjections of d-amphetamine into the nucleus accumbens. *Psychopharmacology*, 84(3):405–412, 1984.

[241] J. R. Taylor and T. W. Robbins. 6-hydroxydopamine lesions of the nucleus accumbens, but not of the caudate nucleus, attenuate enhanced responding with reward-related stimuli produced by intra-accumbens d-amphetamine. *Psychopharmacology*, 90(3):390–397, 1986.

[242] O. Tchernichovski, Y. Benjamini, and I. Golani. The dynamics of long-term exploration in the rat. *Biological Cybernetics*, 78(6):423–432, July 1998.

[243] P. Thalau, T. Ritz, K. Stapput, R. Wiltschko, and W. Wiltschko. Magnetic compass orientation of migratory birds in the presence of a 1.315 MHz oscillating field. *Naturwissenschaften*, 92(2):86–90, 2005.

[244] D. Thistlethwaite. A critical review of latent learning and related experiments. *Psychological Bulletin*, 48(2):97–129, 1951.

[245] C. A. Thorn, H. Atallah, M. Howe, and A. M. Graybiel. Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron*, 66(5):781–795, 2010.

[246] E. L. Thorndike. Animal intelligence. *Nature*, 58(1504):390–390, 1898.

[247] N. Tinbergen. *The study of instinct*. Pygmalion Press, an imprint of Plunkett Lake Press, 2020.

[248] E. Tolman and C. Honzik. Degrees of hunger, reward and non-reward, and maze learning in rats. *University of California Publications in Psychology*, 4:241–256, 1930.

[249] E. C. Tolman. The determiners of behavior at a choice point. *Psychological Review*, 45:1–41, 1938.

[250] E. C. Tolman. Cognitive maps in rats and men. *Psychological Review*, 55(4):189–208, 1948.

[251] Z. Turan. *Life without cortex: Subcortical circuits in naturalistic behaviors*. PhD thesis, California Institute of Technology, 2021.

[252] N. Uchida and Z. F. Mainen. Speed and accuracy of olfactory discrimination in the rat. *Nature Neuroscience*, 6(11):1224–1229, Nov. 2003.

[253] H. J. Uster, K. Bättig, and H. H. Nägeli. Effects of maze geometry and experience on exploratory behavior in the rat. *Animal Learning & Behavior*, 4(1):84–88, Mar. 1976.

[254] R. Vale, D. A. Evans, and T. Branco. Rapid spatial learning controls instinctive defensive behavior in mice. *Current Biology*, 27(9):1342–1349, 2017.

[255] M. T. van Dijk and A. A. Fenton. On how the dentate gyrus contributes to memory discrimination. *Neuron*, 98(4):832–845, 2018.

[256] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, and D. Silver. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

[257] R. Wagner. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Appleton-Centry-Crofts*, 1972.

[258] C. X. Wang, I. A. Hilburn, D.-A. Wu, Y. Mizuhara, C. P. Cousté, J. N. Abrahams, S. E. Bernstein, A. Matani, S. Shimojo, and J. L. Kirschvink. Transduction of the geomagnetic field as evidenced from alpha-band activity in the human brain. *eneuro*, 2019.

[259] M. Watabe-Uchida, N. Eshel, and N. Uchida. Neural circuitry of reward prediction error. *Annual Review of Neuroscience*, 40:373–394, 2017.

[260] B. Webb and A. Wystrach. Neural mechanisms of insect navigation. *Current Opinion in Insect Science*, 15:27–39, June 2016.

[261] J. N. Weber, B. K. Peterson, and H. E. Hoekstra. Discrete genetic modules are responsible for complex burrow evolution in peromyscus mice. *Nature*, 493(7432):402–405, Jan. 2013.

[262] R. Wehner, B. Michel, and P. Antonsen. Visual navigation in insects: Coupling of egocentric and geocentric information. *Journal of Experimental Biology*, 199(1):129–140, Jan. 1996.

[263] C. Wernicke. The symptom complex of aphasia. In *Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968*, pages 34–97. Springer, 1969.

[264] O. L. White, D. D. Lee, and H. Sompolinsky. Short-term memory in orthogonal neural networks. 2004.

[265] J. C. Whittington, T. H. Muller, S. Mark, G. Chen, C. Barry, N. Burgess, and T. E. Behrens. The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183(5):1249–1263, 2020.

[266] B. Widrow and M. E. Hoff. Adaptive switching circuits. Technical report, Stanford Univ Ca Stanford Electronics Labs, 1960.

[267] M. A. Wilson and B. L. McNaughton. Dynamics of the hippocampal ensemble code for space. *Science (New York, N.Y.)*, 261(5124):1055–1058, Aug. 1993.

[268] R. C. Wilson, Y. K. Takahashi, G. Schoenbaum, and Y. Niv. Orbitofrontal cortex as a cognitive map of task space. *Neuron*, 81(2):267–279, 2014.

[269] W. Wiltschko and R. Wiltschko. Magnetic orientation and magnetoreception in birds and other animals. *Journal of comparative physiology A*, 191(8):675–693, 2005.

[270] R. A. Wise and M. A. Bozarth. A psychomotor stimulant theory of addiction. *Psychological Review*, 94(4):469–492, 1987.

[271] D. M. Wolpert, R. C. Miall, and M. Kawato. Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2(9):338–347, Sept. 1998.

[272] R. A. Wood, M. Bauza, J. Krupic, S. Burton, A. Delekate, D. Chan, and J. O'Keefe. The honeycomb maze provides a novel test to study hippocampal-dependent spatial navigation. *Nature*, 554(7690):102–105, Feb. 2018.

[273] H. Woodrow. The problem of general quantitative laws in psychology. *Psychological Bulletin*, 39(1):1–27, 1942.

[274] L.-Q. Wu and J. D. Dickman. Neural correlates of a magnetic sense. *science*, 336(6084):1054–1057, 2012.

[275] M. M. Yartsev and N. Ulanovsky. Representation of three-dimensional space in the hippocampus of flying bats. *Science*, 340(6130):367–372, 2013.

[276] M. M. Yartsev, M. P. Witter, and N. Ulanovsky. Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature*, 479(7371):103–107, 2011.

[277] M. Yilmaz and M. Meister. Rapid innate defensive responses of mice to looming visual stimuli. *Current Biology*, 23(20):2011–5, 2013.

[278] R. M. Yoder, B. J. Clark, J. E. Brown, M. V. Lamia, S. Valerio, M. E. Shinder, and J. S. Taube. Both visual and idiothetic cues contribute to head direction cell stability during navigation along complex routes. *Journal of Neurophysiology*, 105(6):2989–3001, Mar. 2011.

[279] J. H. Yoo, V. Zell, N. Gutierrez-Reed, J. Wu, R. Ressler, M. A. Shenasa, A. B. Johnson, K. H. Fife, L. Faget, and T. S. Hnasko. Ventral tegmental area glutamate neurons co-release GABA and promote positive reinforcement. *Nature communications*, 7(1):1–13, 2016.