eLife

Computational and Systems Biology, Neuroscience

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

Reviewed Preprint posted
24 March 2023

Sent for peer review
8 November 2022

Posted to bioRxiv
10 October 2022

Endotaxis: A neuromorphic algorithm for mapping, goal-learning, navigation, and patrolling

Tony Zhang, Matthew Rosenberg, Pietro Perona, Markus Meister

Division of Biology and Biological Engineering, California Institute of Technology • Division of Engineering and Applied Science, California Institute of Technology

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 (55), a pond (35), or a desert (37) – and much has been learned about neural signals in diverse brain areas under these conditions (11; 45). However, many natural environments are highly complex, such as a system of burrows, or of intersecting paths through the underbrush. The same applies to many cognitive tasks, that typically allow only a limited set of actions at any given stage in the process. Here we propose an algorithm that learns the structure of a complex 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 (4). 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.

eLife assessment

This **useful** work proposes a framework inspired by chemotaxis for understanding how the brain might implement behaviours related to navigating towards a goal. The evidence supporting the conceptual claim is **convincing**, but some technical aspects are **incomplete**. The manuscript proposes a hypothesis that would be of interest to the broad systems neuroscience community, but the reviewers noted the relationship to existing similar hypotheses was not made sufficiently clear.

1 Introduction

Animals navigate their environment to look for resources – such as shelter, food, or a mate – and to exploit such resources once they are found. Efficient navigation requires knowing the structure of the environment: which locations are connected to which others (53). 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 (4). 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 (1; 28).

The specifics of chemotaxis, namely the methods for finding an odor and tracking it, vary by species, but the toolkit always includes a search strategy based on trial-and-error: Try various actions that you have available, then settle on the one that makes the odor stronger (4). 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 (6). We propose that this universal behavioral module for chemotaxis can be harnessed to solve general problems of search and navigation in a complex environment, even when tell-tale odors are not available.

For concreteness, consider a mouse exploring a labyrinth of tunnels (Fig 1A). The maze may contain a source of food that emits an odor (Fig 1A1). 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 1A2-3). 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.

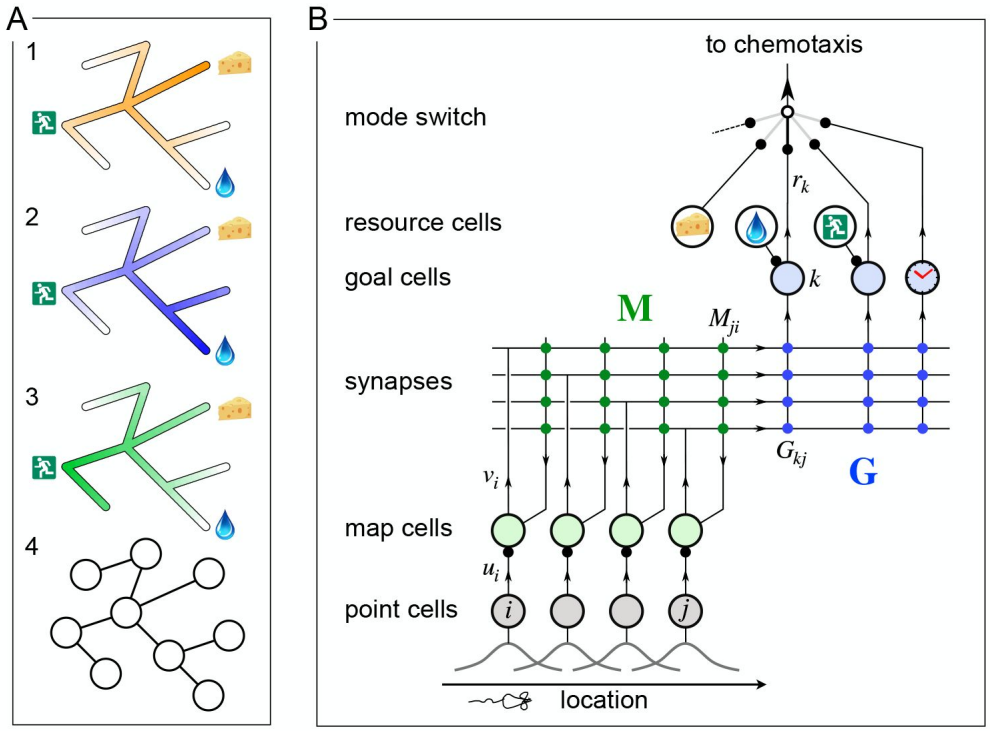

Figure 1:

A mechanism for endotaxis.

A: A constrained environment of tunnels linked by intersections, with special locations offering food, water, and the exit. **1:** A real odor emitted by the food source decreases with distance (shading). **2:** A virtual odor tagged to the water source. **3:** A virtual odor tagged to the exit. **4:** Abstract representation of this environment by a graph of nodes (intersections) and edges (tunnels). **B:** A neural circuit to implement endotaxis. Open circles: four populations of neurons that represent "resource", "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). Resource cells signal the presence of a resource, e.g. cheese, water, or the exit. Map synapses and goal synapses are modified by activity-dependent plasticity. 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). Another goal cell (clock) may report how recently the agent has visited a location. The output of the mode switch 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, **M** is the matrix of synaptic weights among map cells, **G** are the synaptic weights from the map cells onto goal cells, and r_k is the output of goal cell k.

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*.

The developments reported here were inspired by a recent experimental study with mice navigating a complex labyrinth (43) that includes 63 three-way junctions. Among other things, we observed that mice could learn the location of a resource in the labyrinth after encountering it just once, and perfect a direct route to that target location after ~10 encounters. Furthermore, they could navigate back out of the labyrinth using a direct route they had not traveled before, even on the first attempt. Finally, the animals spent most of their waking time patrolling the labyrinth, even long after they had perfected the routes to rewarding locations. These patrols covered the environment efficiently, avoiding repeat visits to the same location. All this happened within a few hours of the animal's first encounter with the labyrinth. Our modeling efforts here are aimed at explaining these remarkable phenomena of rapid spatial learning in a new environment: one-shot learning of a goal location, zero-shot learning of a return route, and efficient patrolling of a complex

maze. In particular we want to do so with a biologically plausible mechanism that could be built out of neurons.

2 A neural circuit to implement endotaxis

Figure 1B presents 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: resource cells, point cells, map cells, and goal cells.

Resource cells

These are sensory neurons that fire when the animal encounters an interesting resource, for example water or food, that may form a target for future navigation. Each resource cell is selective for a specific kind of stimulus. The circuitry that produces these responses is not part of the model.

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 a local plasticity rule. After learning, they represent the topology of the environment.

Goal cells

Each goal cell serves to mark the locations of a special resource in the map of the environment. The goal cell receives excitatory input from a resource cell, which gets activated whenever that resource is present. It also receives excitatory synapses from map cells. Those synapses are strengthened when the presynaptic map cell is active at the same time as the resource cell.

After the map and goal synapses have been learned, each goal cell carries a virtual odor signal for its assigned resource. The signal increases systematically as the animal moves closer to a location with that resource. A mode 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 location.

2.1 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 the target location. This relationship holds even if the environment is constrained with a complex connectivity graph (Fig 1A4). Here we explain how this comes about, with mathematical details to follow.

As the animal explores a new environment, when it moves from one location to an adjacent one, those two point cells fire in rapid succession. That leads to a Hebbian strengthening of the excitatory synapses between the two corresponding map cells (Fig 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 (14; 21), which in turn reflects the adjacency matrix of the graph (Eqn 1). Now the brain can use this adjacency information to find the shortest path to a target.

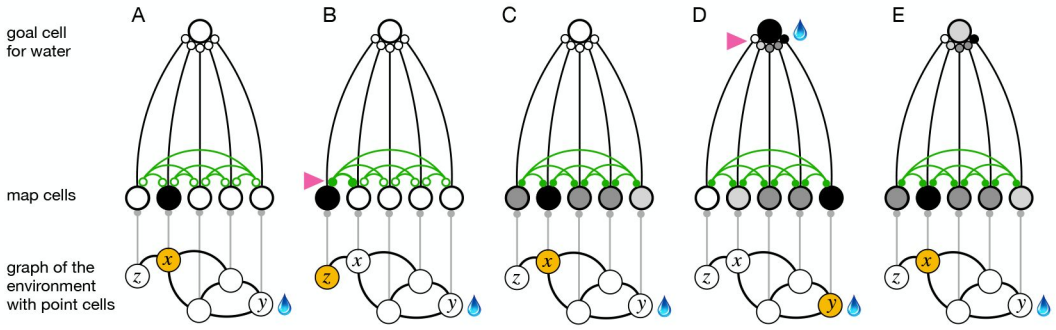

Figure 2:

The phases of endotaxis during exploration, goal-tagging, and navigation.

A portion of the circuit in Figure 1 is shown, including a single goal cell that responds to the water resource. Bottom shows a graph of the environment, with nodes linked by edges, and 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 (empty circles). **B:** During exploration the agent moves between two adjacent nodes on the graph, and that strengthens the connection between their corresponding map cells (arrowhead, filled circles). **C:** After exploration the map synapses reflect the connectivity of the graph. Now the map cells have an extended profile of activity (darker = more active), 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, and this triggers plasticity at its input synapses (arrowhead). 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 water location.

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 of the environment (Fig 2C). If the animal moves to a different location *y*, the map output will change to another hump of activity, now centered on *y* (Fig 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 water resource cell fires, triggering synaptic learning in the goal synapses. That stores the current profile of map activity $v_i(y)$ in the synapses G_{ki} onto the goal cell k that responds to water (Fig 2D), Eqn 8). When the animal subsequently moves to a different location x, the goal cell k receives the current map output $\mathbf{v}(x)$ filtered through the previously stored synaptic template $\mathbf{v}(y)$ (Fig 2E). This is the desired measure of overlap (Eqn 9). Under suitable conditions this goal signal declines monotonically with the shortest graph-distance between x and y, as we will demonstrate both analytically and in simulations (Sections 3, 4, 7).

3 Theory of endotaxis

Here we formalize the processes of Figure 2 in a concrete mathematical model. The model is simple enough to allow some exact predictions for its behavior. The present section develops an analytical understanding of endotaxis that will help guide the numerical simulations in subsequent parts.

The environment is modeled as a directed graph consisting of n nodes, with adjacency matrix

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

Movements of the agent are modeled as a sequence of steps along that graph. During exploration, the agent performs a walk that tries to cover the entire environment; in the simplest case a random walk. During navigation, the agent is instead guided at each intersection by maximizing a goal signal.

We implement the circuit of Fig 1B as a textbook linear rate model (14). The point neurons are one-hot encoders of location: A point neuron fires if the agent is at that location; all the others are silent:

$$u_i(x) = \text{firing rate of point cell } i \text{ with the agent at node } x \tag{2}$$

$$= \delta_{ix} \tag{3}$$

where δ_{ix} is the Kronecker delta.

A map neuron sums synaptic input linearly from its point cell and the other map units; its output is simply proportional to that input:

$$v_i = \gamma \left(u_i + \sum_j M_{ij} v_j \right) \tag{4}$$

So the vector of all map outputs is

$$\mathbf{v} = \gamma\left(\mathbf{u} + \mathbf{M}\mathbf{v}\right) = \left(\frac{1}{\gamma}\mathbf{1} - \mathbf{M}\right)^{-1}\mathbf{u} \tag{5}$$

where γ is the gain of the map units, and $\mathbf{u}$ is the one-hot input from point cells.

Now consider goal cell number k that is associated to a particular location y, because its resource is present at that node. The goal cell sums input from all the map units v_i, weighted by its goal synapses G_{ki}. So with the agent at node x the goal signal r_k is:

$$r_k(x) = \sum_i G_{ki} \cdot v_i(x) = \mathbf{g}_k \cdot \mathbf{v}(x) = \mathbf{g}_k \cdot \left(\frac{1}{\gamma}\mathbf{1} - \mathbf{M}\right)^{-1}\mathbf{u}(x) \tag{6}$$

where we write $\mathbf{g}_k$ for the k^{th} row vector of the goal synapse matrix $\mathbf{G}$. This is the set of synapses from all map cells onto the specific goal cell in question.

Suppose now that the agent has learned the structure of the environment perfectly, such that the map synapses are a copy of the graph's adjacency matrix (1),

$$\mathbf{M} = \mathbf{A} \tag{7}$$

Similarly, suppose that the agent has acquired the goal synapses perfectly, namely proportional to the map output at the goal location y:

$$\mathbf{g}_k = \mathbf{v}(y) \tag{8}$$

Then as the agent moves to another location x, the goal cell reports a signal

$$r_k(x) = \mathbf{g}_k \cdot \mathbf{v}(x) = \mathbf{v}(y) \cdot \mathbf{v}(x) \equiv E_{xy} \tag{9}$$

where the matrix

$$\mathbf{E} = \left(\frac{1}{\gamma}\mathbf{1} - \mathbf{A}\right)^{-1^{\top}}\left(\frac{1}{\gamma}\mathbf{1} - \mathbf{A}\right)^{-1} \tag{10}$$

One can show (Section A.1) that for small $\gamma \ll 1$ this matrix $\mathbf{E}$ reflects the shortest distance between nodes on the graph, namely

$$E_{xy} \sim \gamma^{-D_{xy}} \tag{11}$$

where D_{xy} is the smallest number of steps needed to get from node y to node x.

Under the assumptions stated, the goal signal E_{xy} between nodes x and y declines monotonically with their distance. Figure 3 illustrates this with numerical results on a binary tree graph. As expected, for small γ the goal signal decays exponentially with graph distance (Fig 3A). Therefore an agent that makes turning decisions to maximize that goal signal will reach the goal by the shortest possible path.

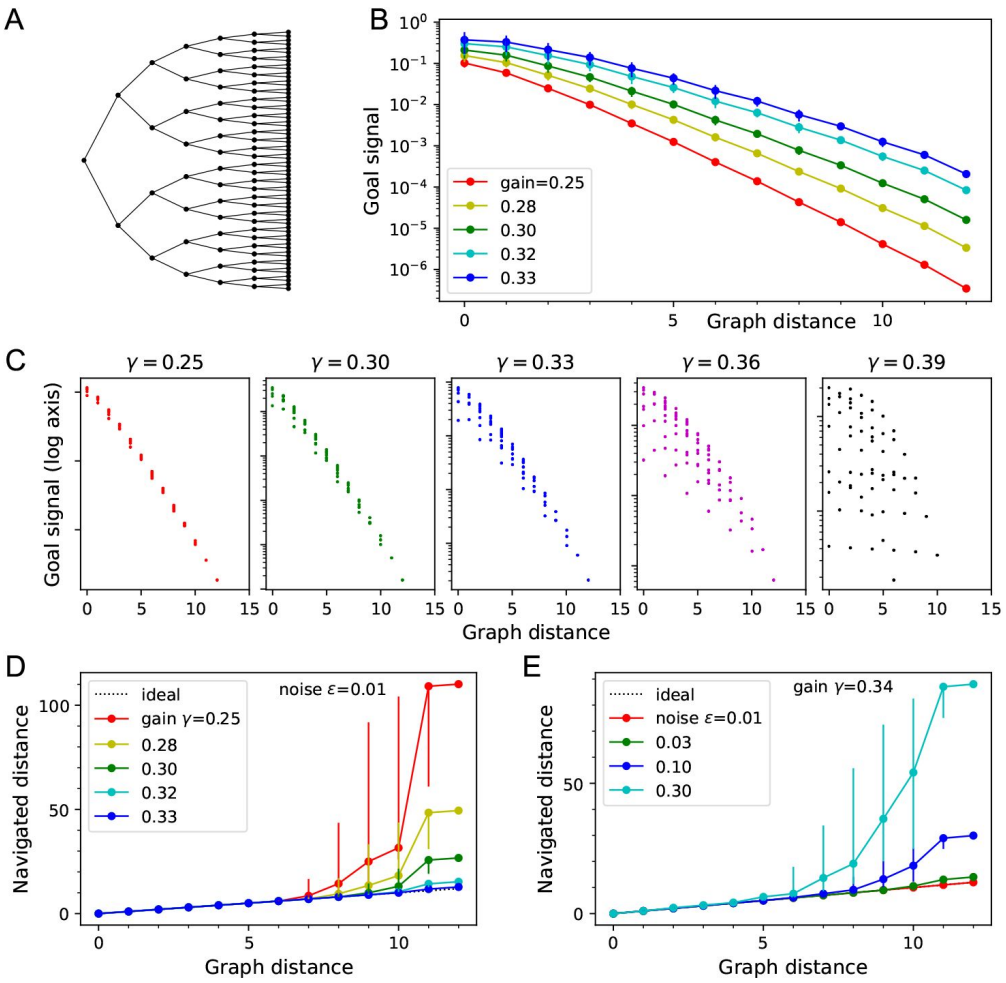

Figure 3:

Theory of the goal signal.

Dependence of the goal signal on graph distance, and the consequences for endotaxis navigation. **A:** The graph representing a binary tree labyrinth (43) serves for illustration. Suppose the endotaxis model has acquired the adjacency matrix perfectly: **M** = **A**. We compute the goal signal E_{xy} between any two nodes on the graph, and compare the results at different values of the map gain γ. **B:** Dependence of the goal signal E_{xy} on the graph distance D_{xy} between the two nodes. Mean ± SD, error bars often smaller than markers. Note logarithmic vertical axis. The signal decays exponentially over many log units. At high γ the decay distance is greater. **C:** A detailed look at the goal signal, each point is for a pair of nodes (x, y). For low γ the decay with distance is strictly monotonic. At high γ there is overlap between the values at different distances. As γ exceeds the critical value $\gamma_c = 0.38$ the distance-dependence breaks down. **D:** Using the goal signal for navigation. For every pair of start and end nodes we navigate the route by following the goal signal and compare the distance traveled to the shortest graph distance. For all routes with the same graph distance we plot the median navigated distance with 10% and 90% quantiles. Variable gain at a constant noise value of $\epsilon = 0.01$. At gains $\gamma > 0.35$ navigation failed for some point pairs. **E:** As in panel (D) but varying the noise at a constant gain of $\gamma = 0.34$.

The exponential decay of the goal signal represents a challenge for practical implementation with biological circuits. Neurons have a finite signal-to-noise ratio, so detecting minute differences in the firing rate of a goal neuron will be unreliable. Because the goal signal changes by a factor of γ across every link in the graph, one wants to set the map neuron gain γ as large as possible. Unfortunately there is a strict upper limit for that gain:

$$\gamma < \gamma_c \equiv \frac{1}{\text{largest absolute eigenvalue of } \mathbf{A}} \qquad (12)$$

For larger $\gamma > \gamma_c$ the goal signal E_{xy} no longer represents graph distances (Section A.2). The largest eigenvalue of the adjacency matrix in turn is related to the number of edges per

To implement the finite dynamic range explicitly, we add some noise to the goal signal of Eqn 9:

$$r_k(x) = \mathbf{g}_k \cdot \mathbf{v}(x) + \eta \qquad (13)$$

where the noise η is uniformly distributed with range ϵ:

$$\eta \in [-\epsilon/2, \epsilon/2]$$

The scale ϵ of this noise is expressed relative to the maximum value of the goal signal. What is a reasonable value for this noise? For reference, humans and animals can routinely discriminate sensory stimuli that differ by only 1%, for example the pitch of tones or the intensity of a light, especially if they occur in close succession. Clearly the neurons all the way from receptors to perception must represent those small differences. Thus we will use ϵ = 0.01 as a reference noise value in many of the results presented here. [3]

The process of navigation towards a chosen goal signal is formalized in Algorithm 1. At each node the agent inspects the goal signal that would be obtained at all the neighboring nodes, corrupted by the readout noise η. Then it steps to the neighbor with the highest value. Suppose the agent starts at node x and navigates following the goal signal for node y. The resulting navigation route $x = s_0, s_1, \ldots, s_n = y$ has $L_{xy} = n$ steps. Navigation is perfect if this equals the shortest graph distance, $L_{xy} = D_{xy}$. We will assess deviations from perfect performance by the excess length of the routes.

Algorithm 1 Navigation

Parameters: gain γ, noise ϵ
Input: map synapse matrix $\mathbf{M}$, goal synapse vector $\mathbf{g}$

$s \leftarrow x$ ▷ start navigation at node x
while not at goal **do** ▷ stop when goal resource is found
  for all nodes j that neighbor s **do**
    $\mathbf{u}(j)_i \leftarrow \delta_{i,j}$ for every point cell i ▷ point cell output with agent at node j
    $\mathbf{v}(j) \leftarrow \left(\frac{1}{\gamma}\mathbf{1} - \mathbf{M}\right)^{-1} \mathbf{u}(j)$ ▷ map output
    $r(j) \leftarrow \mathbf{g}\,\mathbf{v}(j) + \eta(j)$ ▷ noisy goal signal, $\eta \sim \mathrm{Unif}(\epsilon)$
  end for
  $s \leftarrow \arg\max_j r(j)$ ▷ choose the neighbor node with the highest goal signal
end while

Figure 3D-E illustrates how the navigated path distance L_{xy} depends on the noise level ϵ and the gain γ. For small gain or high noise the goal signal extends only over a graph distance of 5-6 links. Beyond that the navigated distance L_{xy} begins to exceed the graph distance D_{xy}. As the gain increases, the goal signal extends further through the graph and navigation becomes reliable over longer distances (Fig 3D). Eventually, however, the goal signal loses its monotonic distance dependence (Fig 3C). At that stage, navigation across the graph may fail because the agent gets trapped in a local maximum of the goal signal. This can happen even before the critical gain value is reached (Fig 3C). For the example in Fig 3 the highest useful gain is γ = 0.34 whereas γ_c = 0.383.

For any given value of the gain, navigation improves with lower noise levels, as expected (Fig 3E). At the reference value of $\epsilon = 0.01$, navigation is perfect even across the 12 links that separate the most distant points on this graph.

In summary, this analysis spells out the challenges that need to be met for endotaxis to work properly. First, during the learning phase, the agent must reliably extract the adjacency matrix of the graph, and copy it into its map synapses. Second, during the navigation phase, the agent must evaluate the goal signal with enough resolution to distinguish the values at alternative nodes. The neuronal gain γ plays a central role: With γ too small, the goal signal decays rapidly with distance and vanishes into the noise just a few steps away from the goal. But at large γ the network computation becomes unstable.

4 Acquisition of map and targets during exploration

As discussed above, the goal of learning during exploration is that the agent acquires a copy of the graph's adjacency matrix in its map synapses, $\mathbf{M} \approx \mathbf{A}$, and stores the map output at a goal location y in the goal synapses $\mathbf{g} \approx \mathbf{v}(y)$. Here we explore how the rules for synaptic plasticity in the map and goal networks allow that to happen. Algorithm 2 spells out the procedure we implemented for learning from a random walk through the environment.

The map synapses M_{ij} start out at zero strength. When the agent moves from node $j = s(t)$ at time t to node $i = s(t + 1)$, the map cell j is excited before the step, and map cell i after the step. When that happens, the agent potentiates the synapse between those two neurons to $M_{ij} = 1$. Of course, a map cell can also get activated through the recurrent network, and we must distinguish that from direct input from its point cell. We found that a simple threshold criterion is sufficient. Here θ is a threshold applied to both the pre- and post-synaptic activity, and the map synapse gets established only if both neurons respond above threshold. The tuning requirements for this threshold are discussed below.

The map learning rule produces a full strength synapse after a single step: This allows the agent to learn a route after the first traversal, which is needed to explain the rapid learning observed in experimental animals. Note also that the potentiation depends on temporal sequence: the pre-synaptic neuron must be active before the post-synaptic neuron. This allows the agent to learn a directed graph, in which links can be traversed in only one direction. For learning on undirected graphs it can be useful to relax the time-dependent rule (see Section 7).

Algorithm 2 Map and goal learning

Parameters: γ, θ, α
Input: adjacency matrix $\mathbf{A}$, resource signals $\mathbf{F}$

$\mathbf{M} \leftarrow 0$ ▷ initiate map synapses at 0
$\mathbf{G} \leftarrow 0$ ▷ initiate goal synapses at 0
$t \leftarrow 0$ ▷ t counts the steps
$s(t) \leftarrow x$ ▷ start random walk at x
while learning **do**
  $t \leftarrow t + 1$
  $s(t) \leftarrow$ a random neighbor of $s(t-1)$ ▷ continue the random walk
  $u_i(t) \leftarrow \delta_{i,s(t)}$ for every point cell i ▷ point cell output
  $\mathbf{v}(t) \leftarrow \left(\frac{1}{\gamma}\mathbf{1} - \mathbf{M}\right)^{-1} \mathbf{u}(t)$ ▷ map cell output
  for all map cell pairs (i, j) **do**
    if $v_j(t-1) > \theta$ and $v_i(t) > \theta$ **then** ▷ threshold on pre- and post-synaptic activity
      $M_{ij} \leftarrow 1$ ▷ on undirected graphs can also increment M_{ji}
    end if
  end for
  $\mathbf{r} \leftarrow \mathbf{G}\mathbf{v}(t)$ ▷ goal signals
  for every goal neuron k **do**
    if $F_{k,s(t)} > 0$ **then** ▷ the agent is at a location that contains resource k
      for every map neuron j **do**
        $G_{kj} \leftarrow G_{kj} + \alpha(F_{k,s(t)} - r_k)v_j(t)$ ▷ update goal synapses
      end for
    end if
  end for
end while

The goal synapses G_{kj} similarly start out at zero strength. Consider a particular goal cell k, and suppose its corresponding resource cell has activity F_{ky} when the agent is at location y. When a positive resource signal arrives, that means the agent is at a goal location. If the goal signal r_k received from the map output is smaller than the resource signal F_{ky}, then the goal synapses get incremented by something proportional to the current map output. Learning at the goal synapses saturates when the goal signal correctly predicts the resource signal. The learning rate α sets how fast that will happen. Note that both the learning rules for map and goal synapses are Hebbian and strictly local: Each synapse is modified based only on signals available in the pre- and post-synaptic neurons.

To illustrate the process of map and goal learning we simulate an agent exploring a simple ring graph by a random walk (Fig 4). At first, there are no targets in the environment that can deliver a resource (Fig 4A). Then we add one target location, and later a second one. Finally we add a new link to the graph that makes a connection clear across the environment. As the agent explores the graph, we will track how its representations evolve by monitoring the map synapses and the profile of the goal signal.

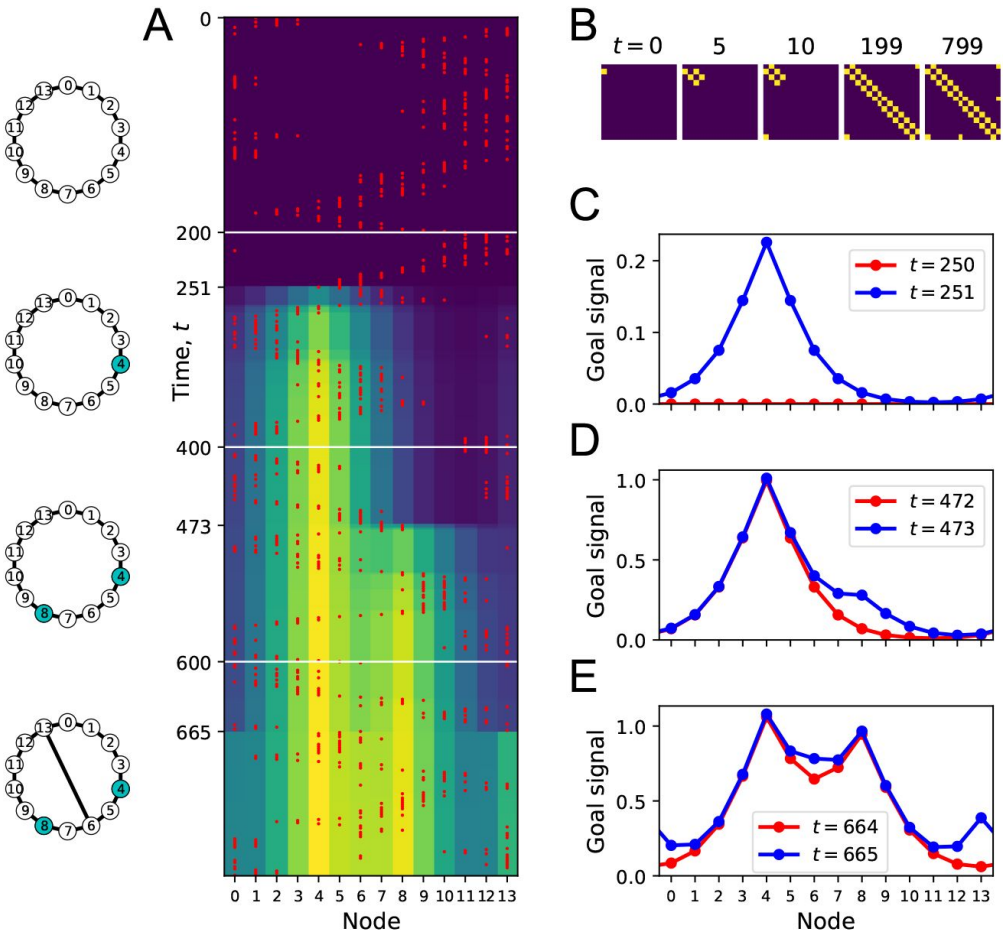

Figure 4:

Learning the map and the targets during exploration.

(A) Simulation of a random walk on a ring with 14 nodes. Left: Layout of the ring, with resource locations marked in blue. The walk progresses in 800 time steps (top to bottom); with the agent's position marked in red (nodes 0-13, horizontal axis). At each time the color map shows the goal signal that would be produced if the agent were at position 'Node'. White horizontal lines mark the appearance of a target at t = 200, a second target with the same resource at t = 400, and a new link across the ring at step t = 600. **(B)** The matrix **M** of map synapses at various times. The pixel in row i and column j represents the matrix element M_{ij}. Color purple = 0. Note the first few steps (number above graph) each add a new synapse. Eventually, **M** reflects the adjacency matrix of nodes on the graph. **(C)** Goal signals just before and just after the agent encounters the first target. **(D)** Goal signals just before and just after the agent encounters the second target. **(E)** Goal signals just before and just after the agent travels the new link for the first time.

At the outset, every time the agent steps to a new node, the map synapse corresponding to that link gets potentiated (Fig 4B). After enough steps, the agent has executed every link on the graph, and the matrix of map synapses resembles the full adjacency matrix of the graph (Fig 4B). At this stage the agent has learned the connectivity of the environment.

Once a target appears in the environment it takes the agent a few random steps to encounter it. At that moment the goal synapses get potentiated for the first time, and suddenly a goal signal appears in the goal cell (Fig 4C). The profile of that goal signal is fully formed and spreads through the entire graph thanks to the pre-established map network. By following this goal signal uphill the agent can navigate along the shortest path to the target from any node on the graph. Note that the absolute scale of the goal signal grows a little every time the agent visits the goal (Fig 4A) and eventually saturates.

Some time later we introduce a second target elsewhere in the environment (Fig 4D). When the agent encounters it along its random walk, the goal synapses get updated, and the new goal signal has two peaks in its profile. Again, this goal signal grows during subsequent visits. By following that signal uphill from any starting point, the agent will be led to a nearby target by the shortest possible path.

When a new link appears, the agent eventually discovers it on its random walk. At that point the goal signal changes instantaneously to incorporate the new route (Fig 4E). An agent following the new goal signal from node 13 on the ring will now be led to a target location in just 3 steps, using the shortcut, whereas previously it took 5 steps.

This simulation illustrates how the structure of the environment is acquired separately from the location of resources. The agent can explore and learn the map of the environment even without any resources present (Fig 4B). This learning takes place among the map synapses in the endotaxis circuit (Fig 1B). When a resource is found, its location gets tagged within that established map through learning by the goal synapses. The resulting goal signal is available immediately without the need for further learning (Fig 4C). If the distribution of resources changes, the knowledge in the map remains unaffected (Fig 4D) but the goal synapses can change quickly to incorporate the new target. Vice versa, if the graph of the environment changes, the map synapses get updated, and that adapts the goal signal to the new situation even without further change in the goal synapses (Fig 1E).

What happens if a previously existing link disappears from the environment, for example because one corridor of the mouse burrow caves in? Ideally the agent would erase that link from the cognitive map. The learning algorithm Alg 2 is designed for rapid and robust acquisition of a cognitive map starting from zero knowledge, and does not contain a provision for forgetting. However, one can add a biologically plausible rule for synaptic depression that gradually erases memory of a link if the agent never travels it. Details are presented in Supplement section A.5 (Fig 9). For sake of simplicity we continue the present analysis of endotaxis based on the simple 3-parameter algorithm presented above (Alg 2).

5 Navigation using the learned goal signal

We now turn to the “exploitation” component of endotaxis, namely use of the learned information to navigate towards targets. In the simulations of Figure 5 we allow the agent to explore a graph. Every node on the graph drives a separate resource cell, thus the agent simultaneously learns goal signals to every node. After a random walk sufficient to cover the graph several times, we test the agent’s ability to navigate to the goals by ascending on the learned goal signal. For that purpose we teleport the agent to an arbitrary start node in the graph and ask how many steps it takes to reach the goal node.

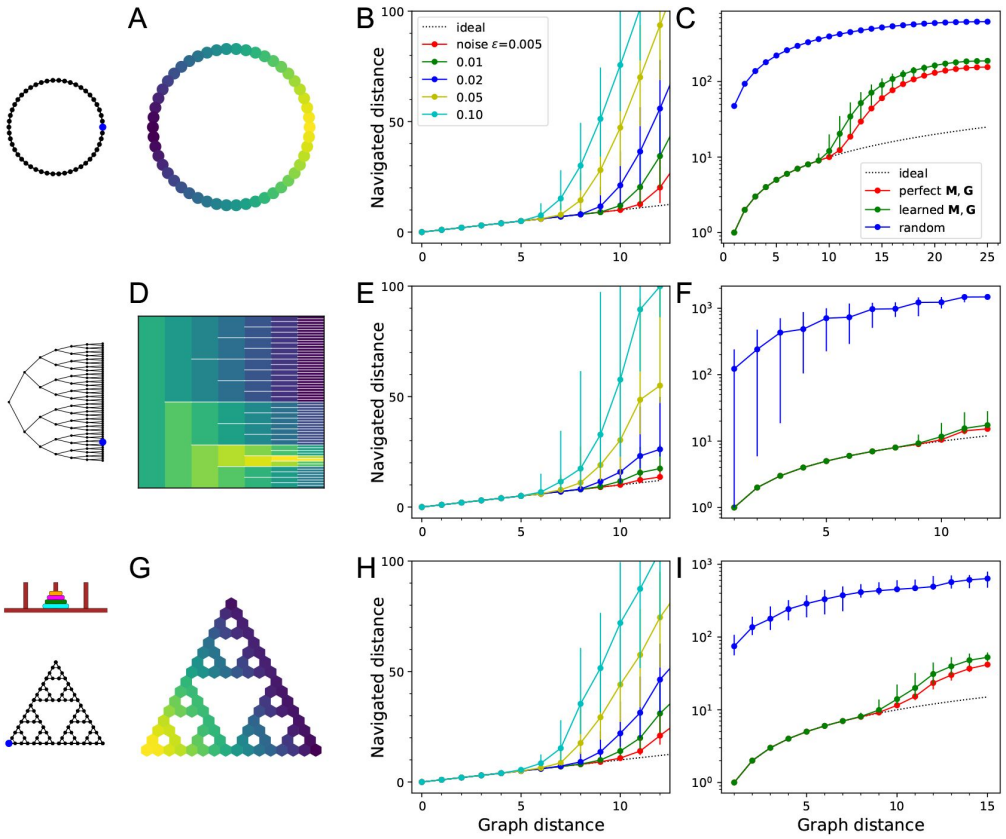

Figure 5:

Navigation using the learned map and targets.

(A-C) Ring with 50 nodes. **(A)** Goal signal for a single target location (blue dot on left icon), after learning during random exploration with 10,000 steps. Color scale is logarithmic, yellow=high. Note monotonic decay of the goal signal with graph distance from the target. **(B)** Results of all-to-all navigation where every node is a separate goal. For all pairs of nodes this shows the navigated distance vs the graph distance. Median ± 10/90 percentiles for all routes with the same graph distance. "Ideal" navigation would follow the identity. The actual navigation is ideal over short distances, then begins to deviate from ideal at a critical distance that depends on the noise level ϵ. **(C)** As in (B) over a wider range, note logarithmic axis. Noise ϵ = 0.01. Includes comparison to navigation by a random walk; and navigation using the optimal goal signal based on knowledge of the graph structure and target location. γ = 0.41, θ = 0.39, α = 0.3. **(D-F)** As in (A-C) for a binary tree graph with 127 nodes. **(D)** Goal signal to the node marked on the left icon. This was the reward port in the labyrinth experiments of (43). White lines separate the branches of the tree. γ = 0.32, θ = 0.27, α = 0.3. **(G-I)** As in (A-C) for a "Tower of Hanoi" graph with 81 nodes. γ = 0.29, θ = 0.27, α = 0.3.

Figure 5A-C shows results on a ring graph with 50 nodes. With suitable values of the model parameters (γ, θ, α) – more on that later – the agent learns a goal signal that declines monotonically with distance from the target node (Fig 5A). The ability to ascend on that goal signal depends on the noise level ϵ, which determines whether the agent can sense the difference in goal signal at neighboring nodes. At a high noise level ϵ = 0.1 the agent finds the target by the shortest route from up to 5 links away (Fig 5B); beyond that range some navigation errors creep in. At a low noise level of ϵ = 0.005 navigation is perfect up to 10 links away. Every factor of two increase in noise seems to reduce the range of navigation by about one link.

How does the process of learning the map of the environment affect the ultimate navigation performance? Figure 5C makes that comparison by considering an agent with oracular knowledge of the graph structure and target location (Eqns 7 and 8). Interestingly this adds only 1 link to the distance range for perfect navigation. Here we also compare to an agent with zero knowledge of the environment that performs a random walk. On this graph that takes about 40 times longer than by using endotaxis.

The ring graph is particularly simple, but how well does endotaxis learn in a more realistic environment? Figure 5D-F shows results on a binary tree graph with 6 levels: This is the

structure of a maze used in a recent study on mouse navigation (43). In those experiments, mice learned quickly how to reach the reward location (blue dot in Fig 5D) from anywhere within the maze. Indeed, the endotaxis agent can learn a goal signal that declines monotonically with distance from the reward port (Fig 5D). At a noise level of $\epsilon = 0.01$ navigation is perfect over distances of 9 links, and close to perfect over the maximal distance of 12 links that occurs in this maze (Fig 5E). Again, the challenge of having to learn the map affects the performance only slightly (Fig 5F). Finally, comparison with the random agent shows that endotaxis shortens the time to target by a factor of 100 on this graph (Fig 5F).

Figure 5G-I shows results for a more complex graph that represents a cognitive task, namely the game "Tower of Hanoi". Disks of different sizes are stacked on four 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 the center 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 5G). The player starts at the top node (all disks on the center peg) and the two possible solutions correspond to the bottom left and right corners. Again, random exploration leads the endotaxis agent to learn the connectivity of the game and to discover the solutions. The resulting goal signal decays systematically with graph distance from the solution (Fig 5G). At a noise of $\epsilon = 0.01$ navigation is perfect once the agent gets to within 9 moves of the target (Fig 5H). This is not quite sufficient for an error-free solution from the starting position, which requires 15 moves. However, compared to an agent executing random moves, endotaxis speeds up the solution by a factor of 10 (Fig 5I). If the game is played with only 3 disks, the maximal graph distance is 7, and endotaxis solves it perfectly at $\epsilon = 0.01$.

These results show that endotaxis functions well in environments with very different structure: linear, tree-shaped, and cyclic. Random exploration in conjunction with synaptic learning can efficiently acquire the connectivity of the environment and the location of targets. With a noise level of 1%, the resulting goal signal allows perfect navigation over distances of ~9 steps, independent of the nature of the graph. This is a respectable range: Personal experience suggests that we rarely learn routes that involve more than 9 successive decisions. Chess openings, which are often played in a fast and reflexive fashion, last about 10 moves. Nonetheless, we explored ways to extend this range.

One potential approach is to counteract the decay of the goal signal across links in the map network. If the goal signal were to decay more gently then it could reach farther before getting corrupted by noise. To this end, we experimented with a nonlinear input-output function for the map cells, for example introducing a tanh() nonlinearity in Eqn 4. This boosts small output values and saturates at large values (16), but did not improve the overall performance of endotaxis. Instead, learning of the map was perturbed, because the learning algorithm (Alg 2) requires a substantial difference between direct activation of a map cell from a point cell and indirect activation of the neighboring map cell.

A more promising approach is to cover the environment by multiple maps with a hierarchy of length scales. In the example of Fig 5G, endotaxis can lead the agent to the solution once it enters the correct third of the graph. So one could envision a second map network with much coarser point cells and fewer links that guides the agent roughly into the right region, from where the fine map can take over. This comes with its own challenges – for example the time scale of synaptic plasticity must be extended to allow for longer travel times – but the concept is worth exploring further.

6 Parameter sensitivity

The endotaxis model has only 3 parameters: the gain γ of map units, the threshold θ for learning at map synapses, and the learning rate α at goal synapses. How does performance

depend on these parameters? Do they need to be tuned precisely? And does the optimal tuning depend on the spatial environment? There is a natural hierarchy to the parameters if one separates the process of learning from that of navigation. Suppose the circuit has learned the structure of the environment perfectly, such that the map synapses reflect the adjacencies (Eqn 7), and the goal synapses reflect the map output at the goal (Eqn 8). Then the optimal navigation performance of the endotaxis system depends only on the gain γ and the noise level ϵ. For a given γ, in turn, the precision of map learning depends only on the threshold θ. Finally, if the gain is set optimally and the map was learned properly, the identification of targets depends only on the goal learning rate α. Figure 6 explores these relationships in turn.

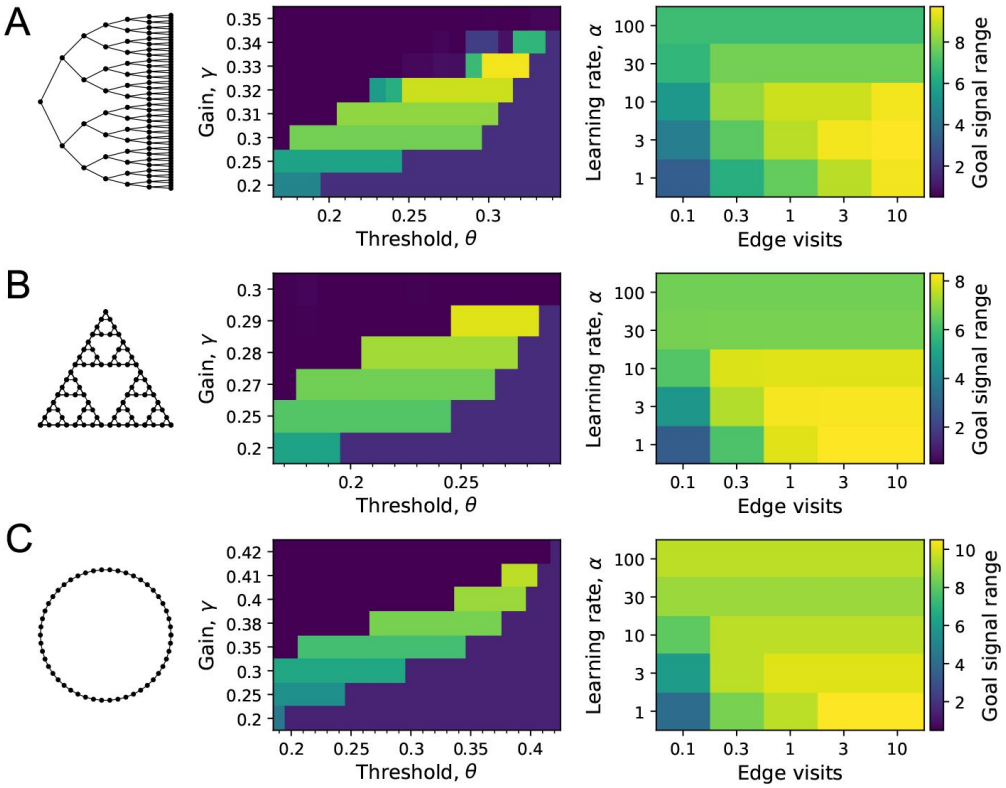

Figure 6:

Sensitivity of performance to the model parameters.

On each of the three graphs we simulated endotaxis for all-to-all navigation, where each node serves as a start and a goal node. The performance measure was the range of the goal signal, defined as the graph distance over which at least half the navigated routes follow the shortest path. The exploration path for synaptic learning was of medium length, visiting each edge on the graph approximately 10 times. The noise was set to $\epsilon = 0.01$. **(A)** Binary tree maze with 127 nodes. **Left:** Dependence of the goal signal range on the gain γ and the threshold θ for learning map synapses. Performance increases with higher gain until it collapses beyond the critical value. For each gain there is a sharply defined range of useful thresholds, with lower values at lower gain. **Right:** Dependence of the goal signal range on the learning rate α at goal synapses, and the length of the exploratory walk, measured in visits per edge of the graph. For a short walk (1 edge visit) a high learning rate is best. For a long walk (100 edge visits) a lower learning rate wins out. **(B)** As in (A) for the Tower of Hanoi graph with 81 nodes. **(C)** As in (A) for a Ring graph with 50 nodes.

We simulated the learning phase of endotaxis as in the preceding section (Fig 5B, E, H), using a noise level of $\epsilon = 0.01$, and systematically varying the model parameters (γ, θ, α). For each parameter set we measured the graph distance over which at least half of the navigated routes were perfect. We defined this distance as the range of the goal signal.

For example, on the binary tree graph with 127 nodes (Fig 6A) the signal range improves with gain, until performance collapses beyond a maximal gain value. This is just as predicted by the theory (Fig 3), except that the maximal gain $\gamma_{max} = 0.34$ is slightly below the critical value $\gamma_c = 0.383$. Clearly the added complications of having to learn the map and goal locations take their toll at high gain. Below the maximal cutoff, the dependence of

performance on gain is rather gentle: For example a 10% change in gain from 0.30 to 0.33 leads to a 23% change in performance. At any given gain value, there is a range of values for the threshold θ that deliver the identical performance. With θ in this range, the map is essentially learned perfectly. Note that this range is generous and does not require precise adjustment: For example, under a near-maximal gain of 0.32, the threshold can vary freely over a 20% range.

Once the gain and synaptic threshold are set so as to acquire the map synapses, the quality of goal learning depends only on the learning rate α. With large α, a single visit to the goal fully potentiates the goal synapses so they don't get updated further. This allows for a fast acquisition of that target, but at the risk of imperfect learning, because the map may not be fully explored yet. A small α will update the synapses only partially over many successive visits to the goal. This leads to a poor performance after short exploration, because the weak goal signal competes with noise, but superior performance after long explorations: a trade-off between speed of learning and accuracy. Precisely this speed-accuracy tradeoff is seen in the simulations (Fig 6A, right): A high learning rate is optimal for short explorations, but for longer ones a small learning rate wins out. An intermediate value of $\alpha = 1$ delivers a good compromise performance.

We found qualitatively similar behavior for the other two environments studied here: The Tower of Hanoi graph (Fig 6B) and a ring graph (Fig 6C). In each case, the maximal usable gain is slightly below the critical value γ_c of that graph. A learning rate of $\alpha = 1$ delivers intermediate results. For long explorations a lower learning rate is best.

In summary this sensitivity analysis shows that the optimal parameter set for endotaxis does depend on the environment. This is not altogether surprising: Every neural network needs to adapt to the distribution of inputs it receives so as to perform optimally. At the same time, the required tuning is rather generous, allowing at least 10-20% slop in the parameters for reasonable performance. Furthermore, a single parameter set of $\gamma = 0.29$, $\theta = 0.26$, $\alpha = 1$ performs quite well on both the binary maze and the Tower of Hanoi graphs, which are dramatically different in character.

7 Navigating a partial map: homing behavior

We have seen that endotaxis can learn both connections in the environment and the locations of targets after just one visit (Fig 6.) This suggests that the agent can navigate well on whatever portion of the environment it has already seen, before covering it exhaustively. To illustrate this we analyze an ethologically relevant instance.

Consider a mouse that enters an unfamiliar environment for the first time, such as a labyrinth constructed by fiendish graduate students (43). Given the uncertainties about what lurks inside, the mouse needs to retain the ability to flee back to the entrance as fast as possible. For concreteness take the mouse trajectory in Figure 7A. The animal has entered the labyrinth (location 1), made its way to one of the end nodes (3), then explored further to another end node (4). Suppose it needs to return to the entrance now. One way would be to retrace all its steps. But the shorter way is to take a left at and cut out the unnecessary branch to (3). Experimentally we found that mice indeed take the short direct route instead of retracing their path (43). They can do so even on the very first visit of an unfamiliar labyrinth. Can endotaxis explain this behavior?

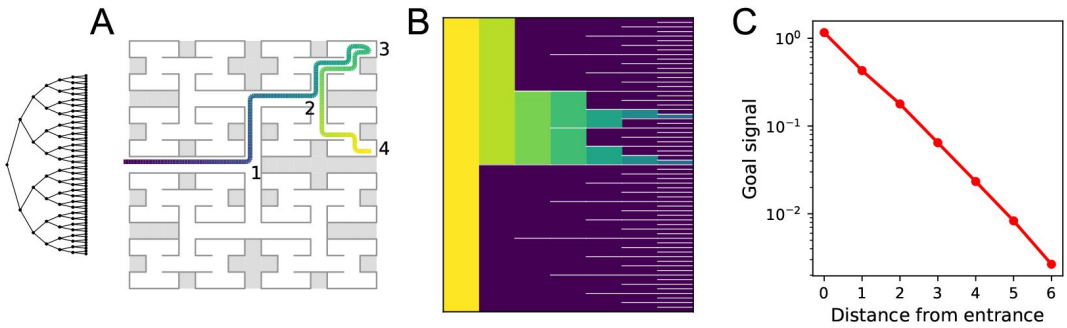

Figure 7:

Homing by endotaxis.

(A) A binary tree maze as used in (43). A simulated mouse begins to explore the labyrinth (colored trajectory, purple=early, yellow=late), traveling from the entrance to one of the end nodes (3), then to another end node (4). Can it return to the entrance from there using endotaxis? **(B)** Goal signal learned by the end of the walk in (A), displayed as in Fig 5D, purple=0. Note the goal signal is non-zero only at the nodes that have been encountered so far. From all those nodes it increases monotonically toward the entrance. **(C)** Detailed plot of the goal signal along the shortest route for homing. Parameters $\gamma = 0.32$, $\theta = 0.27$, $\alpha = 10$, $\epsilon = 0.01$.

We assume that the entrance is a salient location, so the agent dedicates a goal cell to the root node of the binary tree. Figure 7B plots the goal signal after the path in panel A, just as the agent wants to return home. The goal signal is non-zero only at the locations the agent has visited along its path. It clearly increases monotonically towards the entrance (Fig 7C). At a noise level of $\epsilon = 0.01$ the agent can navigate to the entrance by the shortest path without error. Note specifically that the agent does not retrace its steps when arriving at location (2), but instead turns toward (1).

One unusual aspect of homing is that the goal is identified first, before the agent has even entered the environment to explore it. That strengthens the goal synapse from the sole map cell that is active at the entrance. Only subsequently does the agent build up map synapses that allow the goal signal to spread throughout the map network. Another key assumption behind homing is that any link on the graph can be traversed in both directions. For route-finding in a spatial environment, that is often assured. [4] To enable a return home along a path that has only been traveled in the forward direction, we loosened the learning rule for map synapses (Alg 2) to be independent of the activation sequence, so that synapses in both directions get enhanced when two map neurons are active in near coincidence. In general, this variant of the learning rule helps speed up the learning of undirected graphs.

8 Efficient patrolling

Beside exploring and exploiting, a third mode of navigating the environment is patrolling. At this stage the animal knows the lay of the land, and has perhaps discovered some special locations, but continues to patrol the environment for new opportunities or threats. In our study of mice freely interacting with a large labyrinth, the animals spent more than 85% of the time patrolling the maze (43). This continued for hours after they had perfected the targeting of reward locations and the homing back to the entrance. Presumably the goal of

patrolling is to cover the entire environment quickly so as to spot any changes as soon as they develop. So the ideal path in patrolling would visit every node on the graph in the smallest number of steps possible. In the binary-tree maze used for our experiments, that optimal patrol path takes 252 steps: It visits every end node of the labyrinth exactly once without any repeats (Fig 8A).

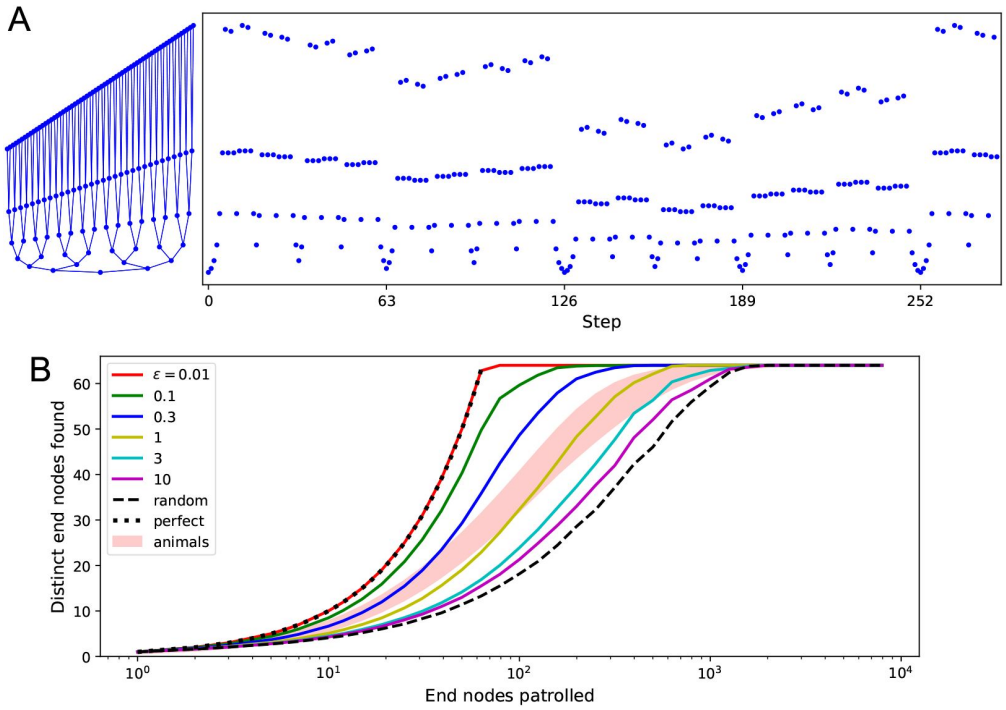

Figure 8:

Patrolling by endotaxis.

(A) Left: A binary tree maze as used in (43), plotted here so every node has a different vertical offset. **Right:** A perfect patrol path through this environment. It visits every node in 252 steps, then starts over. **(B)** Patrolling efficiency of different agents on the binary tree maze. The focus here is on the 64 end nodes of the labyrinth. We ask how many distinct end nodes are found (vertical axis) as a function of the number of end nodes visited (horizontal axis). For the perfect patrolling path, that relationship is the identity ('perfect'). For a random walk, the curve is shifted far to the right ('random', note log axis). Ten mice in (43) showed patrolling behavior within the shaded range. Solid lines are the endotaxis agent, operating at different noise levels ϵ. Note $\epsilon = 0.01$ produces perfect patrolling; in fact, panel A is a path produced by this agent. Higher noise levels lead to lower efficiency. The behavior of mice corresponds to $\epsilon \approx 1$. Gain $\gamma = 0.32$, habituation $\beta = 1.2$, with recovery time $\tau = 100$ steps.

Real mice don't quite execute this optimal path, but their patrolling behavior is much more efficient than random (Fig 8B). They avoid revisiting areas they have seen recently. Could endotaxis implement such an efficient patrol of the environment? The task is to steer the agent to locations that haven't been visited recently. One can formalize this by imagining a resource called "neglect" distributed throughout the environment. At each location neglect increases with time, then resets to zero the moment the agent visits there. To use this in endotaxis one needs a goal cell that represents neglect.

We add to the core model a goal cell that receives excitation from every map cell, via synapses that are equal and constant in strength (see clock symbol in Fig 1B). This produces a goal signal that is approximately constant everywhere in the environment. Now suppose that the point neurons undergo a form of habituation: When a point cell fires because the agent walks through its field, its sensitivity decreases by some habituation factor. That habituation then decays over time until the point cell recovers its original sensitivity. As a result, the most recently visited points on the graph produce a smaller goal signal. Endotaxis based on this goal signal will therefore lead the agent to the areas most in need of a visit.

Figure 8B illustrates that this is a powerful way to implement efficient patrols. Here we modeled endotaxis on the binary-tree labyrinth, using the standard parameters useful for

exploration, exploitation, and homing in previous sections. To this we added a habituation in the point cells with exponential recovery dynamics. Formally, the procedure is defined by Algorithm 3.

Algorithm 3 Patrolling

Parameters: gain γ, noise ϵ, habituation β, recovery time τ
Input: map synapses $\mathbf{M}$

$h_i \leftarrow 1$, for all point cells i ▷ starting sensitivity of point cell at node i
$s \leftarrow x$ ▷ begin patrolling at node x
while patrolling **do**
  $h_s \leftarrow h_s \, \mathrm{e}^{-\beta}$ ▷ habituation of point cell s
  $h_i \leftarrow 1 - (1 - h_i) \, \mathrm{e}^{-1/\tau}$, for all i ▷ resensitization of all point cells
  for all nodes j that neighbor s **do**
    $u_i(j) \leftarrow \delta_{i,j} h_j$ for all point cells i ▷ point cell output with agent at node j
    $\mathbf{v}(j) \leftarrow \left(\frac{1}{\gamma}\mathbf{1} - \mathbf{M}\right)^{-1} \mathbf{u}(j)$ ▷ map output
    $p(j) \leftarrow \frac{1}{Z} \sum_i v_i(j) + \eta$ ▷ sum of map output with noise, Z chosen so max = 1
  end for
  $s \leftarrow \arg\max_j p(j)$ ▷ choose the neighbor node with the highest patrol signal
end while

With appropriate choices of habituation β and recovery time τ the agent does in fact execute a perfect patrol path on the binary tree, traversing every edge of the graph exactly once, and then repeating that sequence indefinitely (Fig 8A). For this to work, some habituation must persist for the time taken to traverse the entire tree; in this simulation we used τ = 100 steps on a graph that requires 252 steps. As in all applications of endotaxis, the performance also depends on the readout noise ϵ. For increasing readout noise, the agent's behavior transitions gradually from the perfect patrol to a random walk (Fig 8B). The patrolling behavior of real mice is situated about halfway along that range, at an equivalent readout noise of ϵ = 1 (Fig 8B).

Finally, this suggests a unified explanation for exploration and patrolling: The agent follows the output of the "neglect" cell, which is just the sum total of the map output. However, in the early exploration phase, when the agent is still assembling the cognitive map, it gives the neglect signal zero or low weight, so the turning decisions are dominated by readout noise and produce something close to a random walk. Lateron, the agent assigns a higher weight to the neglect signal, which shifts the behavior towards systematic patrolling. In our simulations, an intrinsic readout noise of ϵ = 0.01 is sufficiently low to enable even a perfect patrol path (Fig 8B).

In summary, the core model of endotaxis can be enhanced by adding a basic form of habituation at the input neurons. In the endotaxis model this allows the agent to implement an effective patrolling policy that steers towards regions which have been neglected for a while. Of course, habituation among point cells will also change the dynamics of map learning during the exploration phase. We found that both map and goal synapses are still learned effectively, and navigation to targets is only minimally affected by habituation (Suppl Fig 10).

9 Discussion

9.1 Summary of claims

We have presented a biologically plausible neural mechanism that can support learning, navigation, and problem solving in complex environments. The algorithm, called *endotaxis*, offers an end-to-end solution for assembling a cognitive map (Fig 4), locating interesting targets within that map, navigating to those targets (Fig 5), as well as accessory functions like instant homing (Fig 7) and effective patrolling (Fig 8). Conceptually, it is related to chemotaxis, namely the ability to follow an odor signal to its source, which is shared universally by most or all motile animals. The endotaxis network creates an internal "virtual odor" which the animal can follow to reach any chosen target location (Fig 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). After sufficient exploration the agent can then navigate back to those target locations from any point in the environment (Fig 5). Beyond spatial navigation, endotaxis can also learn the solution to purely cognitive tasks (Fig 5) that can be formulated as search on a graph (Sec 3). In the following sections we consider how these findings relate to some well-established phenomena and results on animal navigation.

9.2 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 types of navigation 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 or a bat 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. As a concrete example, a mouse wishing to exit from deep inside a labyrinth (Fig 7A, (43)) can draw little benefit from knowing the distance and heading of the entrance.

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 discovered on a previous foray. It explains how an animal can return home from inside a complex environment on the first attempt (43), or navigate to a special location after encountering it just once (Figs 6, 7). But it does not implement more advanced routines of spatial reasoning, such as stringing a habitual sequence of actions together into one, or deliberating internally to plan entire routes. Clearly, given enough time in an environment, animals may develop 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 (41). Lashley (31), 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". 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. Similar behaviors

occur in arthropods (51) and humans (44) 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 seems to use the same policy of spatial sampling that it would apply to a real odor signal.

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 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.

Once the animal gains familiarity with the environment, it performs fewer of the vicarious trial and error movements, and instead moves smoothly through multiple intersections in a row (41). 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. Eventually the animal may also develop an internal forward model for the effects of its own actions, which would allow for prospective planning of an entire route (40). 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.

9.3 Brain circuits

The proposed circuitry (Fig 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 (5; 17). 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 (5; 56). 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 (25). 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 (30; 47). 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 (9; 48; 54). 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 1). [5] 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 (3), perhaps through their projections to the central complex (32), an area critical to the animal's turning behavior. Conceivably, 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 (38; 42). However, simulations showed that the synapses are substantially strengthened only among immediately adjacent place fields (39; 42), thus limiting the utility for global navigation across the environment. The learning algorithm (Alg 2) implements this local connectivity. We show that a useful global distance function emerges from the *output* of the recurrent network (Eqn 11), even though its synaptic structure is strictly local. Further, we offer a biologically realistic circuit (Fig 1B) that can read out this distance function for subsequent navigation.

9.4 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 (36), 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:

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 (29; 55). 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 (27). Both of those characteristics are expected of the goal cells in the endotaxis model.

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 (19; 55). 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 doesn't match any of the locations in the current environment (2; 15). 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 (38).

Goal cells, on the other hand, are expected to have large place fields, centered on a goal location, but extending over much of the environment, so the animal can follow the gradient of their activity (10). Indeed such cells have been reported in rat cortex (26). In the endotaxis model, a goal cell appears suddenly when the animal first arrives at a memorable location, the input synapses from the map network are potentiated, and the neuron immediately develops a place field (Fig 4). 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 (8). 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 (33). The endotaxis model relies on just such a plasticity rule for map learning (Alg 2), that can correlate events at subsequent nodes on the agent's trajectory.

9.5 Learning theories

Endotaxis can be seen as a form of reinforcement learning (50): The agent learns from rewards or punishments in the environment and develops a policy that allows for subsequent navigation to special locations. The goal signal in endotaxis plays the role of a value function in reinforcement learning theory. From experience the agent learns to compute that value function for every location and control its actions accordingly. Within the broad universe of reinforcement learning algorithms, endotaxis combines some special features as well as limitations that are inspired by empirical phenomena of animal learning, and also make it suitable for a biological implementation.

First, most of the learning happens without any reinforcement. During the exploratory random walk, endotaxis learns the topology of the environment, specifically by updating the synapses in the map network (**M** in Fig 1B). Rewards are not needed for this map learning, and indeed the goal signal remains zero during this period (Fig 4). Once a reward is encountered, the goal synapses (**G** in Fig 1B) get set, and the goal signal instantly spreads through the known portion of the environment. Thus, the agent learns how to navigate to the goal location from a single reinforcement (Fig 6). This is possible because the ground has been prepared, as it were, by learning a map. In animal behavior the acquisition of a cognitive map without rewards is called *latent learning*. Early debates in animal psychology pitched latent learning and reinforcement learning as alternative explanations (52). Instead, in the endotaxis algorithm, neither can function without the other, as the goal signal explicitly depends on both the map and goal synapses (Eqn 13, Alg 1).

In the context of reinforcement learning, the map represents a simple model of the environment on which the value function can be computed (34; 49). The neural signals in endotaxis bear some similarity to the so-called *successor representation* (12; 13; 46). 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. In that representation, there is a neuron for every state of the agent, and the activity of neuron j is the time-discounted probability that the agent will find itself at state j in the future. Similarly, the output of the endotaxis map network is related to future states of the agent (Eqns 5, 18). However, there is an important difference: The successor representation (at least as currently discussed) is designed to improve learning under a particular policy (13; 16; 22). By contrast the endotaxis map network is independent of policy; it just reflects the objective connectivity of the environment. Knowing that connectivity is a foundation for developing any specific policy. The algorithm for learning the map (Alg 2) is insensitive to what policy the agent uses: A synapse between map cells gets formed when a particular link is traveled, regardless of why it is traveled. A systematic walk through the environment (Fig 8) learns the exact same map synapses as a random walk.

Second, endotaxis does not tabulate the list of available actions at each state. That information remains externalized in the environment: The agent simply tries whatever actions are available at the moment, then picks the best one. This is a characteristically biological mode of action and most organisms have a behavioral routine that executes such

trial-and-error. This "externalized cognition" simplifies the learning task: For any given navigation policy the agent needs to learn only one scalar function of location, namely the goal signal. By comparison, many machine learning algorithms develop a value function for state-action pairs, which then allows more sophisticated planning (34; 50). The relative simplicity of the endotaxis circuit depends on the limitation to learning only state functions.

Finally, endotaxis is "always on". There is no separation of learning from recall. The map and goal synapses can continue to update even while the agent is navigating, homing, or patrolling. Learning continues to happen automatically "under the hood". In fact, many policies are learned simultaneously: Each goal cell represents a different value function, and their synapses all are updated in parallel as the agent encounters different targets. Meanwhile the animal pursues its current needs by choosing one of the goal signals (with the mode switch in Fig 1B) and feeding it to the chemotaxis module for decision making.

9.6 Outlook

Burgess and O'Keeffe (10) pointed out some time ago the benefits of modeling spatial learning with explicit neural circuits rather than purely conceptual arguments: For one, it tests whether a proposed explanation actually works inside of biological realism; second it can offer an interpretation of the profusion of different kinds of place cells one might find in any given brain (24). An analogy to the visual system is useful here: There is a profusion of neurons with visual receptive fields. In principle these are all "light cells", but by now it is well understood that they appear at different levels of the visual circuitry and play entirely different roles. At the bottom of the hierarchy are photoreceptors that respond when light appears at a particular location. Towards the end of the visual system are neurons that respond selectively to faces independent of viewpoint (20). Sophisticated circuit models exist to explain the processing all the way from the retina to IT cortex (23; 57). A simple place cell is like a photoreceptor: It responds when the animal is at a particular location. How does the brain perform sophisticated spatial cognition based on that elementary input? To reach an understanding comparable to that of the visual system, we should invest further in end-to-end models for navigation that use biologically plausible neural circuits.

A Supplement

A.1 A neuromorphic function to compute the shortest distance on a graph

Here we prove some of the assertions in the text about the relationship between endotaxis goal signals and the distance between two points on a graph. We begin with a more general discussion of graph distance. For an agent navigating on a graph it is very useful to know the shortest graph distance between any two nodes

$$D_{ij} = \text{minimum number of steps needed to reach node } i \text{ from node } j \tag{14}$$

Given this information, one can navigate the shortest route from x to y: for each of the neighbors of x, look up its distance to y and step to the neighbor with the shortest distance. Then repeat that process until y is reached. Thus the shortest route can be navigated one step at a time without any high-level advanced planning. This is the core idea behind endotaxis.

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) (58). Generally, an APSP algorithm delivers a matrix containing the distances D_{ij} for all pairs of nodes. The Floyd-Warshall algorithm (18) 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} (Eqn 1), then with a suitably small positive value of γ the shortest path distances are given by

$$D_{ij} = \left\lceil \frac{\log\left[(\mathbf{1} - \gamma\,\mathbf{A})^{-1}\right]_{ij}}{\log\gamma} \right\rceil \tag{15}$$

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

$$N_{ij}^{(k+1)} = \sum_l A_{il} N_{lj}^{(k)} \tag{16}$$

The RHS corresponds to multiplication by **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

$$D_{ij} = \min_k N_{ij}^{(k)} > 0 \tag{17}$$

Now consider the Taylor series

$$\begin{aligned}\mathbf{Y} &= (\mathbf{1} - \gamma\mathbf{A})^{-1} \\ &= \mathbf{1} + \gamma\mathbf{A} + \gamma^2\mathbf{A}^2 + \dots\end{aligned} \tag{18}$$

Then

$$Y_{ij} = \sum_{k=0}^{\infty} N_{ij}^{(k)}\gamma^k = N_{ij}^{(D_{ij})}\gamma^{D_{ij}} + N_{ij}^{(D_{ij}+1)}\gamma^{D_{ij}+1} + \dots \tag{19}$$

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

$$\gamma^{D_{ij}} < Y_{ij} < \gamma^{D_{ij}-1} \tag{20}$$

The left inequality is obvious from Eqn 19 because $N_{ij}^{(D_{ij})} \geq 1$ by Eqn 17.

To understand the right inequality, note first that $N_{ij}^{(k)}$ is bounded by a geometric series. From Eqn 16 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 19

$$Y_{ij} < (q\gamma)^{D_{ij}} + (q\gamma)^{D_{ij}+1} + \dots = \frac{(q\gamma)^{D_{ij}}}{1 - q\gamma} \tag{21}$$

This expression is $< \gamma^{D_{ij}-1}$ (Eqn 20) as long as

$$\gamma < \frac{1}{q + q^{D_{ij}}} \tag{22}$$

In addition, because

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

this is satisfied if one chooses γ such that

$$\gamma < \frac{1}{q + q^n} \tag{23}$$

With that condition on γ, the inequality (20) holds, and taking the logarithm on both sides leads to the desired result:

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

As shown in the text (Eqn 10), the endotaxis network, in its linear rate approximation, computes a goal signal equal to the scalar products of the column-vectors in $\mathbf{Y}$, namely

$$E_{ij} = \text{ goal signal from node } j \text{ to } i = \gamma^2 \sum_k Y_{ki} Y_{kj} \tag{24}$$

To understand how that goal signal E_{ij} varies with distance, one can follow arguments parallel to those that led to Eqn 19. Using the upper bound by the geometric series (Eqn 21) and inserting in Eqn 24 one finds again that it is possible to choose a γ small enough to satisfy

$$\gamma^{D_{ij}} < \frac{E_{ij}}{\gamma^2} < \gamma^{D_{ij}-1} \tag{25}$$

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

In summary, a recurrent neural network seems ideally suited to compute the distance between nodes on a graph, if the nodes are sparsely represented in the network's inputs, and the recurrent connections reflect the connections of the graph. Ultimately, this derives from the correspondence between the network's transfer function (Eqn 5) and the function that delivers APSP on a graph (Eqn 15).

A.2 The critical gain value

As elaborated in Section 3, there is a benefit to raising the gain γ of the map neurons, so as to limit the sharp decline of the goal signal across distance. However, there is an upper limit. Recall that the argument linking the recurrent network function to graph distances traces back to the Taylor expansion in Eqn 18:

$$(\mathbf{1} - \gamma \mathbf{A})^{-1} = \mathbf{1} + \gamma \mathbf{A} + \gamma^2 \mathbf{A}^2 + \dots$$

For a real function $(1 - x)^{-1}$, this Taylor series has a convergence radius of $|x| < 1$. The corresponding condition for the matrix series is that the spectral radius ρ of $\gamma\mathbf{A}$ must be < 1:

$$1 > \rho(\gamma \mathbf{A}) = \gamma\, \rho(\mathbf{A}) = \gamma \max_i |\lambda_i|$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\mathbf{A}$. So the critical upper bound on γ is

$$\gamma < \gamma_c \equiv \frac{1}{\rho(\mathbf{A})} = \frac{1}{\text{largest absolute eigenvalue of } \mathbf{A}}$$

A.3 Average navigated distance

In the text we often assess the performance of an endotaxis agent by considering point-to-point navigation between all pairs of points on a graph. Given the readout noise ϵ that affects the goal signal, navigation is a stochastic process with many random decisions along the route. Different random instantiations of the process will produce routes of different lengths. Fortunately, there is a way to calculate the expectation value of the route length without any Monte-Carlo simulation.

Consider navigation to goal node y. From the state of the network (**M** and **G**) we compute the goal signal E_{yj} at every node j. When the agent is at node j it chooses among the neighbor nodes the one with the highest sum of goal signal and noise (1). Based on the goal signal E_{yj} and the noise ϵ one can compute the probability for each such possible step from j. This leads to a transition matrix for the random walk

$$T_{ij}^{(y)} = \text{probability of stepping to } i \text{ when at } j \text{ while in pursuit of } y$$

Subsequent decisions along the route are independent of each other. Hence the process is a Markov chain. Then we make use of a well-known result for first-capture times on a Markov chain to compute the expected number of steps to arrival at y starting from any node x.

Note the method assumes that the process is stationary Markov, such that the goal signal E_{xy} does not change in the course of navigation. In our analysis of patrolling (Figs 8 and 10) this assumption is violated, because the habituation state of the point cells depends on what path the agent took to the current node. In those cases we resorted to Monte Carlo simulations to estimate the distribution of route lengths.

A.4 Simulations

Numerical simulations were performed as described, see Algorithms 1, 2, 3, 4. Parameter settings are listed in the text and figure captions. The sensitivity to parameters is reported in Figure 6. Code that produced all the results is available in a public repository.

A.5 Forgetting of links and resources

In section 4 we discuss the learning algorithm that acquires the connectivity of the environment and the locations of resources. It reacts rapidly to the appearance of new links in the environment: As soon as the agent travels from one point to another, the synapse between the corresponding map cells gets established. Suppose now that a previously existing link becomes blocked: How can one remove the corresponding synapse from the map? A simple solution would be to let all synapses decay over time, balanced by strengthening whenever a link gets traveled. In that case the entire map would be forgotten when the animal goes to sleep for a few hours, whereas it is clear that animals retain such maps over many days. Instead, one wants a mode of *active* forgetting: Memory of the link from node i to j should be weakened only if the agent find itself at node i and repeatedly chooses not to go to j. One can formalize this in the following algorithm, which differs only slightly from Alg 2:

Algorithm 4 Learning and forgetting

Parameters: gain γ, threshold θ, goal learning rate α, forgetting rate δ
Input: adjacency matrix $\mathbf{A}$, resource signals $\mathbf{F}$

$\mathbf{M} \leftarrow 0$ ▷ initiate map synapses at 0
$\mathbf{G} \leftarrow 0$ ▷ initiate goal synapses at 0
$t \leftarrow 0$ ▷ t counts the steps
$s(t) \leftarrow x$ ▷ start random walk at x
while learning **do**
  $t \leftarrow t + 1$
  $s(t) \leftarrow$ a random neighbor of $s(t-1)$ ▷ continue the random walk
  $u_i(t) \leftarrow \delta_{i,s(t)}$ for every point cell i ▷ point cell output
  $\mathbf{v}(t) \leftarrow \left(\frac{1}{\gamma}\mathbf{1} - \mathbf{M}\right)^{-1} \mathbf{u}(t)$ ▷ map cell output
  for all map cell pairs (i, j) **do**
    if $v_j(t-1) > \theta$ **then** ▷ if pre-synaptic high
      if $v_i(t) > \theta$ **then** ▷ if post-synaptic also high
        $M_{ij} \leftarrow 1$ ▷ potentiate the synapse
      else ▷ if post-synaptic low
        $M_{ij} \leftarrow \mathrm{e}^{-\delta} M_{ij}$ ▷ depress the synapse
      end if
    end if
  end for
  $\mathbf{r} \leftarrow \mathbf{G}\mathbf{v}(t)$ ▷ goal signals
  for every goal neuron k **do**
    $D \leftarrow F_{k,s(t)} - r_k$ ▷ difference between resource signal and prediction from the map
    if $D > 0$ **then** ▷ if the resource signal exceeds the prediction from the map
      for every map neuron j **do**
        $G_{kj} \leftarrow G_{kj} + \alpha D v_j(t)$ ▷ potentiate goal synapses
      end for
    else ▷ if resource signal less than prediction
      for every map neuron j **do**
        $G_{kj} \leftarrow \mathrm{e}^{-\delta v_j} G_{kj}$ ▷ depress goal synapses
      end for
    end if
  end for
end while

Here the added parameter δ determines how much a map synapse gets depressed each time the corresponding link is not chosen. Similarly, goal synapses decay if their prediction for a resource exceeds the resource signal received by the goal cell. The synaptic learning rule resembles the BCM rule (7): Synaptic modification is conditional on presynaptic activity, and leads to either potentiation or depression depending on the level of post-synaptic activity.

Figure 9 illustrates this process with a simulation analogous to Fig 4. The agent explores a ring graph by a random walk. At some point a new link appears clear across the ring. Later on that link disappears again. Acquisition of the link happens very quickly, within a single time step (Fig 9A, C). Forgetting that link takes longer, on the order of several hundred steps (Fig 9A, D, E). In this simulation δ = 0.1, so the map synapses decay by about 10% whenever a link is not traveled. One could of course accelerate that with a higher δ, but at the cost of destabilizing the entire map. Even the synapses for intact links get depressed frequently (Fig 9E), because the random choices of the agent lead it to take any given link only a fraction of the time.

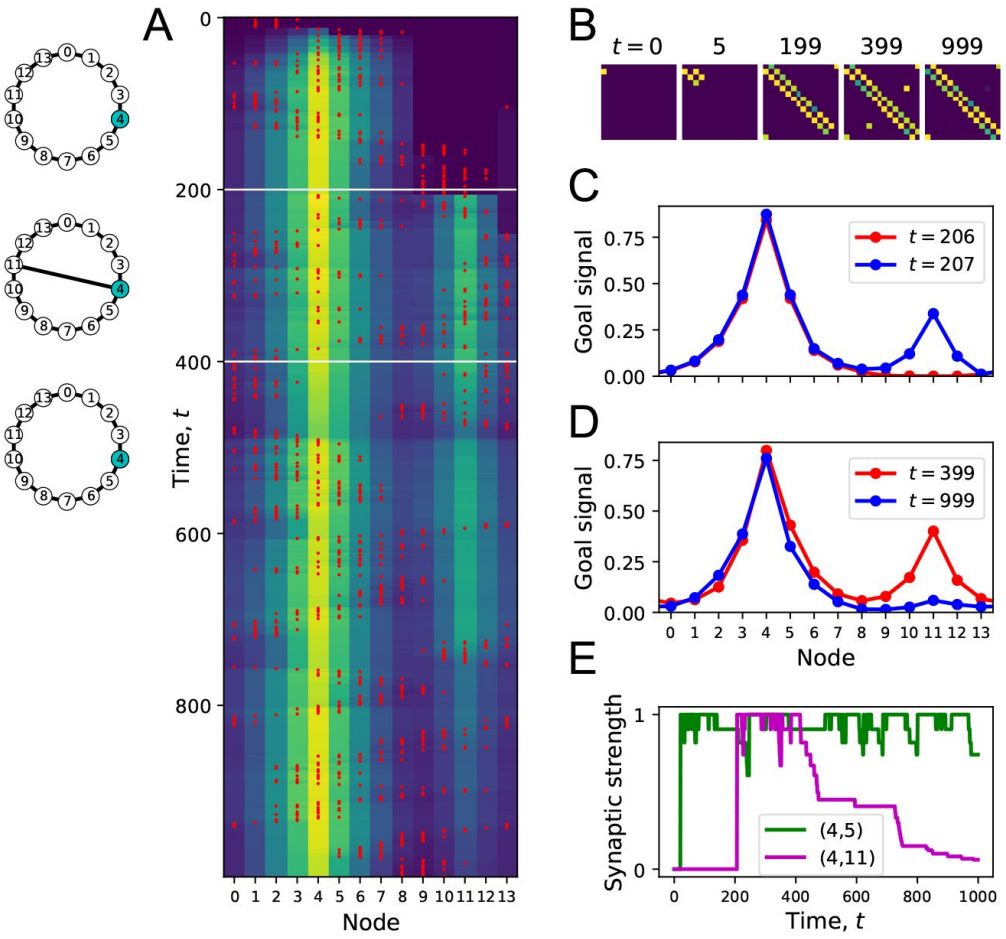

Figure 9:

Forgetting a link during exploration.

(A) Simulation of a random walk on a ring with 14 nodes as in Fig 4. Left: Layout of the ring, with resource locations marked in blue. The walk progresses in 1000 time steps (top to bottom); with the agent's position marked in red (nodes 0-13, horizontal axis). At each time the color map shows the goal signal that would be produced if the agent were at position 'Node'. White horizontal lines mark the appearance of a new link between nodes 4 and 11 at t = 200, and disappearance of that link at t = 400. **(B)** The matrix **M** of map synapses at various times. The pixel in row i and column j represents the matrix element M_{ij}. Color purple = 0. Note the first few steps (number above graph) each add a new synapse. Eventually, **M** reflects the adjacency matrix of nodes on the graph, and changes as a link is added and removed. **(C)** Goal signals just before and just after the agent travels the new link. **(D)** Goal signals just before the link disappears and at the end of the walk. **(E)** Strength of two synapses in the map, $M_{4,5}$ and $M_{4,11}$, plotted against time during the random walk. Model parameters: γ = 0.32, θ = 0.27, α = 1, δ = 0.1.

One limitation of the endotaxis agent is that it does not keep a record of what actions are available at each node. Instead, it leaves that information in the environment (see Discussion) and simply tries all the actions that are available. When faced with a blocked tunnel, the endotaxis agent does not know that this was previously available. Clearly, a more advanced model of the world that includes a state-action table would allow more effective editing of the cognitive map.

A.6 Habituation in point cells

In section 8 we discuss an extension of the core endotaxis model in which a point neuron undergoes habituation after the agent passes through its node. With every visit, the neuron's sensitivity declines by a factor $e^{-\beta}$. Between visits the sensitivity gradually returns towards 1 with an exponential recovery time of τ steps, see Algorithm 3.

This addition to the model changes the dynamics of the network input throughout the phases of exploration, navigation, and patrolling. We explored how the resulting

and comparing to the basic model with no habituation (β = 0). During the learning phase, when the map and goal synapses are established via a random walk, the main change is that it takes somewhat longer to learn the map. This is because synaptic updates happen only when both pre- and post-synaptic map cells exceed a threshold (see Alg 2), and that requires that both of the respective point neurons be in a high-sensitivity state. In our simulations we extended the random walk for exploration by a factor of 3. Remarkably all the parameter settings (γ, θ, α) that support learning and navigating under standard conditions (Fig 6), work well with habituation as well.

To illustrate the overall effect that habituation has on performance, we simulated navigation between all pairs of nodes on the binary-tree graph of Fig 8. For every pair of start and end nodes we asked how the actual navigated distance compared to the shortest graph distance. Figure 10 shows that performance is affected only slightly. At the standard noise value ϵ = 0.01 used in other simulations, the range of navigation extends over 10 steps under both conditions.

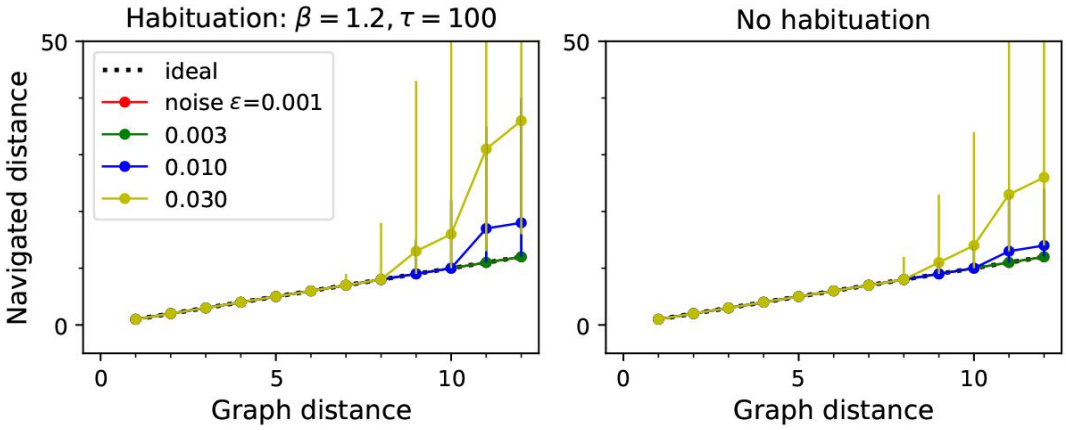

Figure 10:

Navigation performance with and without habituation.

Navigated distance on the binary-tree maze, displayed as in Fig 5E. **Left:** An agent with strong habituation: β = 1.2, τ = 100. **Right:** no habituation: β = 0. The agent learned the map and the goal signals for every node during a random walk with 30,000 steps. Then the agent navigated between all pairs of points on the maze. Graphs show the median ± 10/90 percentile of the navigated distance for all routes with the same graph distance. Other model parameters: γ = 0.32, θ = 0.27, α = 1, ϵ as listed.

Data and code availability

Data and code to reproduce the reported results are available at https://github.com/markusmeister/Endotaxis-2022. Following acceptance of the manuscript they will be archived in a permanent public repository.

Acknowledgements

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.

Author contributions

Conception of the study TZ, MR, PP, MM; Numerical work TZ, PP, MM; Analytical work MM; Drafting the manuscript MM; Revision and approval TZ, MR, PP, MM.

Competing interests

The authors declare no competing interests.

Colleagues

We thank Kyu Hyun Lee and Ruben Portugues for comments.

References

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

[2] Alme C. B. , Miao C. , Jezek K. , Treves A. , Moser E. I. , Moser M.-B. (2014) **Place cells in the hippocampus: Eleven maps for eleven rooms 111:**18428–18435

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

[4] Baker K. L. , Dickinson M. , Findley T. M. , Gire D. H. , Louis M. , Suver M. P. , Verhagen J. V. , Nagel K. I. , Smear M. C. (2018) **Algorithms for Olfactory Search across Species 38:**9383–9389

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

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

[7] Bienenstock E. L. , Cooper L. N. , Munro P. W. (1982) **Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex 2:**32–48

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

[9] Buehlmann C. , Wozniak B. , Goulard R. , Webb B. , Graham P. , Niven J. E. (2020) **Mushroom Bodies Are Required for Learned Visual Navigation, but Not for Innate Visual Behavior, in Ants 30:**3438–3443

[10] Burgess N. , O'Keefe J. (1996) **Neuronal computations underlying the firing of place cells and their role in navigation 6:**749–762

[11] Collett T. S. , Collett M. (2002) **Memory use in insect visual navigation 3:**542–552

[12] Corneil D. S. , Gerstner W. (2015) **Attractor Network Dynamics Enable Preplay and Rapid Path Planning in Maze–like Environments** *In Advances in Neural Information Processing Systems*

[13] Dayan P. (1993) **Improving generalization for temporal difference learning: The successor representation 5:**613–624

[14] Dayan P. , Abbott L. F. (2001) **Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems** *Computational Neuroscience*

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

[16] Fang C. , Aronov D. , Abbott L. F. , Mackevicius E. (2022) **Neural learning rules for generating flexible predictions and computing the successor representation**

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

[18] Floyd R. W. (1962) **Algorithm 97: Shortest path 5:**345

[19] Frank L. M. , Stanley G. B. , Brown E. N. (2004) **Hippocampal Plasticity across Multiple Days of Exposure to Novel Environments 24:**7681–7689

[20] Freiwald W. A. , Tsao D. Y. (2010) **Functional compartmentalization and viewpoint generalization within the macaque face-processing system** *Science* **330:**845–51

[21] Galtier M. , Faugeras O. , Bressloff P. (2012) **Hebbian Learning of Recurrent Connections: A Geometrical Perspective** *Neural computation* **24:**2346–83

[22] Geerts J. P. , Chersi F. , Stachenfeld K. L. , Burgess N. (2020) **A general model of hippocampal and dorsal striatal learning and decision making 117:**31427–31437

[23] Gollisch T. , Meister M. (2010) **Eye smarter than scientists believed: Neural computations in circuits of the retina** *Neuron* **65:**150–64

[24] Grieves R. M. , Jeffery K. J. (2017) **The representation of space in the brain** *Behavioural Processes* **135:**113–131

[25] Heisenberg M. (2003) **Mushroom body memoir: From maps to models 4:**266–275

[26] Hok V. , Save E. , Lenck-Santini P. P. , Poucet B. (2005) **Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex 102:**4602–4607

[27] Hollup S. A. , Molden S. , Donnett J. G. , Moser M.-B. , Moser E. I. (2001) **Accumulation of Hippocampal Place Fields at the Goal Location in an Annular Watermaze Task 21:**1635–1644

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

[29] Kjelstrup K. B. , Solstad T. , Brun V. H. , Hafting T. , Leutgeb S. , Witter M. P. , Moser E. I. , Moser M.-B. (2008) **Finite scale of spatial representation in the hippocampus 321:**140–143

[30] Knaden M. , Graham P. , Berenbaum M. R. (2016) **The Sensory Ecology of Ant Navigation: From Natural Environments to Neural Mechanisms** *Annual Review of Entomology* **61:**63–76

[31] Lashley K. S. (1912) **Visual discrimination of size and form in the albino rat 2:**310–331

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

[33] Magee J. C. , Grienberger C. (2020) **Synaptic Plasticity Forms and Functions 43:**95–117

[34] Moerland T. M. , Broekens J. , Jonker C. M. (2020) **Model-based reinforcement learning: A survey**

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

[36] Moser M.-B. , Rowland D. C. , Moser E. I. (2015) **Place Cells, Grid Cells, and Memory 7:**

[37] Müller M. , Wehner R. (1988) **Path integration in desert ants, Cataglyphis fortis 85:**5287–5290

[38] Muller R. U. , Kubie J. L. , Saypoff R. (1991) **The hippocampus as a cognitive graph (abridged version) 1:**243–246

[39] Muller R. U. , Stead M. , Pach J. (1996) **The hippocampus as a cognitive graph 107:**663–694

[40] Nyberg N. , Duvelle É. , Barry C. , Spiers H. J. (2022) **Spatial goal coding in the hippocampal formation 110:**394–422

[41] Redish A. D. (2016) **Vicarious trial and error 17:**147–159

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

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

[44] Santos-Pata D. , Verschure P. F. M. J. (2018) **Human Vicarious Trial and Error Is Predictive of Spatial Navigation Performance** *Frontiers in Behavioral Neuroscience* **12:**237

[45] Sosa M. , Giocomo L. M. (2021) **Navigating for reward** 1–16

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

[47] Steck K. , Hansson B. S. , Knaden M. (2009) **Smells like home: Desert ants, Cataglyphis fortis, use olfactory landmarks to pinpoint the nest 6:**5

[48] Sun X. , Yue S. , Mangan M. (2020) **A decentralised neural model explaining optimal integration of navigational strategies in insects** *eLife* **9:**

[49] Sutton R. S. (1990) **Integrated architectures for learning, planning, and reacting based on approximating dynamic programming** *In Machine Learning Proceedings 1990* 216–224

[50] Sutton R. S. , Barto A. G. (2018) **Reinforcement Learning: An Introduction**

[51] Tarsitano M. (2006) **Route selection by a jumping spider (Portia labiata) during the locomotory phase of a detour 72:**1437–1442

[52] Thistlethwaite D. (1951) **A critical review of latent learning and related experiments 48:**97–129

[53] Tolman E. C. (1948) **Cognitive maps in rats and men 55:**189–208

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

[55] Wilson M. A. , McNaughton B. L. (1993) **Dynamics of the hippocampal ensemble code for space 261:**1055–1058

[56] Wolpert D. M. , Miall R. C. , Kawato M. (1998) **Internal models in the cerebellum 2:**338–347

[57] Zhuang C. , Yan S. , Nayebi A. , Schrimpf M. , Frank M. C. , DiCarlo J. J. , Yamins D. L. K. (2021) **Unsupervised neural network models of the ventral visual stream 118:**

[58] Zwick U. , auf der Heide F. M. (2001) **Exact and approximate distances in graphs — A survey** *Algorithms — ESA 2001* 33–48

Author information

1. **Tony Zhang**
   Division of Biology and Biological Engineering, California Institute of Technology
2. **Matthew Rosenberg**
   Division of Biology and Biological Engineering, California Institute of Technology
3. **Pietro Perona**
   Division of Engineering and Applied Science, California Institute of Technology
4. **Markus Meister**
   Division of Biology and Biological Engineering, California Institute of Technology

   For correspondence:
   meister@caltech.edu

Editors

Reviewing Editor
Srdjan Ostojic
Ecole Normale Superieure Paris, France

Senior Editor
Timothy Behrens
University of Oxford, United Kingdom

Reviewer #1 (Public Review):

This paper presents a highly compelling and novel hypothesis for how the brain could generate signals to guide navigation towards remembered goals. Under this hypothesis, which the authors call "Endotaxis", the brain co-opts its ancient ability to navigate up odor gradients (chemotaxis) by generating a "virtual odor" that grows stronger the closer the animal is to a goal location. This idea is compelling from an evolutionary perspective and a mechanistic perspective. The paper is well-written and delightful to read.

The authors develop a detailed model of how the brain may perform "Endotaxis", using a variety of interconnected cell types (point, map, and goal cells) to inform the chemotaxis system. They tested the ability of this model to navigate in several state spaces, representing

both physical mazes and abstract cognitive tasks. The Endotaxis model performed reasonably well across different environments and different types of goals.

The authors further tested the model using parameter sweeps and discovered a critical level of network gain, beyond which task performance drops. This critical level approximately matched analytical derivations.

My main concern with this paper is that the analysis of the critical gain value (gamma_c) is incomplete, making the implications of these analyses unclear. There are several different reasonable ways in which the Endotaxis map cell representations might be normalized, which I suspect may lead to different results. Specifically, the recurrent connections between map cells may either be an adjacency matrix, or a normalized transition matrix. In the current submission, the recurrent connections are an un-normalized adjacency matrix. In a previous preprint version of the Endotaxis manuscript, the recurrent connections between the map cells were learned using Oja's rule, which results in a normalized state-transition matrix (see "Appendix 5: Endotaxis model and the successor representation" in "Neural learning rules for generating flexible predictions and computing the successor representation", your reference 17). The authors state "In summary, this sensitivity analysis shows that the optimal parameter set for endotaxis does depend on the environment". Is this statement, and the other conclusions of the sensitivity analysis, still true if the learned recurrent connections are a properly normalized state-transition matrix?

Overall, this paper provides a very compelling model for how neural circuits may have evolved the ability to navigate towards remembered goals, using ancient chemotaxis circuits.

This framework will likely be very important for understanding how the hippocampus (and other memory/navigation-related circuits) interfaces with other processes in the brain, giving rise to memory-guided behavior.

Reviewer #2 (Public Review):

The manuscript presents a computational model of how an organism might learn a map of the structure of its environment and the location of valuable resources through synaptic plasticity, and how this map could subsequently be used for goal-directed navigation.

The model is composed of 'map cells', which learn the structure of the environment in their recurrent connections, and 'goal-cell' which stores the location of valued resources with respect to the map cell population. Each map cell corresponds to a particular location in the environment due to receiving external excitatory input at this location. The synaptic plasticity rule between map cells potentiates synapses when activity above a specified threshold at the pre-synaptic neuron is followed by above-threshold activity at the post-synaptic neuron. The threshold is set such that map neurons are only driven above this plasticity threshold by the external excitatory input, causing synapses to only be potentiated between a pair of map neurons when the organism moves directly between the locations they represent. This causes the weight matrix between the map neurons to learn the adjacency for the graph of locations in the environment, i.e. after learning the synaptic weight matrix matches the environment's adjacency matrix. Recurrent activity in the map neuron population then causes a bump of activity centred on the current location, which drops off exponentially with the diffusion distance on the graph. Each goal cell receives input from the map cells, and also from a 'resource cell' whose activity indicates the presence or absence of a given values resource at the current location. Synaptic plasticity potentiates map-cell to goal-cell synapses in proportion to the activity of the map cells at time points when the resource cell is active. This causes goal cell activity to increase when

the activity of the map cell population is similar to the activity where the resource was obtained. The upshot of all this is that after learning the activity of goal cells decreases exponentially with the diffusion distance from the corresponding goal location. The organism can therefore navigate to a given goal by doing gradient ascent on the activity of the corresponding goal cell. The process of evaluating these gradients and using them to select actions is not modelled explicitly, but the authors point to the similarity of this mechanism to chemotaxis (ascending a gradient of odour concentration to reach the odour source), and the widespread capacity for chemotaxis in the animal kingdom, to argue for its biological plausibility.

The ideas are interesting and the presentation in the manuscript is generally clear. The two principle limitations of the manuscript are: i) Many of the ideas that the model implements have been explored in previous work. ii) The mapping of the circuit model onto real biological systems is pretty speculative, particularly with respect to the cerebellum.

Regarding the novelty of the work, the idea of flexibly navigating to goals by descending distance gradients dates back to at least Kaelbling (Learning to achieve goals, IJCAI, 1993), and is closely related to both the successor representation (cited in manuscript) and Linear Markov Decision Processes (LMDPs) (Piray and Daw, 2021, 2023 eLife. https://doi.org/10.1038/s41467-021-25123-3, Todorov, 2009 2023 eLife. https://doi.org/10.1073/pnas.0710743106). The specific proposal of navigating to goals by doing gradient descent on diffusion distances, computed as powers of the adjacency matrix, is explored in Baram et al. 2018 (2023 eLife. https://doi.org/10.1101/421461), and the idea that recurrent neural networks whose weights are the adjacency matrix can compute diffusion distances are explored in Fang et al. 2022 (2023 eLife. https://doi.org/10.1101/2022.05.18.492543). Similar ideas about route planning using the spread of recurrent activity are also explored in Corneil and Gerstner (2015, cited in manuscript). Further exploration of this space of ideas is no bad thing, but it is important to be clear where prior literature has proposed closely related ideas.

Regarding whether the proposed circuit model might plausibly map onto a real biological system, I will focus on the mammalian brain as I don't know the relevant insect literature. It was not completely clear to me how the authors think their model corresponds to mammalian brain circuits. When they initially discuss brain circuits they point to the cerebellum as a plausible candidate structure (lines 520-546). Though the correspondence between cerebellar and model cell types is not very clearly outlined, my understanding is they propose that cerebellar granule cells are the 'map-cells' and Purkinje cells are the 'goal-cells'. I'm no cerebellum expert, but my understanding is that the granule cells do not have recurrent excitatory connections needed by the map cells. I am also not aware of reports of place-field-like firing in these cell populations that would be predicted by this correspondence. If the authors think the cerebellum is the substrate for the proposed mechanism they should clearly outline the proposed correspondence between cerebellar and model cell types and support the argument with reference to the circuit architecture, firing properties, lesion studies, etc.

The authors also discuss the possibility that the hippocampal formation might implement the proposed model, though confusingly they state 'we do not presume that endotaxis is localized to that structure' (line 564). A correspondence with the hippocampus appears more plausible than the cerebellum, given the spatial tuning properties of hippocampal cells, and the profound effect of lesions on navigation behaviours. When discussing the possible relationship of the model to hippocampal circuits it would be useful to address internally generated sequential activity in the hippocampus. During active navigation, and when animals exhibit vicarious trial and error at decision points, internally generated sequential activity of hippocampal place cells appears to explore different possible routes ahead of the animal (Kay et al. 2020, 2023 eLife. https://doi.org/10.1016/j.cell.2020.01.014, Reddish 2016, 2023 eLife. https://doi.org/10.1038/nrn.2015.30). Given the emphasis the model places on

sampling possible future locations to evaluate goal-distance gradients, this seems highly relevant. Also, given the strong emphasis the authors place on the relationship of their model to chemotaxis/odour-guided navigation, it would be useful to discuss brain circuits involved in chemotaxis, and whether/how these circuits relate to those involved in goal-directed navigation, and the proposed model.

Finally, it would be useful to clarify two aspects of the behaviour of the proposed algorithm:

1. When discussing the relationship of the model to the successor representation (lines 620-627), the authors emphasise that learning in the model is independent of the policy followed by the agent during learning, while the successor representation is policy dependent. The policy independence of the model is achieved by making the synapses between map cells binary (0 or 1 weight) and setting them to 1 following a single transition between two locations. This makes the model unsuitable for learning the structure of graphs with probabilistic transitions, e.g. it would not behave adaptively in the widely used two-step task (Daw et al. 2011, 2023 eLife. https://doi.org/10.1016/j.neuron.2011.02.027) as it would fail to differentiate between common and rare transitions. This limitation should be made clear and is particularly relevant to claims that the model can handle cognitive tasks in general. It is also worth noting that there are algorithms that are closely related to the successor representation, but which learn about the structure of the environment independent of the subjects policy, e.g. the work of Kaelbling which learns shortest path distances, and the default representation in the work of Piray and Daw (both referenced above). Both these approaches handle probabilistic transition structures.

2. As the model evaluates distances using powers of adjacency matrix, the resulting distances are diffusion distances not shortest path distances. Though diffusion and shortest path distances are usually closely correlated, they can differ systematically for some graphs (see Baram et al. cited above).

Reviewer #3 (Public Review):

This paper argues that it has developed an algorithm conceptually related to chemotaxis that provides a general mechanism for goal-directed behaviour in a biologically plausible neural form.

The method depends on substantial simplifying assumptions. The simulated animal effectively moves through an environment consisting of discrete locations and can reliably detect when it is in each location. Whenever it moves from one location to an adjacent location, it perfectly learns the connectivity between these two locations (changes the value in an adjacency matrix to 1). This creates a graph of connections that reflects the explored environment. In this graph, the current location gets input activation and this spreads to all connected nodes multiplied by a constant decay (adjusted to the branching number of the graph) so that as the number of connection steps increases the activation decreases. Some locations will be marked as goals through experiencing a resource of a specific identity there, and subsequently will be activated by an amount proportional to their distance in the graph from the current location, i.e., their activation will increase if the agent moves a step closer and decrease if it moves a step further away. Hence by making such exploratory movements, the animal can decide which way to move to obtain a specified goal.

I note here that it was not clear what purpose, other than increasing the effective range of activation, is served by having the goal input weights set based on the activation levels when the goal is obtained. As demonstrated in the homing behaviour, it is sufficient to just have a goal connected to a single location for the mechanism to work (i.e., the activation at that

location increases if the animal takes a step closer to it); and as demonstrated by adding a new graph connection, goal activation is immediately altered in an appropriate way to exploit a new shortcut, without the goal weights corresponding to this graph change needing to be relearnt.

Given the abstractions introduced, it is clear that the biological task here has been reduced to the general problem of calculating the shortest path in a graph. That is, no real-world complications such as how to reliably recognise the same location when deciding that a new node should be introduced for a new location, or how to reliably execute movements between locations are addressed. Noise is only introduced as a 1% variability in the goal signal. It is therefore surprising that the main text provides almost no discussion of the conceptual relationship of this work to decades of previous work in calculating the shortest path in graphs, including a wide range of neural- and hardware-based algorithms, many of which have been presented in the context of brain circuits.

The connection to this work is briefly made in appendix A.1, where it is argued that the shortest path distance between two nodes in a directed graph can be calculated from equation 15, which depends only on the adjacency matrix and the decay parameter (provided the latter falls below a given value). It is not clear from the presentation whether this is a novel result. No direct reference is given for the derivation so I assume it is novel. But if this is a previously unknown solution to the general problem it deserves to be much more strongly featured and either way it needs to be appropriately set in the context of previous work.

Once this principle is grasped, the added value of the simulated results is somewhat limited. These show: 1) in practical terms, the spreading signal travels further for a smaller decay but becomes erratic as the decay parameter (map neuron gain) approaches its theoretical upper bound and decreases below noise levels beyond a certain distance. Both follow the theory. 2) that different graph structures can be acquired and used to approach goal locations (not surprising) .3) that simultaneous learning and exploitation of the graph only minimally affects the performance over starting with perfect knowledge of the graph. 4) that the parameters interact in expected ways. It might have been more impactful to explore whether the parameters could be dynamically tuned, based on the overall graph activity.

Perhaps the most biologically interesting aspect of the work is to demonstrate the effectiveness, for flexible behaviour, of keeping separate the latent learning of environmental structure and the association of specific environmental states to goals or values. This contrasts (as the authors discuss) with the standard reinforcement learning approach, for example, that tries to learn the value of states that lead to reward. Examples of flexibility include the homing behaviour (a goal state is learned before any of the map is learned) and the patrolling behaviour (a goal cell that monitors all states for how recently they were visited). It is also interesting to link the mechanism of exploration of neighbouring states to observed scanning behaviours in navigating animals.

The mapping to brain circuits is less convincing. Specifically, for the analogy to the mushroom body, it is not clear what connectivity (in the MB) is supposed to underlie the graph structure which is crucial to the whole concept. Is it assumed that Kenyon cell connections perform the activation spreading function and that these connections are sufficiently adaptable to rapidly learn the adjacency matrix? Is there any evidence for this? As discussed above, the possibility that an algorithm like 'endotaxis' could explain how the rodent place cell system could support trajectory planning has already been explored in previous work so it is not clear what additional insight is gained from the current model.