Supplemental Information for "Emergent Elasticity in the Neural Code for Space"

Code available at https://github.com/ganguli-lab/EmergentElasticityAnalysisAndSimulations

OVERVIEW OF SUPPLEMENTAL MATERIALS

In this supplemental information we first provide a reference for all mathematical symbols used (Sec. I). In Sec. II we present, in full mathematical detail, a general family of models that combines attractor networks, velocity-conjunctive cells, and plastic error correcting landmark cells, that can all work together to form maps of space in one dimension. In Sec. III, we then perform model reduction on this model to derive simplified low dimensional equations which capture the essential combined neural and synaptic dynamics of the full model. By linearizing, we derive an even simpler model in Sec. IV, which can be applied towards understanding how a simple geometry is learned through exploration in Sec. V.

In Sec. VI we provide a simple specific example of a neural model which yields analytic formulas for the effective reduced dynamics. The reader is encouraged to refer to this section to build intuition.

In Sec. VII, we extend the above mathematical formalism to two-dimensional attractor models yielding grid cells in two dimensions, and show how linearization of the exploration process yields a mechanical "particles-on-springs" model in Sec. VIII. We outline several connections to experiments in Sec. IX, and then discuss several extensions and lemmas used for the theory in Sec. X, as well as explain details of simulations and data analysis in Sec. XII.

CONTENTS

www.pnas.org/cgi/doi/10.1073/pnas.1805959115

I. TABLES OF SYMBOLS

A. Table of all symbols used in main paper

TABLE I: Table of symbols used in main paper.

Variable Name	**Symbol**	**Type**	**Units**
Exploration/dynamics time scale	t	Scalar	Time
Neural ring position	u	Angle (defined modulo 2π)	Neur. sheet length (Dimensionless)
Synaptic activation of attractor cells	$s(u)$	Scalar function of neural ring position	Firing rate
Pairwise cellular interactions	$\mathrm{J}(u-u')$	Scalar function of neural ring position	1/(Neur. sheet length× Time)
Leak time	τ_m	Scalar	Time
Firing nonlinearity	$\mathcal{G}$	Scalar function	Firing rate/Time
Attractor state	$\phi^{\mathrm{A}}(t)$	Angle (also neural ring position) function of time	Dimensionless
Steady bump pattern	$s^{*}(u-\phi^{\mathrm{A}})$	Scalar function of neural ring position	Firing rate
One-dimensional animal position	$x(t)$	Scalar function of time	Physical length
One-dimensional running velocity	$v(t)$	Scalar function of time	Physical length/Time
1D path integration constant	k	Scalar	Angle/Physical length
Landmark cell index	i	Integer label	Dimensionless
Landmark cell firing rate	$s_i^{\mathrm{L}}(t)$	Scalar function of time	Firing rate
Landmark cell firing field	$\mathrm{H}_i(x)$	Scalar function of animal position	Firing rate
Landmark cell synaptic weights (neural ring basis)	$\mathrm{W}_i(u)$	Scalar function of neural ring position	1/Neur. sheet length
Weight component in attract. basis	ϕ^{L}	Angle (defined modulo 2π)	Dimensionless
Attractor force law	$\mathcal{F}(\phi^{\mathrm{P}}-\phi^{\mathrm{A}})$	Function of difference of neural ring position	Dimensionless
Landmark strength	ω	Scalar	1/Time
Training time	T	Scalar	Training sessions
Linearized landmark pinning phase	θ_i^{L}	Unrolled angle (not defined modulo 2π)	Dimensionless
Box width	L	Scalar	Physical length
Width of wall cues	$\mathrm{L}_{\mathrm{Wall}}$	Scalar	Physical length
Box traversal time	τ	Scalar	Time
Animal running speed	v_0	Positive scalar (speed not velocity)	Physical length/Time
Magnitude of effect of east landmark on west landmark (spring constant)	M_{WE}	Scalar	1/Training session
Path integration amount between west wall and east wall	$\Delta\mathcal{X}^{\mathrm{A}}_{\mathrm{W}\to\mathrm{E}}$	Scalar	Physical length
Position self-estimate	$\mathcal{X}^{\mathrm{A}}[x(t),t]$	Scalar functional of path history and time	Physical length
Landmark position estimate	$\mathcal{X}^{\mathrm{L}}_{\mathrm{E}},\ \mathcal{X}^{\mathrm{L}}_{\mathrm{W}}$	Scalar	Physical length
Landmark pinning phase	$\theta^{\mathrm{L}}_{\mathrm{E}},\ \theta^{\mathrm{L}}_{\mathrm{W}}$	Unrolled phase	Dimensionless
2D neural sheet position	$\mathbf{u}$	Position on periodic rhombus, defined mod. $(0,2\pi),(\sqrt{3}\pi,\pi)$	Neur. sheet length
Pairwise cellular interactions	$\mathbf{J}(\lvert\mathbf{u}-\mathbf{u}'\rvert)$	Function of neural sheet distance	(Neur. sheet length)$^{-2}$/ Time
Synaptic activation	$s(\mathbf{u})$	Scalar function of 2D neural sheet position	Firing rate
Steady bump pattern	$s^{*}(\mathbf{u}-\boldsymbol{\phi}^{\mathbf{A}})$	Scalar function of 2D neural sheet position	Firing rate
2D attractor state (location of firing bump)	$\boldsymbol{\phi}^{\mathbf{A}}$	Position on periodic rhombus, defined mod. $(0,2\pi),(\sqrt{3}\pi,\pi)$	Neur. sheet length (Dimensionless)
2D attractor force law	$\boldsymbol{\mathcal{F}}(\boldsymbol{\phi}^{\mathbf{P}}-\boldsymbol{\phi}^{\mathbf{A}})$	2D vector function of neural sheet separation	Dimensionless
2D running velocity	$\mathbf{v}(t)$	2D vector function of time	Physical length/Time
2D path integration constant	$\mathbf{K}$	2×2 matrix (2D animal velocity → 2D neural sheet velocity)	Angle/Physical length
2D position	$\mathbf{r}(t)$	2D vector	Physical length
Synaptic weights (neural sheet basis)	$\mathrm{W}_i(\mathbf{u})$	Scalar function of neural sheet position	(Neur. sheet length)$^{-2}$
Synaptic weights (attractor basis)	$\tilde{\mathrm{W}}_i(\boldsymbol{\phi}^{\mathbf{L}})$	Scalar function of position on periodic rhombus	(Neur. sheet length)$^{-2}$
Synaptic weight component in attractor basis	$\boldsymbol{\phi}^{\mathbf{L}}$	Position on periodic rhombus, defined mod. $(0,2\pi),(\sqrt{3}\pi,\pi)$	Neur. sheet length (Dimensionless)
2D position self-estimate	$\boldsymbol{\mathcal{R}}^{\mathbf{A}}$	2D vector	Physical length
2D landmark position estimate	$\boldsymbol{\mathcal{R}}^{\mathbf{L}}_i$	2D vector	Physical length
Magnitude of effect of landmark j on landmark i	M_{ij}	Scalar	1/Training session
Mean path integration between landmark i and landmark j	$\Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{j\to i}$	2D vector	Physical length

1. Table of all symbols used in SI

A. Table for one-dimensional model and its reduction

TABLE II: Table of Symbols for One-Dimensional Model(Sec. II) and its Reduction(Sec. III, Sec. IV)

Variable Name	**Symbol**	**Type**	**Units**
Exploration/Neural dynamics time scale	t	Scalar	Seconds
Neural ring position	u	Angle (defined modulo 2π)	Neur. sheet length (Dimensionless)
Synaptic activation of non-conjunctive attractor cells	$s(u)$	Scalar function of neural ring position	Firing rate
Pairwise cellular interactions	$\mathrm{J}(u-u')$	Scalar function of neural ring position	1/(Neur. sheet length× Time)
Leak time	τ_m	Scalar	Time
Firing nonlinearity	$\mathcal{G}$	One-variable function	Firing rate/Time
Attractor dynamics	$\mathcal{D}_{\mathrm{A}}[s]$	Functional of synaptic activation; both inputs and outputs a function of neural ring position	Firing rate/Time
Attractor state (location of firing bump)	$\phi^{\mathrm{A}}(t)$	Angle (defined modulo 2π) function of time	Neur. sheet length (Dimensionless)
Steady bump pattern	$s^{*}(u-\phi^{\mathrm{A}})$	Scalar function of neural ring position	Firing rate
Jacobian of dynamics around attractor state ϕ^{A}	$\mathrm{Jac}_{\phi^{\mathrm{A}}}(u,u')$	Matrix of scalars, indexed by pairs of neural sheet positions (u,u')	(Neur. sheet length)$^{-2}$ $\times$Time^{-1}
Generic perturbation to attractor state	$\Delta s(u)$	Scalar function of neural ring position	Firing rate
Generic perturbation strength	ϵ	Scalar	1/Time
Generic perturbation to network dynamics	$\epsilon\ \delta_s(u-\phi^{\mathrm{P}})$	Scalar function of neural ring position	Firing rate/Time
Path integration perturbation strength	ϵ_{PI}	Scalar	1/Time
Offset of outgoing connections of velocity-conjunctive cells	$\Delta\phi_{\mathrm{PI}}$	Angle (defined modulo 2π)	Dimensionless
Synaptic activation of velocity-conjunctive attractor cells	$s_{\mathrm{EC}}(u)$, $s_{\mathrm{WC}}(u)$	Scalar function of neural ring position	Firing rate
Conjunctive characteristic speed	v_{C0}	Positive scalar (speed not velocity)	Physical length/Time
Landmark cell perturbation strength	ϵ_{LM}	Scalar	Neur. sheet length× (Firing rate) /Time
Shift in outgoing connections of velocity conjunctive cells	$\Delta\phi_{\mathrm{PI}}$	Difference in neural ring position	Neur. sheet length
Center of perturbation to network	ϕ^{P}	Angle (defined modulo 2π)	Neur. sheet length
One-dimensional position	$x(t)$	Scalar function of time	Physical length
Attractor force law	$\mathcal{F}(\phi^{\mathrm{P}}-\phi^{\mathrm{A}})$	Function of difference of neural ring position	Dimensionless
One-dimensional running velocity	$v(t)$	Scalar function of time	Physical length/Time
One-dimensional animal position	$x(t)$	Scalar function of time	Physical length
1D path integration constant	k	Scalar	Angle/Physical length
Landmark cell index	i	Integer label	Dimensionless
Landmark cell firing rate	$s_i^{\mathrm{L}}(t)$	Scalar function of time	Firing rate
Landmark cell firing field	$\mathrm{H}_i(x)$	Scalar function of animal position	Firing rate
Synaptic weights (neural ring basis)	$\mathrm{W}_i(u)$	Scalar function of neural ring position	1/Neur. sheet length
Synaptic weights (attractor basis)	$\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})$	Scalar function of angle	1/Neur. sheet length
Synaptic weight component in attractor basis	ϕ^{L}	Angle (defined modulo 2π)	Dimensionless
Training time	T	Scalar	Training sessions
Landmark strength	ω	Scalar	1/Time
Linearized landmark pinning phase	θ_i^{L}	Unrolled angle (not defined modulo 2π)	Dimensionless

B. *Table for simplest environment case*

TABLE III: Table of Symbols for Simplest Environment Case (Sec. V)

Variable Name	**Symbol**	**Type**	**Units**
Box width	L	Scalar	Physical length
Width of wall cues	$\mathrm{L}_{\mathrm{Wall}}$	Scalar	Physical length
Width of cue-depleted zone	$\mathrm{L}_{\mathrm{Int}}$	Scalar	Physical length
Box traversal time	τ	Scalar	Time
Animal running speed	v_0	Positive scalar (speed not velocity)	Physical length/Time
Magnitude of effect of east landmark on west landmark	M_{WE}	Scalar	1/Training session
Path integration amount between west wall and east wall	$\Delta\mathcal{X}^{\mathrm{A}}_{\mathrm{W}\to\mathrm{E}}$	Scalar	Physical length
Position self-estimate	$\mathcal{X}^{\mathrm{A}}[x(t),t]$	Scalar functional of path history and time	Physical length
Landmark pinning phase	θ^{L}_i	Unrolled phase	Dimensionless
Landmark position estimate	$\mathcal{X}^{\mathrm{L}}_i$	Scalar	Physical length

C. *Table of units for mechanical framework*

TABLE IV: Table of units mechanical framework (Sec. VIII)

Variable Name	**Symbol**	**Type**	**Units**
Position estimate given a path history	$\boldsymbol{\mathcal{R}}^{\mathbf{A}}[\mathbf{r}(t),t]$	2D vector functional of path history and time	Physical length
Mean position self-estimate at a given position	$\bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}(\mathbf{r})$	2D vector function of animal position	Physical length
Mean position self-estimate across landmark field i	$\bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}_i$	2D vector	Physical length
Total landmark strength at a given position	$\omega(\mathbf{r})$	Scalar function of animal position	1/Time
Effect of landmark forcing at $\mathbf{r}'$ on position self estimate at $\mathbf{r}$	$\mathrm{S}(\mathbf{r}_{\mathrm{B}},\mathbf{r}_{\mathrm{A}})$	Scalar function of animal position pairs (depends on path statistics)	1/(Physical length^2)
Reverse animal trajectory	$\mathbf{r}_{\mathrm{rev}}(t)$	2D vector function of time	Physical length
Effect of landmark forcing at time t' on self-estimate at time t	$\mathrm{F}[\mathbf{r}(t),t,t']$	2D vector functional of time pairs and path history	1/Time
Degree of memory of input from time t' at time t	$\mathrm{Mem}[\mathbf{r}(t),t,t']$	2D vector functional of time pairs and path history	Dimensionless

D. *Table of symbols for two-dimensional model, reduction, linearization, and mechanical proof*

TABLE V: Table of symbols for two-dimensional model (Sec. VII)

Variable Name	**Symbol**	**Type**	**Units**
2D neural sheet position	$\mathbf{u}$	Position on periodic rhombus, defined modulo $(0, 2\pi), (\sqrt{3}\pi, \pi)$	Neur. sheet length (Dimensionless)
Firing nonlinearity	$\mathcal{G}$	Scalar function	Firing rate/Time
Pairwise cellular interactions	$\mathbf{J}(\lvert\mathbf{u}-\mathbf{u}'\rvert)$	Function of neural sheet distance	(Neur. sheet length)$^{-2}$
Synaptic activation	$s(\mathbf{u})$	Scalar function of 2D neural sheet position	Firing rate
Steady bump pattern	$s^*(\mathbf{u}-\boldsymbol{\phi}^{\mathbf{A}})$	Scalar function of 2D neural sheet position	Firing rate
2D attractor state (location of firing bump)	$\boldsymbol{\phi}^{\mathbf{A}}$	Position on periodic rhombus, defined modulo $(0, 2\pi), (\sqrt{3}\pi, \pi)$(function of time)	Neur. sheet length (Dimensionless)
Attractor force law	$\boldsymbol{\mathcal{F}}(\boldsymbol{\phi}^{\mathbf{P}}-\boldsymbol{\phi}^{\mathbf{A}})$	2D vector function of neural sheet separation	Dimensionless
Attractor dynamics	$\mathcal{D}_{\mathrm{A}}[s(\mathbf{u})]$	Functional of synaptic activation; both inputs and outputs a function of neural sheet position	Time derivative of synaptic activation
2D running velocity	$\mathbf{v}(t)$	2D vector function of time	Physical length/Time
Two dimensional path integration constant	$\mathbf{K}$	2 × 2 Matrix (transforms 2D animal velocity to 2D attractor state velocity)	Angle/Physical length
2D animal position	$\mathbf{r}(t)$	2D vector function of time	Physical length
Landmark cell index	i	Integer label	Dimensionless
Landmark cell firing rate	$s_i^{\mathbf{L}}(t)$	Scalar	Firing rate
Landmark cell firing field	$\mathrm{H}_i(\mathbf{r})$	Scalar function of 2D physical position	Firing rate
Synaptic weights (neural sheet basis)	$\mathrm{W}_i(\mathbf{u})$	Scalar function of neural sheet position	(Neur. sheet length)$^{-2}$
Synaptic weights (attractor basis)	$\tilde{\mathrm{W}}_i(\boldsymbol{\phi}^{\mathbf{L}})$	Scalar function of position on periodic rhombus	(Neur. sheet length)$^{-2}$
Synaptic weight component in attractor basis	$\boldsymbol{\phi}^{\mathbf{L}}$	Position on periodic rhombus, defined modulo $(0, 2\pi), (\sqrt{3}\pi, \pi)$	Neur. sheet length (Dimensionless)
Landmark pinning phase	$\boldsymbol{\theta}_i^{\mathbf{L}}$	Unrolled 2D phase	Dimensionless
Training time	T	Scalar	Training sessions
Landmark strength	ω	Scalar	1/Time
Position self-estimate	$\boldsymbol{\mathcal{R}}^{\mathbf{A}}$	2D vector	Physical length
Landmark position estimate	$\boldsymbol{\mathcal{R}}_i^{\mathbf{L}}$	2D vector	Physical length
Average landmark position estimate at a position	$\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r})$	2D vector function of physical position	Physical length
Total landmark strength at a position	$\omega(\mathbf{r})$	Scalar function of physical position	1/Time
Magnitude of effect of landmark j on landmark i	M_{ij}	Scalar	1/(Training session)
Mean path integration between landmark i and landmark j	$\Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{j\to i}$	2D vector	Physical length

E. Table of units for experiments and simulations

TABLE VI: Table of symbols for experiments and simulations (Sec. XII)

Variable Name	**Symbol**	**Type**	**Units**
Head direction unit vector	$\hat{\mathbf{HD}}(t)$	2D unit vector (function of time)	Dimensionless
Experimentally observed animal position (center of diodes)	$\mathbf{r}(t)$	2D vector	Dimensionless
Path condition	$\mathcal{C}$	Boolean functional of time and path history $\mathbf{r}(t)$ (path condition is satisfied or not satisfied at a certain time given the path history)	Dimensionless
Grid cell label	GC	Integer label	Dimensionless
Variational step distance for cross-correlation	$\Delta\mathbf{r}_{\mathbf{C}}$	2D vector	Physical length
Conditional firing rate of cell GC given path condition $\mathcal{C}$	$s_{\mathrm{GC}}^{\mathcal{C}}(\mathbf{r})$	Scalar function of 2D animal position	Physical length
Cross-correlation between two path conditions $\mathcal{C}1$, $\mathcal{C}2$ for grid cell GC	$\mathbf{C}_{\mathrm{GC}}^{\mathcal{C}1\mathcal{C}2}(\Delta\mathbf{r}_{\mathbf{C}})$	Scalar function of 2D difference in animal position	Dimensionless
Firing field center	$\mathbf{r}_{\mathbf{ff}}$	2D animal position	Physical length
Firing field	$\mathbf{ff}$	Set of 2D animal positions	Physical length
Spike shift for path condition $\mathcal{C}$, grid cell GC, and firing field $\mathbf{ff}$	$\mathbf{S}_{\mathcal{C},\mathrm{GC},\mathbf{ff}}$	2D difference vector	Physical length
Spike position	$\mathbf{r}_{\mathrm{Spk}}$	2D vector	Physical length

II. FULL HIGH-DIMENSIONAL MODEL OF DYNAMICS AND LEARNING FOR PATH INTEGRATION AND ENVIRONMENTAL EXPLORATION

We first consider a one-dimensional attractor network consisting of a large population of neurons whose connectivity is determined by their position on an abstract ring, as in Fig. SI.1. For analytical simplicity, we take a neural field approach [1], so that position on the ring is described by a continuous coordinate u, with the activation, or firing rate of a neuron at each position u given by $s(u)$ (see list of symbols and units in Table II).

In the interest of generality, we do not commit to any particular functional form for the neural field model presented in this section. To help build intuition, the reader is encouraged to refer to Sec. VI, where a simple, exactly solvable model is presented.

1. Cell Types

Our model has three types of cells.

A. Non-conjunctive attractor network cells

Non-conjunctive cells that form an attractor network are uniformly distributed on a neural ring, and have a firing rate $s(u)$, where u is the position on the neural ring. Non-conjunctive attractor network cells have outgoing connections to each other with a weight of $\mathrm{J}(u' - u)$, yielding some set of attractor dynamics. These additionally receive inputs from conjunctive attractor network cells and landmark cells. In general, to yield a steady bump pattern, J will have local excitation and mid-range inhibition (Fig. SI.1A).

B. Conjunctive-velocity sensitive cells

The model also has a ring of east-conjunctive (EC) and a ring of west-conjunctive (WC) cells. The firing rates of these cells at a position u on their respective rings depend instantaneously on the firing rates $s(u)$ along the non-conjunctive attractor ring and the animal running velocity through the functions

$$s_{\mathrm{EC}}(u) = \left(\frac{v_{\mathrm{East}}}{v_{\mathrm{C0}}}\right) s(u), \qquad s_{\mathrm{WC}}(u) = \left(\frac{v_{\mathrm{West}}}{v_{\mathrm{C0}}}\right) s(u). \tag{SI.II.1}$$

Here $v_{\mathrm{East}},\ v_{\mathrm{West}}$ correspond to the east and west components of the animal velocity v:

$$v_{\mathrm{East}} = \begin{cases} v & v > 0 \\ 0 & v \leq 0 \end{cases}, \qquad v_{\mathrm{West}} = \begin{cases} 0 & v \geq 0 \\ -v & v < 0 \end{cases}, \qquad v_{\mathrm{East}} - v_{\mathrm{West}} = v,$$

and v_{C0} is the characteristic speed at which conjunctive cell firing rates equal the non-conjunctive attractor network cell firing rates.

The east-conjunctive cells at u have outgoing weights onto the attractor ring so that the peak synaptic output is biased, and centered at a point $u+\Delta\phi_{\mathrm{PI}}$, where $\Delta\phi_{\mathrm{PI}}$ is the bias in the direction of outgoing weights. West-conjunctive cells have outgoing weights with opposite bias to $u - \Delta\phi_{\mathrm{PI}}$. In general, weights from a conjunctive cell at u will have an outgoing synaptic strength profile that peaks at $u + \Delta\phi_{\mathrm{PI}}$ for east-conjunctive cells (Fig. SI.1B) and $u - \Delta\phi_{\mathrm{PI}}$ for west-conjunctive cells (Fig. SI.1C). Here, for simplicity, we have each east (west) conjunctive cell at u connect to a *single* non-conjunctive attractor cell at $u \pm \Delta\phi_{\mathrm{PI}}$. See Sec. X 1 for a more realistic case in which a conjunctive cell connects to multiple non-conjunctive attractor network cells.

C. Landmark cells

Landmark cells do not live on a neural ring like the attractor cells and the velocity-conjunctive cells. Instead, each landmark cell i has a position-dependent normalized firing rate of $s_i^{\mathrm{L}}(t) = \mathrm{H}_i(x(t))$, where $\mathrm{H}_i(x)$ is some function of immediate animal position. They do, however, have plastic synaptic connections onto every non-conjunctive cell in the attractor network with neural ring position-dependent strength $\mathrm{W}_i(u)$ (Fig. SI.1D).

2. Attractor dynamics

Over the short exploration time scale t, the equations (extended version of Eq. 1(Main)) for neural dynamics [2] are (Fig. SI.1):

$$\frac{ds\,(u)}{dt} = -\frac{s(u)}{\tau_m} + \underbrace{\mathcal{G}\left(\int_{u'} \mathrm{J}(u-u')s(u')\right)}_{\text{Non-conjunctive (Fig. SI.1A)}} + \underbrace{\epsilon_{\mathrm{PI}}\left(s_{\mathrm{EC}}\left(u-\Delta\phi_{\mathrm{PI}}\right)\right)}_{\text{East cells (Fig. SI.1B)}} + \underbrace{\epsilon_{\mathrm{PI}} s_{\mathrm{WC}}\left(u+\Delta\phi_{\mathrm{PI}}\right)}_{\text{West cells (Fig. SI.1C)}} + \underbrace{\epsilon_{\mathrm{LM}}\sum_i \left(\mathrm{W}_i(u)s_i^{\mathrm{L}}(t)\right)}_{\text{Landmark Cell Inputs (Fig. SI.1D)}} \tag{SI.II.2}$$

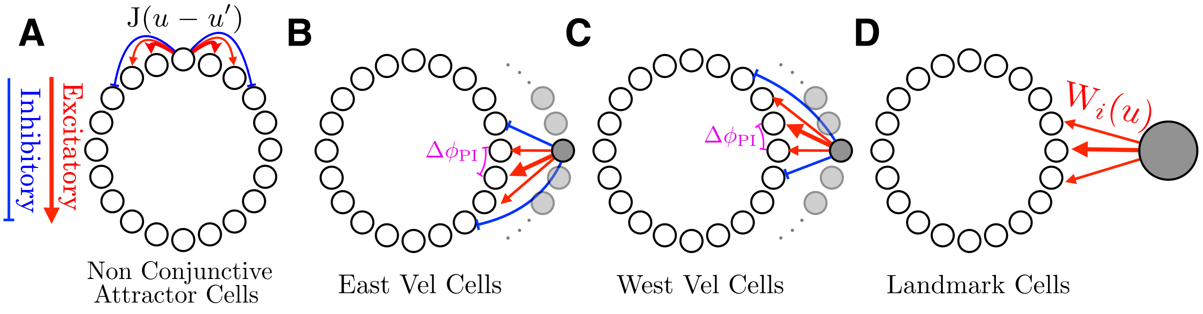
A
J$(u-u')$
Excitatory
Inhibitory
Non Conjunctive Attractor Cells
B
$\Delta\phi_{\mathrm{PI}}$
East Vel Cells
C
$\Delta\phi_{\mathrm{PI}}$
West Vel Cells
D
$W_i(u)$
Landmark Cells

FIG. SI.1:

A) Schematic of a ring attractor with short-range excitation (red arrows) and longer range inhibition (blue arrows). **B)** East-conjunctive cells with clockwise biased outgoing connections ($u \to u + \Delta\phi_{\mathrm{PI}}$) **C)** West-conjunctive cells with counterclockwise biased outgoing connections ($u \to u - \Delta\phi_{\mathrm{PI}}$). While in Eq. SI.II.2, we assume a simple uniform offset for velocity-conjunctive cells, this constraint can be generalized to more realistic outgoing connections as well (Sec. X 1). **D)** Landmark cells. The landmark cell doesn't live on the neural sheet, but has outgoing connections to attractor neurons with a strength of $\mathrm{W}_i(u)$.

Here $\mathrm{J}(u-u')$ defines the synaptic weight from a cell at position u' to a cell at u, τ_m is the "leak time" and $\mathcal{G}$ is a nonlinearity (See Sec. X 1 for how variants of these dynamics can be solved.) The firing rates of east-west conjunctive cells are given by the non-conjunctive firing rates and the animal velocity [3]:

$$s_{\mathrm{EC}}(u) = \left(\frac{v_{\mathrm{East}}}{v_{\mathrm{C0}}}\right) s(u), \qquad s_{\mathrm{WC}}(u) = \left(\frac{v_{\mathrm{West}}}{v_{\mathrm{C0}}}\right) s(u),$$

and the firing rates of the landmark cells are:

$$s_i^{\mathrm{L}}(t) = \mathrm{H}_i(x(t)).$$

3. Plasticity dynamics

The long term learning (Eq. 5(Main)) is mediated by the updates of the Hebbian weights $\mathrm{W}_i(u)$ from the landmark cells to the attractor network:

$$\frac{d\mathrm{W}_i(u)}{d\mathrm{T}} = \langle s(u)|i \text{ Firing}\rangle - \mathrm{W}_i(u) = \frac{\int_t s(u,t)s_i^{\mathrm{L}}(t)}{\int_t s_i^{\mathrm{L}}(t)} - \mathrm{W}_i(u) \tag{SI.II.3}$$

There is a separation of timescales between the navigation dynamics (Eq. SI.II.2) and the learning dynamics (Eq. SI.II.3). The integral $\int_t$ represents the average *within* a single training session [4] (assumed to be much longer than the time it takes to traverse the environment), while $d/d\mathrm{T}$ represents plasticity *across* training sessions [5], or at least over several traversals through the environment. We note that the equations for learning (Eq. SI.II.3) and dynamics (Eq. SI.II.2) implicitly depend on (1) the environmental geometry, (2) the landmark firing fields $\mathrm{H}(x)$, and (3) the distribution of animal trajectories $x(t)$. We will see below how these three ingredients interact to determine learned circuit outcomes.

III. MODEL REDUCTION OF HIGH-DIMENSIONAL NEURAL DYNAMICS TO A REDUCED PHASE DYNAMICS

In this section (List of symbols and units in Table II) we will first show how the state of the ring attractor can be mapped onto a single periodic variable (Sec. III 1) representing the peak of the bump pattern. Then, using this low-dimensional representation, we develop a framework to understand the effect of perturbations on the low-dimensional attractor state (Sec. III 2), and then use this framework to understand the effect of path integration (Sec. III 3) and landmarks (Sec. III 4) on this state. Finally, in Sec. III 5, we also map the Hebbian learning rule to this reduced representation. In Sec. VI, we present an exactly solvable model which reduces to analytically solvable functions for effective dynamics.

1. Reducing the ring attractor network state to a single phase variable

We will refer to the non-velocity, non-landmark dynamics as $\mathcal{D}_{\mathrm{A}}[s]$:

$$\mathcal{D}_{\mathrm{A}}[s] = -\frac{s(u)}{\tau_m} + \mathcal{G}\left(\int_{u'} \mathrm{J}(u' - u')s(u')\right).$$

When there is no external sensory input, Eq. SI.II.2 reduces to:

$$\frac{ds\,(u)}{dt} = \mathcal{D}_{\mathrm{A}}[s].$$

How can we characterize these dynamics?

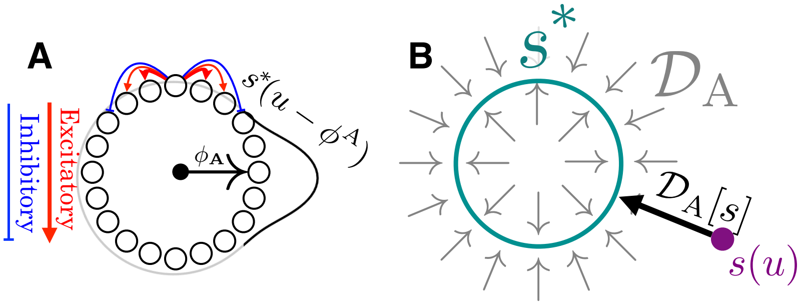

FIG. SI.2:

A) The ring attractor dynamics $ds(u)/dt = \mathcal{D}_{\mathrm{A}}[s]$ yield a 1D family of bump-attractor states $s^*(u - \phi^{\mathrm{A}})$, which are mapped onto a single periodic variable ϕ^{A} representing the peak of the bump pattern. B) Manifold schematic of attractor dynamics. In the state space of $s(u)$, there exists a one-dimensional manifold of stable attractor dynamics s^* (Teal circle). A state $s(u)$ not on the manifold will eventually be pulled towards some steady state $s^*(u - \phi^{\mathrm{A}})$ on the manifold.

Many appropriate choices of J and $\mathcal{G}$, corresponding for example to short range excitation and long range inhibition, will yield a stable, or steady state, localized bump activity pattern [6, 7]. We assume we have chosen J to be an even function whose shape admits such stable bump solutions. In Sec. VI we describe a specific choice of J that leads to an exactly solvable model. Because the dynamics are translation-invariant, every translation of a steady bump pattern is *also* a steady bump pattern. We call these stable bump patterns patterns $s^*(u - \phi^{\mathrm{A}})$, parameterized by the position of their peak firing ϕ^{A} [6, 7].

This one-dimensional family of stable bump activity patterns can itself be thought of as ring of stable firing patterns in the space of all possible firing patterns. Just as u indexes a family of neurons on the neural sheet, the coordinate ϕ^{A} indexes the different stable neural activity patterns, with a particular value of ϕ^{A} corresponding to a stable bump on the neural ring centered at coordinate $u = \phi^{\mathrm{A}}$. For simplicity we set units such that the coordinate u along the neural ring, and the coordinate ϕ^{A} along the ring of stable attractor patterns are both angles, defined modulo 2π. Thus u and ϕ^{A} are phase variables denoting position along the neural ring and ring of bump attractor patterns respectively.

2. Effects of perturbations on the reduced attractor state

The dynamics of Eq. SI.II.2 are:

$$\frac{ds\,(u)}{dt} = \mathcal{D}_{\mathrm{A}}[s] + \epsilon_{\mathrm{PI}}\underbrace{\left(s_{\mathrm{EC}}\,(u - \Delta\phi_{\mathrm{PI}})\right)}_{\text{Input from east cells}} + \epsilon_{\mathrm{PI}}\underbrace{s_{\mathrm{WC}}\,(u + \Delta\phi_{\mathrm{PI}})}_{\text{Input from west cells}} + \epsilon_{\mathrm{LM}}\sum\nolimits_i \underbrace{\left(\mathrm{W}_i(u)s_i^{\mathrm{L}}(t)\right)}_{\text{Landmark cell inputs}} \;,$$

where s_{EC}, s_{EC} are defined in Eq. SI.II.1. While we have characterized the stable fixed points of $\dot{s} = \mathcal{D}_{\mathrm{A}}[s]$ in the absence of landmark and self-motion cues, we have not yet shown how the attractor network responds to these extra inputs. Assuming that the intrinsic dynamics are much stronger than the inputs applied from landmark and conjunctive cells, we can treat these inputs as small perturbations to the intrinsic dynamics. We can then describe how these small input perturbations cause the attractor bump to move around, without changing shape. To do so, we first derive a reduced description for how a general weak external feedforward input to the attractor network modifies its dynamics.

A. *Attractor dynamics under small perturbations*

Here, we examine how an attractor network with a steady-state firing pattern $s^{*}(u-\phi^{\mathrm{A}})$ responds when its dynamics are perturbed by some small additional input $\epsilon\delta_s(u-\phi^{\mathrm{P}})$, which is centered at ϕ^{P}. This perturbed dynamics is given by

$$\frac{ds(u)}{dt} = \mathcal{D}_{\mathrm{A}}[s] + \epsilon\delta_s(u-\phi^{\mathrm{P}}). \tag{SI.III.4}$$

In order to understand the effect of this small perturbation, we need to linearize the dependence of dynamics on synaptic activation. The linearization of dynamics around an *arbitrary* state $s_0(u)$ with a generic perturbation to the attractor state $\Delta s(u)$ is defined by the functional derivative:

$$\mathcal{D}_{\mathrm{A}}[s_0 + \Delta s] \approx \mathcal{D}_{\mathrm{A}}[s_0] + \left(\int_u \frac{\delta\mathcal{D}_{\mathrm{A}}}{\delta s\,(u)}\Delta s(u)\right)\Bigg|_{s=s_0} \tag{SI.III.5}$$

We can write Eq. SI.III.5 in matrix notation using the Jacobian Jac_{s_0} around s_0:

$$\mathcal{D}_{\mathrm{A}}[s_0 + \Delta s] \approx \mathcal{D}_{\mathrm{A}}[s_0] + \mathrm{Jac}_{s_0}\cdot\Delta s, \qquad \text{where} \qquad \mathrm{Jac}_{s_0} = \frac{\delta\mathcal{D}_{\mathrm{A}}}{\delta s\,(u)}\Bigg|_{s=s_0} \tag{SI.III.6}$$

Because $s(u)$ will always be close to a steady state under a small perturbation, we *specifically* need to understand $\mathrm{Jac}_{\phi^{\mathrm{A}}}$, the Jacobian of these dynamics around the point $s^{*}(u-\phi^{\mathrm{A}})$ [8].

B. *Linearized response of the attractor network to perturbations*

Because

$$\mathcal{D}_{\mathrm{A}}[s^{*}(u-\phi^{\mathrm{A}})] = 0 \text{ For all } \phi^{\mathrm{A}}, \tag{SI.III.7}$$

Eq. SI.III.4 reduces to:

$$\frac{ds}{dt} = \mathcal{D}_{\mathrm{A}}\left[s_{\phi^{\mathrm{A}}} + \Delta s\right] + \epsilon\delta_s \approx \overbrace{\cancel{\mathcal{D}_{\mathrm{A}}\left[s_{\phi^{\mathrm{A}}}\right]}}^{0} + \mathrm{Jac}_{\phi^{\mathrm{A}}} \cdot \Delta s + \epsilon\delta_s.$$

Likewise, Eq. SI.III.7 tells us that the sliding mode $s^{*\prime}(u - \phi^{\mathrm{A}})$ (the spatial derivative of $s^*\left(u - \phi^{\mathrm{A}}\right)$) is a zero-eigenvector of the Jacobian $\mathrm{Jac}_{\phi^{\mathrm{A}}}$ [9]. This can be seen through the definition of the Jacobian:

$$\mathrm{Jac}_{\phi^{\mathrm{A}}} \cdot \left[s^{*\prime}\left(u - \phi^{\mathrm{A}}\right)\right] = -\left(\left.\frac{\delta\mathcal{D}_{\mathrm{A}}[s\left(u\right)]}{\delta s\left(u\right)}\right|_{s(u)=s^*(u-\phi^{\mathrm{A}})}\right) \cdot \left(\frac{ds^*(u-\phi^{\mathrm{A}})}{d\phi^{\mathrm{A}}}\right) = -\frac{d}{d\phi^{\mathrm{A}}}\overbrace{\cancel{\left(\mathcal{D}_{\mathrm{A}}\left[s^*\left(u-\phi^{\mathrm{A}}\right)\right]\right)}}^{0 \text{ For all } \phi^{\mathrm{A}}} = 0.$$

Moreover, because $s^*(u - \phi^{\mathrm{A}})$ is a stable one-dimensional family of solutions of $\mathcal{D}_{\mathrm{A}}$, $\mathrm{Jac}_{\phi^{\mathrm{A}}}$ must be a negative semidefinite matrix, where the sliding mode $s^{*\prime}(u - \phi^{\mathrm{A}})$ is the *only* eigenvector of $\mathrm{Jac}_{\phi^{\mathrm{A}}}$ with a non-negative eigenvalue [10].

C. *Mathematical background: response of linear dynamical systems to perturbations*

How will a dynamical system, with a negative semidefinite Jacobian having a *single* zero mode, respond to a weak perturbation? To answer this question, we first review some mathematical background. Consider a simple case of a two-dimensional line attractor with a state $\mathbf{s} = (s_0, s_1)$. The dynamics $d\mathbf{s}/dt = \mathcal{D}_{\mathrm{A}}(\mathbf{s})$ are:

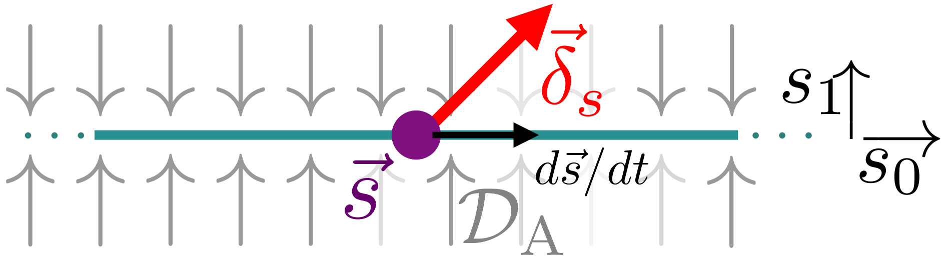

FIG. SI.3:

Schematic of perturbed line attractor dynamics (Eq. SI.III.9). Any applied perturbation will the projected onto the $\hat{\mathbf{s}}_0$ direction, which is the zero-eigenvector of the Jacobian.

$$\frac{d}{dt}\begin{pmatrix} s_0 \\ s_1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -\omega_{\mathcal{D}} \end{pmatrix} \cdot \begin{pmatrix} s_0 \\ s_1 \end{pmatrix}.$$

Adding a perturbation of $\boldsymbol{\delta}_{\mathbf{s}} = (\delta_{s0}, \delta_{s1})$ yields the dynamics:

$$\frac{d}{dt}\begin{pmatrix} s_0 \\ s_1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -\omega_{\mathcal{D}} \end{pmatrix} \cdot \begin{pmatrix} s_0 \\ s_1 \end{pmatrix} + \begin{pmatrix} \delta_{s0} \\ \delta_{s1} \end{pmatrix} = \begin{pmatrix} \delta_{s0} \\ \delta_{s1} - \omega_{\mathcal{D}} s_1 \end{pmatrix}. \tag{SI.III.8}$$

Note that there are no interactions between the two modes. In the limit where the attractor dynamics are very strong ($\omega_{\mathcal{D}} \to \infty$), Eq. SI.III.8 reduces to

$$ds_0/dt = \delta_{s0}, \qquad s_1 = 0, \tag{SI.III.9}$$

i,e. the s_0 mode ("the sliding mode") is free to move, while the s_1 mode is anchored at zero.

In matrix notation, the reduction of Eq. SI.III.8 to Eq. SI.III.9 is the reduction of the full dynamics under a perturbation (Equivalent to Eq. SI.III.8):

$$d\mathbf{s}/dt = \mathrm{Jac} \cdot \mathbf{s} + \boldsymbol{\delta}_{\mathbf{s}} \tag{SI.III.10}$$

to the projection of the perturbation onto the sliding mode (Equivalent to Eq. SI.III.9):

$$d\mathbf{s}/dt = \underbrace{(\hat{\mathbf{s}}_0 \cdot \boldsymbol{\delta}_{\mathbf{s}})}_{\substack{\text{Proj. onto} \\ \text{Sliding Mode}}} \times \underbrace{\hat{\mathbf{s}}_0}_{\text{Sliding Mode}}, \tag{SI.III.11}$$

where $\hat{\mathbf{s}}_0 = (1, 0)$ is the zero-eigenvector of Jac [11].

The reasoning in Sec. III 2 D is an analogous but higher-dimensional generalization of Eq.SI.III.11, applied to functions and functional derivatives. In essence, in the higher dimensional case, an input-perturbation will be restored along all dimensions in firing rate space back to the manifold of attractor bump patterns, *except* in the single direction along the sliding mode. Since the sliding mode is a spatial derivative of the bump pattern, adding in the sliding mode will move the bump pattern along the manifold of attractor states. Moreover, the rate of motion will be proportional to the inner product of the profile of the external input perturbation as a function along the neural ring, and the spatial derivative of the current bump pattern. We now study this inner-product, or projection of the input perturbation onto the attractor network sliding mode to yield a force law for attractor bump motion.

D. *Projection of perturbations onto the sliding mode*

When an external perturbation is small and $\text{Jac}_{\phi^{\text{A}}}$ is symmetric at a given firing rate profile of $s(u)$, e.g.

$$\text{Jac}_{\phi^{\text{A}}}(u', u) = \frac{d\left(\mathcal{D}_{\text{A}}[s](u')\right)}{ds(u)} = \frac{d\left(\mathcal{D}_{\text{A}}[s](u)\right)}{ds(u')} = \text{Jac}_{\phi^{\text{A}}}(u, u'),$$

the eigenvectors of $\text{Jac}_{\phi^{\text{A}}}$ are orthogonal [12]. Therefore, given any small perturbation to the dynamics:

$$ds(u)/dt = \mathcal{D}_{\text{A}}[s] + \epsilon\delta_s(u - \phi^{\text{P}}), \tag{SI.III.12}$$

the effective perturbation will be the projection of the actual perturbation $\delta_s(u - \phi^{\text{P}})$ onto the sliding mode, i.e., the single zero-mode of the Jacobian:

$$\frac{ds(u)}{dt} \approx \epsilon \underbrace{\left[\frac{1}{\mathcal{N}}\int_u s^{*\prime}\left(u - \phi^{\text{A}}\right)\delta_s\left(u - \phi^{\text{P}}\right)\right]}_{\text{Projection onto Sliding Mode}} \underbrace{s^{*\prime}\left(u - \phi^{\text{A}}\right)}_{\text{Sliding Mode}} = \epsilon \underbrace{\left[\frac{1}{\mathcal{N}}\int_u s^{*\prime}(u)\delta_s\left(u - \left[\phi^{\text{P}} - \phi^{\text{A}}\right]\right)\right]}_{\text{Projection onto Sliding Mode}} \underbrace{s^{*\prime}\left(u - \phi^{\text{A}}\right)}_{\text{Sliding mode}}, \tag{SI.III.13}$$

where we have substituted $u \to u + \phi^{\text{A}}$ in the second step [13], and $\mathcal{N}$ is equal to the squared magnitude of the sliding mode [14] to ensure that the projection is properly normalized.

We can define the negative of projection onto the sliding mode to be an effective force law $\mathcal{F}\left(\phi^{\text{P}} - \phi^{\text{A}}\right)$, i.e.

$$\mathcal{F}\left(\phi^{\text{P}} - \phi^{\text{A}}\right) = -\frac{1}{\mathcal{N}}\int_u s^{*\prime}(u)\delta_s\left(u - \left[\phi^{\text{P}} - \phi^{\text{A}}\right]\right)$$

to get (Fig. SI.4):

$$\frac{ds(u)}{dt} = -\epsilon\mathcal{F}(\phi^{\text{P}} - \phi^{\text{A}})\ s^{*\prime}\left(u - \phi^{\text{A}}\right). \tag{SI.III.14}$$

Eq. SI.III.14 shows that at any given time, the *temporal* derivative $ds(u)/dt$ is a multiple of the *spatial* derivative $s^{*\prime}(u - \phi^{\text{A}})$. Therefore, the effect of any perturbation can be reduced to an effective force on the single-variable bump position:

$$\frac{d\phi^{\text{A}}}{dt} = \epsilon\mathcal{F}(\phi^{\text{P}} - \phi^{\text{A}}). \tag{SI.III.15}$$

We can verify that the effective force law of Eq. SI.III.15 yields Eq. SI.III.14:

$$\frac{ds^*\left(u - \phi^{\text{A}}\right)}{dt} = \frac{ds^*\left(u - \phi^{\text{A}}\right)}{d\phi^{\text{A}}}\frac{d\phi^{\text{A}}}{dt} = -s^{*\prime}\left(u - \phi^{\text{A}}\right)\frac{d\phi^{\text{A}}}{dt} = -\underbrace{\epsilon\mathcal{F}(\phi^{\text{P}} - \phi^{\text{A}})}_{d\phi^{\text{A}}/dt}\, s^{*\prime}\left(u - \phi^{\text{A}}\right), \tag{SI.III.16}$$

where we have used the fact that $ds^*(u - \phi^{\text{A}})/d\phi^{\text{A}} = -ds^*(u - \phi^{\text{A}})/du = -s^{*\prime}(u - \phi^{\text{A}})$.

a. Form of the force function When the perturbation takes the form of input from localized Hebbian landmark cells, the perturbation function is simply the attractor bump pattern $s^*(u-\phi^{\rm L})$. Therefore, defining $\Delta\phi = \phi^{\rm L} - \phi^{\rm A}$,

$$\mathcal{F}\left(\Delta\phi\right) = -\int_u s^{*\prime}(u)s^*(u-\Delta\phi) = -\int_u s^{*\prime}(u+\Delta\phi)s^*(u) = -\int_u \frac{d[s^*(u+\Delta\phi)]}{d\Delta\phi}s^*(u) = -\frac{d}{d\Delta\phi}\left[\int_u s^*(u+\Delta\phi)s^*(u)\right].$$

Therefore, the force function is simply the negative derivative of the spatial autocorrelation function of the bump pattern. Because the spatial autocorrelation is even and maximized at $\Delta\phi = 0$ (minimized at $\Delta\phi = \pi$), the force function will be odd, with positive (negative) values for positive (negative) $\Delta\phi$. As long as the bump size is not much smaller than the bump spacing, the autocorrelation will decrease gradually between $\Delta\phi = 0$, $\Delta\phi = \pi$, leading to a long range force function which only approaches zero at, and far from, the origin. This behavior is qualitatively matched by $\mathcal{F}\left(\Delta\phi\right) = \sin(\Delta\phi)$. For simplicity, we define the magnitude of $\mathcal{F}(\Delta\phi)$ to give it a slope of 1 at $\Delta\phi = 0$; all strength information can be absorbed in the factor ϵ it is multiplied by.

b. Dynamics with non-symmetric Jacobians. $\text{Jac}_{\phi^{\rm A}}$ will in general be non-symmetric. In this case, we may use the same techniques as before, except now we must use a non-orthogonal projection onto the sliding mode:

$$\frac{ds(u)}{dt} \approx \epsilon \underbrace{\left[\int_u \mathrm{v}_{\rm proj}\left(u-\phi^{\rm A}\right)\delta_s(u-\phi^{\rm P})\right]}_{\text{Non-orthogonal projection onto sliding mode}} \underbrace{s^{*\prime}\left(u-\phi^{\rm A}\right)}_{\text{Sliding Mode}} = -\epsilon\mathcal{F}(\phi^{\rm P}-\phi^{\rm A})s^{*\prime}\left(u-\phi^{\rm A}\right) \qquad \text{(SI.III.17)}$$

where $\mathrm{v}_{\rm proj}(u-\phi^{\rm A})$ can, in principle, be solved through diagonalization of the Jacobian $\text{Jac}_{\phi^{\rm A}}$.

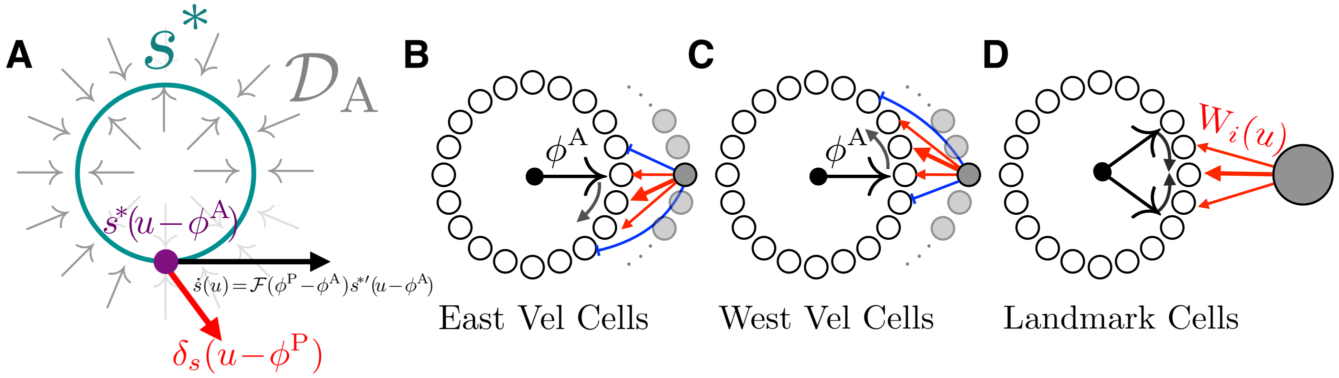

FIG. SI.4:

A) Schematic of how perturbation slides attractor state along manifold. There exists a one-dimensional manifold of steady attractor states s^*(teal circle), which is supported by the attractor dynamics $\mathcal{D}_{\rm A}$(Gray arrows). Any perturbation in the direction of $\delta_s(u-\phi^{\rm P})$ will be projected to the sliding mode $s^{*\prime}(u-\phi^{\rm A})$ along the manifold. **B)** When the animal travels east, the resulting perturbation $s^*(u-[\phi^{\rm A}-\Delta\phi_{\rm PI}])$ rotates the attractor network clockwise at a rate that does not depend on the attractor state. **C)** When the animal travels west, the resulting perturbation $s^*(u+[\phi^{\rm A}-\Delta\phi_{\rm PI}])$ rotates the attractor network clockwise at a rate that does not depend on the attractor state. **D)** Schematic of a landmark cell correcting the attractor bump (Eq. SI.III.22). A single landmark cell will pull the peak of the bump pattern towards the peak of its efferent synaptic strength profile.

3. Proof of recovery of exact path integration

When the animal is moving, there are additional velocity inputs to the attractor network from the east-conjunctive and west-conjunctive cells, yielding:

$$ds(u)/dt = \mathcal{D}_{\rm A}[s] + v_{\rm East}\epsilon_{\rm PI}\; s^*\left(u-\left[\phi^{\rm A}-\Delta\phi_{\rm PI}\right]\right) + v_{\rm West}\epsilon_{\rm PI}\; s^*\left(u-\left[\phi^{\rm A}+\Delta\phi_{\rm PI}\right]\right),$$

where v_{East}, v_{West} are the east and west velocities of the animal. Treating $\delta_s = s^{*}$, model reduction via Eq. SI.III.15 yields (Eq. 2(Main)):

$$\begin{aligned} d\phi^{\mathrm{A}}/dt = v_{\mathrm{East}}\epsilon_{\mathrm{PI}}\mathcal{F}\left(\Delta\phi_{\mathrm{PI}} + \cancel{\phi^{\mathrm{A}} - \phi^{\mathrm{A}}}\right) + v_{\mathrm{West}}\epsilon_{\mathrm{PI}}\mathcal{F}\left(-\Delta\phi_{\mathrm{PI}} + \cancel{\phi^{\mathrm{A}} - \phi^{\mathrm{A}}}\right) &= \\ \underbrace{\left[v_{\mathrm{East}} - v_{\mathrm{West}}\right]}_{v}\underbrace{\epsilon_{\mathrm{PI}}\mathcal{F}\left(\Delta\phi_{\mathrm{PI}}\right)}_{k(\text{Definition})} &= vk, \end{aligned} \tag{SI.III.18}$$

where k is a constant of proportionality that relates animal velocity to the rate of phase advance in the attractor network ($k = 2\pi/$Grid Field Spacing). Solving Eq. SI.III.18 allows us to recover path integration (Fig. SI.5) where the resulting (Eq. 3(Main)) integrated attractor phase is *only* a function of current position $x(t)$:

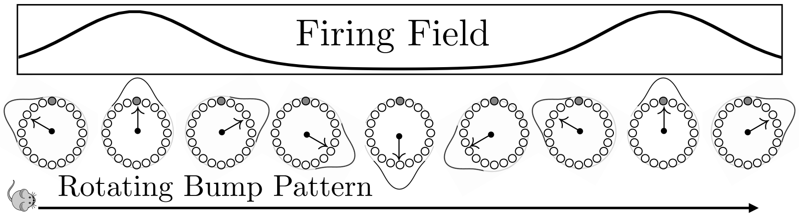

FIG. SI.5:

Solving Eq. SI.III.18 yields an attractor phase (Eq. SI.III.19), and thus individual firing rates (top cell in attractor ring, Eq. SI.III.20) which are *only* a function of current position $x(t)$.

$$\phi^{\mathrm{A}}(t) = \phi^{\mathrm{A}}(0) + k[x(t) - x(0)] \tag{SI.III.19}$$

$$\Rightarrow s(u,t) = s^{*}\left(u - \phi^{\mathrm{A}}(0) - k[x(t) - x(0)]\right). \tag{SI.III.20}$$

Thus the connectivity of the conjunctive-velocity cells in Fig. SI.4B, C ensure that as the animal moves east (west) along a 1D track, the attractor phase moves clockwise (counterclockwise), at a speed proportional to velocity. The collection of neurons in the attractor network then trace out periodic firing patterns as a function of spatial position, all with the same period but different phases.

4. Anchoring of the attractor state to landmark cell synapses

When the animal is in a landmark field, there are additional inputs to the network from landmark cells. Each landmark cell has Hebbian weights $\mathrm{W}_i(u)$ onto neurons at position u on the attractor ring. It is convenient to express these Hebbian weights in the "attractor basis", i.e., as a weighted superposition of attractor bumps with peaks at ϕ^{L} with weighting $\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})$:

$$\mathrm{W}_i(u) = \int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}}) s^{*}(u - \phi^{\mathrm{L}}). \tag{SI.III.21}$$

When a single landmark cell i is firing with rate $s_i^{\mathrm{L}}(t)$, the dynamics become:

$$ds(u)/dt = \mathcal{D}_{\mathrm{A}}[s] + \epsilon_{\mathrm{LM}} s_i^{\mathrm{L}}(t) \int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}}) s^{*}(u - \phi^{\mathrm{L}}).$$

First, we examine the effect on the attractor state from a *single* ϕ^{L}. When $\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})$ is localized near a single point ϕ^{L}, the landmark perturbation is well described by

$$\delta_s(u) = \epsilon_{\mathrm{LM}} s_i^{\mathrm{L}}(t) \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}}) s^{*}(u - \phi^{\mathrm{L}}),$$

and using Eq. SI.III.15 we recover the effective dynamics (Eq. 4(Main)) when the sole input comes from this landmark cell firing:

$$d\phi^{\mathrm{A}}/dt = \omega s_i^{\mathrm{L}}(t) \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}}) \mathcal{F}(\phi^{\mathrm{L}} - \phi^{\mathrm{A}}), \tag{SI.III.22}$$

where $\omega \propto \epsilon_{\mathrm{LM}}$. ϵ_{LM} sets the strength of synapses from the landmark cell to the attractor network, while ω is the emergent strength of the effective force that the landmark cell exerts on the reduced attractor state.

Because we have linearized the effect of input perturbations, the effect of multiple perturbations to the reduced attractor state is additive. Treating the force law $\mathcal{F}[\delta_s]$ as a functional of a perturbation, both Eq. SI.III.13 and Eq. SI.III.17 yield $\mathcal{F}\big[\delta_s^{\mathrm{A}}+\delta_s^{\mathrm{B}}\big] = \mathcal{F}\big[\delta_s^{\mathrm{A}}\big] + \mathcal{F}\big[\delta_s^{\mathrm{B}}\big]$. Therefore, the effect of a landmark cell i with *arbitrary* $\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})$ is:

$$\frac{d\phi^{\mathrm{A}}}{dt} = \omega_i s_i^{\mathrm{L}}(t)\int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})\mathcal{F}(\phi^{\mathrm{L}}-\phi^{\mathrm{A}}). \tag{SI.III.23}$$

a. Combined neural and synaptic dynamics during exploration. Again, taking advantage of the fact that weak input perturbations act additively, we combine the effect of path integration on the attractor phase ϕ^{A} described in Eq. SI.III.18 with the effect of multiple landmark cells with arbitrary learned weights $\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})$, each acting on the phase through Eq. SI.III.23, we obtain the full dynamics of attractor phase driven by both animal velocity and landmark encounters:

$$\frac{d\phi^{\mathrm{A}}}{dt} = vk + \sum_i \omega_i \mathrm{H}_i(x(t))\int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})\mathcal{F}(\phi^{\mathrm{L}}-\phi^{\mathrm{A}}). \tag{SI.III.24}$$

Recall that $s_i^{\mathrm{L}}(t) = \mathrm{H}_i(x(t))$.

Eq. SI.III.24 constitutes significant reduction of the original functional dynamics of Eq. SI.II.2. The attractor state has been reduced from an arbitrary function over the neural sheet $s(u)$ to a scalar ϕ^{A}. The effect of velocity-conjunctive cells has been reduced to ideal path integration, and the effect of landmark cells has been reduced to a distribution of forces "pulling" the attractor state to each synaptic weight peak ϕ^{L}. While these dynamics describe exactly how the reduced attractor dynamics evolve given *fixed* synaptic weights from landmark cells to attractor cells; it does not describe how the attractor weights *themselves* evolve. Next, we perform model reduction on the learning dynamics of Eq. SI.II.3.

5. Hebbian learning of landmark cell synapses

The model assumes Hebbian plasticity with weight decay, of efferent landmark cell synapses during exploration while *both* path integration and landmark cells are active. The synaptic dynamics follow Eq. SI.II.3:

$$\frac{d\mathrm{W}_i(u)}{d\mathrm{T}} = \langle s(u)|i \text{ Firing}\rangle - \mathrm{W}_i(u) = \frac{\int_t s(u,t)s_i^{\mathrm{L}}(t)}{\int_t s_i^{\mathrm{L}}(t)} - \mathrm{W}_i(u). \tag{SI.III.25}$$

Because the attractor firing rates can always be described as some translation of the steady bump pattern, $s(u,t) = s^*(u-\phi^{\mathrm{A}}(t))$, the long term average $\langle s(u)|i \text{ Firing}\rangle$ of attractor patterns $s(u)$ conditioned upon landmark cell i firing can be written as:

$$\langle s(u)|i \text{ Firing}\rangle = \int_{\phi^{\mathrm{L}}} s^*(u-\phi^{\mathrm{L}})\mathrm{Pr}(\phi^{\mathrm{A}}(t) = \phi^{\mathrm{L}}|i \text{ Firing}).$$

Thus all that matters for determining synaptic strength is the distribution of attractor phases that occur when the landmark cell fires. Again, it is convenient to use the "attractor basis", where the Hebbian weights are represented as a weighted superposition of attractor bump patterns $\mathrm{W}_i(u) = \int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})s^*(u-\phi^{\mathrm{L}})$. Representing Eq. SI.II.3 in this basis yields the learning dynamics of the synaptic weighting coefficients (see Sec. XI 1 for a proof):

$$d\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})/d\mathrm{T} = \mathrm{Pr}(\phi^{\mathrm{A}}(t) = \phi^{\mathrm{L}}|i \text{ Firing}) - \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}}). \tag{SI.III.26}$$

Together, Eq. SI.III.24 and Eq. SI.III.26 reflect a complex coupled dynamics between neurons and synapses. In Eq. SI.III.26 the distribution of attractor network activity patterns, or phases, drives plasticity in synapses from landmark cells to the attractor network. In turn, these synaptic weights modify the evolution of the attractor network phase via Eq. SI.III.24.

6. From landmark cell synapses to pinning phases

From Eq. SI.III.23 we know the effect of a *single* landmark cell on the dynamics of the attractor state is given by:

$$\frac{d\phi^{\mathrm{A}}}{dt} = \omega s^{\mathrm{L}}(t) \int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}(\phi^{\mathrm{L}}) \mathcal{F}(\phi^{\mathrm{L}} - \phi^{\mathrm{A}}).$$

We want to characterize the *entire* distribution of landmark cell synapses $\tilde{\mathrm{W}}(\phi^{\mathrm{L}})$ using only two variables: 1) The *average* pinning phase $\theta^{\mathrm{L}}_{\mathrm{Eff}}$ that the landmark pulls the attractor state to, and 2) the effective strength ω_{Eff} with which the landmark cell pulls the attractor state [15]. The approximation problem is to find $\theta^{\mathrm{L}}_{\mathrm{Eff}}$, ω_{Eff} such that the effective force on the attractor state is well approximated when the animal is within the firing field:

$$\omega_{\mathrm{Eff}} \mathrm{H}(x(t)) \mathcal{F}(\theta^{\mathrm{L}}_{\mathrm{Eff}} - \phi^{\mathrm{A}}(t)) \approx \omega \mathrm{H}(x(t)) \int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}(\phi^{\mathrm{L}}) \mathcal{F}(\phi^{\mathrm{L}} - \phi^{\mathrm{A}}(t)), \tag{SI.III.27}$$

where the approximation is averaged over the joint exploration and attractor state statistics $x(t)$, $\phi^{\mathrm{A}}(t)$. This approximation can not be exact in general, but it becomes exact in certain limits.

A. *Single bump state*

When the Hebbian weights are a *single* bump state, i.e. $\tilde{\mathrm{W}}(\phi^{\mathrm{L}}) = \delta(\phi^{\mathrm{L}} - \phi^{\mathrm{L}}_0)$, the approximation becomes *exact*, recovering Eq. SI.III.22, where $\omega_{\mathrm{Eff}} = \omega$, $\theta^{\mathrm{L}}_{\mathrm{Eff}} = \phi^{\mathrm{L}}_0$. More generally, as long as $\tilde{\mathrm{W}}(\phi^{\mathrm{L}})$ is localized, the approximation works well, where $\theta^{\mathrm{L}}_{\mathrm{Eff}}$ is the mean of the distribution $\tilde{\mathrm{W}}(\phi^{\mathrm{L}})$, i.e.

$$\theta^{\mathrm{L}}_{\mathrm{Eff}} = \int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}(\phi^{\mathrm{L}}) \phi^{\mathrm{L}},$$

where the bounds of integration contain the localized bump $\tilde{\mathrm{W}}(\phi^{\mathrm{L}})$. In this limit ω_{Eff} is close to ω and inversely related to the dispersion of $\tilde{\mathrm{W}}(\phi^{\mathrm{L}})$.

B. *Sinusoidal force law*

Another case in which the approximation works exactly and intuition can be built is when the force law is sinusoidal, i.e. $\mathcal{F}(\phi^{\mathrm{L}} - \phi^{\mathrm{A}}) = \sin(\phi^{\mathrm{L}} - \phi^{\mathrm{A}})$ (See Sec. VI for an attractor network where this is the case). Here, we can solve Eq. SI.III.27 exactly for *arbitrary* landmark cell synapses $\tilde{\mathrm{W}}(\phi^{\mathrm{L}})$. We do so by summarizing the landmark cell synapses with a single complex number:

$$z^{\mathrm{L}} = \left(\int \tilde{\mathrm{W}}(\phi^{\mathrm{L}}) e^{i\phi^{\mathrm{L}}} \right).$$

The effective pinning phase $\theta^{\mathrm{L}}_{\mathrm{Eff}}$ is then given by the angle of z^{L}, and the effective landmark strength ω_{Eff} is given by the magnitude of z^{L}

$$e^{i\theta^{\mathrm{L}}_{\mathrm{Eff}}} = z^{\mathrm{L}}/|z^{\mathrm{L}}|, \qquad \omega_{\mathrm{Eff}} = \left|z^{\mathrm{L}}\right| \omega, \qquad z^{\mathrm{L}} = e^{i\theta^{\mathrm{L}}_{\mathrm{Eff}}} \omega_{\mathrm{Eff}}. \tag{SI.III.28}$$

We can verify that the approximation Eq. SI.III.28 is exact:

$$\begin{gathered} \omega \int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}(\phi^{\mathrm{L}}) \sin(\phi^{\mathrm{L}} - \phi^{\mathrm{A}}) = \int_{\phi^{\mathrm{L}}} \mathrm{Imag} \left(\omega \tilde{\mathrm{W}}(\phi^{\mathrm{L}}) e^{i(\phi^{\mathrm{L}} - \phi^{\mathrm{A}})} \right) = \\ \mathrm{Imag} \left(z^{\mathrm{L}} e^{-i\phi^{\mathrm{A}}} \right) = \mathrm{Imag} \left(\omega_{\mathrm{Eff}} e^{i\theta^{\mathrm{L}}_{\mathrm{Eff}}} e^{-i\phi^{\mathrm{A}}} \right) = \omega_{\mathrm{Eff}} \sin \left(\theta^{\mathrm{L}}_{\mathrm{Eff}} - \phi^{\mathrm{A}} \right). \end{gathered} \tag{SI.III.29}$$

Next, in section Sec. IV, we will show a limit in which this approximation can be made exact, where the distribution of synaptic weights $\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})$ as well as the attractor state distribution $\Pr\left(\phi^{\mathrm{A}} | i \text{ firing}\right)$ are localized.

IV. LINEAR APPROXIMATION

While the equations of Eq. SI.III.24 and Eq. SI.III.26 have been significantly reduced from the original equations of Eq. SI.II.2, Eq. SI.II.3, there is an additional source of complexity in that the synaptic weights of each landmark cell are described by an *entire distribution* over attractor phases $\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})$. In this section, we will show how these dynamics can be linearized, yielding a single scalar representation for learned landmark cell synapses which can be used to characterize both exploration and learning of environments (List of symbols and units in Table II).

1. Linearized representation of weights from landmark cells to the attractor network

For landmark cells with localized firing fields (much smaller than the grid spacing), the distribution of synaptic weights $\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})$ tends to become localized [16] ; it does not spread evenly around the entire unit circle, but is rather centered in a region. Recall from Sec. III 6 that we may simplify the force exerted by a landmark cell on the attractor state in the manner of Eq. SI.III.27 by representing each landmark cell's synaptic weight distribution by its weighted average θ_i^{L}.

$$\theta_i^{\mathrm{L}} = \int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})\phi^{\mathrm{L}}. \tag{SI.IV.30}$$

Intuitively, θ_i^{L} is *also* the weighted average synaptic position (Fig. SI.6).

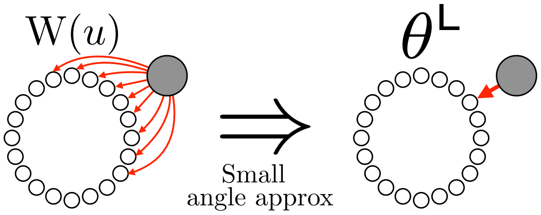

FIG. SI.6:

Schematic of the scalar single variable representation of a landmark cell's synaptic strength profile in Eq. SI.IV.30. The entire distribution of synaptic weights $\mathrm{W}(u)$ can be approximated concisely as the weighted average of efferent synaptic strengths onto the neural attractor ring, yielding a single phase variable θ^{L} description for a landmark cell's synapses. When the landmark cell fires, its effect on the attractor bump is to pull it towards θ^{L} along the neural ring.

2. Linear approximation of force law

After initial steps of learning, when the animal is within the localized firing field of a landmark cell, the bump attractor peak tends to be close to the peak synaptic output of that landmark cell (this corresponds to navigational errors that are much smaller than the spacing between firing fields). If, whenever a landmark cell i is firing, $\phi^{\mathrm{A}}(t)$ is close to all ϕ^{L} for which $\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})$ is significantly greater than zero, we may linearize the force law $\mathcal{F}(\phi^{\mathrm{L}} - \phi^{\mathrm{A}}) \approx (\phi^{\mathrm{L}} - \phi^{\mathrm{A}})$ [17] to obtain a simple linear force proportional to the difference between the attractor phase and the mean position of the landmark cell synapses onto the attractor ring. Therefore, we can simplify the force exerted by a single firing landmark cell exerted on the attractor state:

$$\int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i\left(\phi^{\mathrm{L}}\right)\ \mathcal{F}\left(\phi^{\mathrm{L}} - \phi^{\mathrm{A}}\right) \approx \int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i\left(\phi^{\mathrm{L}}\right)\ \left(\phi^{\mathrm{L}} - \phi^{\mathrm{A}}\right) = \underbrace{\int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})\phi^{\mathrm{L}}}_{\theta_i^{\mathrm{L}}} - \underbrace{\int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})\phi^{\mathrm{A}}}_{\phi^{\mathrm{A}}} = \left(\theta_i^{\mathrm{L}} - \phi^{\mathrm{A}}\right),$$

yielding dynamics (Eq. 6(Main)) of:

$$d\phi^{\mathrm{A}}/dt = kv + \sum_i \omega_i \mathrm{H}_i(x(t)) \left(\theta^{\mathrm{L}}_i - \phi^{\mathrm{A}}\right). \tag{SI.IV.31}$$

3. Linear approximation of learning rule

While the evolution of the learned Hebbian weights $\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})$ (Eq. SI.III.24) depends on the *distribution* of ϕ^{A} conditioned on a landmark cell firing, the evolution of the weighted average θ^{L}_i depends only on the *average* attractor phase conditioned on landmark firing:

$$\frac{d\theta^{\mathrm{L}}_i}{d\mathrm{T}} = \langle \phi^{\mathrm{A}} | i \text{ Firing}\rangle - \theta^{\mathrm{L}}_i. \tag{SI.IV.32}$$

Eq. SI.IV.32 (Eq. 7(Main)) can be verified by combining the attractor basis dynamics for Hebbian learning and the linear approximation for θ^{L}:

$$\begin{aligned}\frac{d\theta^{\mathrm{L}}_i}{d\mathrm{T}} = \frac{d}{d\mathrm{T}}\left(\int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})\phi^{\mathrm{L}}\right) = \int_{\phi^{\mathrm{L}}}\left(\frac{d\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})}{d\mathrm{T}}\right)\phi^{\mathrm{L}} = \int_{\phi^{\mathrm{L}}}\left(\Pr(\phi^{\mathrm{A}} = \phi^{\mathrm{L}} | i \text{ Firing}) - \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})\right)\phi^{\mathrm{L}} \\ = \underbrace{\int_{\phi^{\mathrm{L}}} \Pr(\phi^{\mathrm{A}} = \phi^{\mathrm{L}} | i \text{ Firing})\phi^{\mathrm{L}}}_{\langle \phi^{\mathrm{A}} | i \text{ Firing}\rangle} - \underbrace{\int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})\phi^{\mathrm{L}}}_{\theta^{\mathrm{L}}_i} = \langle \phi^{\mathrm{A}} | i \text{ Firing}\rangle - \theta^{\mathrm{L}}_i.\end{aligned}$$

In essence Eq. SI.IV.31 and Eq. SI.IV.32 constitute a significant model reduction of Eq. SI.II.2 and Eq. SI.II.3. In this reduction, the entire pattern of neural activity of the attractor network is summarized by a single number ϕ^{A}, denoting a point, or phase, on the ring manifold of stable attractor states. Similarly, the entire pattern of synaptic weights $\mathrm{W}_i(u)$ from landmark cell i into the attractor network is summarized by a single number θ^{L}_i, which denotes the mean position of landmark cell synaptic inputs onto the attractor ring (Fig. SI.6).

Intuitively, the reduced Eq. SI.IV.31 describes both path integration and a dynamics whereby each landmark cell i attempts to *pin* the attractor phase ϕ^{A} to the landmark cell's learned synaptic phase θ^{L}_i, each time the physical position $x(t)$ of the animal is within the landmark's firing field $\mathrm{H}_i(x)$. In turn, synaptic plasticity described in Eq. SI.IV.32 aligns the learned pinning phase θ^{L}_i of each landmark cell i to the average of the ensemble of attractor phases ϕ^{A} that occur when the animal is in the firing field of the landmark.

As we will see below, as an animal explores its environment, this coupled dynamics between attractor phase ϕ^{A} and landmark pinning phases θ^{L}_i settle into a self-consistent steady state such that the attractor phase yields an internal estimate of the animal's current position that is, to first order, largely independent of the history of the animal's previous trajectory. Moreover, each landmark cell learns a pinning phase θ^{L}_i, consistent with the location of its firing field in physical space.

V. LEARNING A SIMPLE ENVIRONMENTAL GEOMETRY

To keep the supplementary material self-contained, we largely repeat a section of the main paper describing and illustrating how the above equations yield a self-consistent neural representation of space in a simple 1D environment. A relevant list of symbols and units can be found in Table III.

Consider the linearized dynamics of Eqs. SI.IV.31, SI.IV.32 for the simple case of an animal moving back and forth between the walls of a 1D box of length L, at a constant speed $v_0 = \mathrm{L}/\tau$, yielding a total time of 2τ to complete a full cycle (Fig. 4A). In this environment we assume two landmark cells corresponding to the east (west) walls, with firing fields extending a distance $\mathrm{L}_{\mathrm{Wall}}$ into the environment leaving an empty space $\mathrm{L}_{\mathrm{Int}} = \mathrm{L} - 2\mathrm{L}_{\mathrm{Wall}}$ between (Fig. SI.7). Their pinning phases $\theta^{\mathrm{L}}_{\mathrm{E}}$ ($\theta^{\mathrm{L}}_{\mathrm{W}}$) encode the peak position of their outgoing synaptic weights. How does circuit plasticity yield a consistent environmental representation through exploration?

We will build intuition in the limit where $\mathrm{L}_{\mathrm{Wall}} \to 0$, $\omega \to \infty$; in this regime, landmark cells only act at the very edge, yet *fully* anchor the attractor state when the animal touches the edge. At $t = 0$, the animal starts at the west wall at physical position $x(0) = -\mathrm{L}/2$. Through Eq. 6, the west border cell pins the initial attractor phase so that $\phi^{\mathrm{A}}(0) = \theta^{\mathrm{L}}_{\mathrm{W}}$. At $t = \tau$, the animal travels to the east wall at physical position $x(\tau) = +\mathrm{L}/2$, and the attractor phase

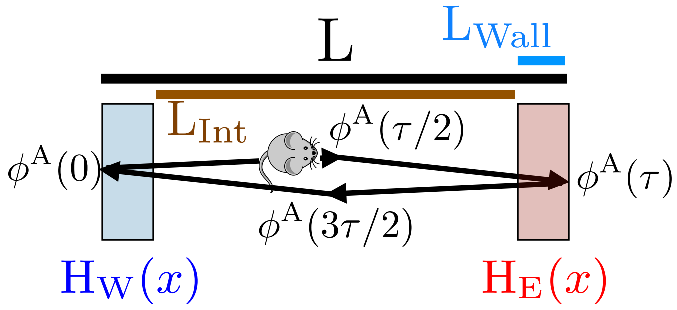

FIG. SI.7:

Schematic of an animal moving between two landmark fields in a simple 1D geometry, along with attractor phase as a function of position and path history $\left(\phi^{\text{A}}(0), \phi^{\text{A}}(\tau/2), \phi^{\text{A}}(\tau), \phi^{\text{A}}(3\tau/2)\right)$. The landmark fields extend a distance of L_{Wall} into the environment leaving an empty space $\text{L}_{\text{Int}} = \text{L} - 2\text{L}_{\text{Wall}}$ between.

advances due to path integration to become $\phi^{\text{A}}(\tau^-) = \theta^{\text{L}}_{\text{W}} + k\text{L}$. However, upon encountering the east wall, the east border cell pins the attractor phase to $\theta^{\text{L}}_{\text{E}}$.

Before any learning, there is no guarantee that the east border cell pinning phase $\theta^{\text{L}}_{\text{E}}$ equals the attractor phase $\theta^{\text{L}}_{\text{W}} + k\text{L}$, obtained by starting at the west wall and moving to the east wall; sensation and path integration might disagree (Fig. 4B). However, plasticity described in Eq. 7 will act so as to move $\theta^{\text{L}}_{\text{E}}$ closer to $\theta^{\text{L}}_{\text{W}} + k\text{L}$. Then as the animal returns to the left wall at time $t = 2\tau$, path integration will retard the attractor phase $\phi^{\text{A}}(2\tau) = \theta^{\text{L}}_{\text{E}} - k\text{L}$, and an encounter with the west wall leads the west border cell to pin the attractor phase to $\theta^{\text{L}}_{\text{W}}$. Again, there is no guarantee that the west border cell pinning phase $\theta^{\text{L}}_{\text{W}}$ agrees with the attractor phase $\theta^{\text{L}}_{\text{E}} - k\text{L}$ obtained by starting at the east wall and traveling to the west wall, but circuit plasticity will change $\theta^{\text{L}}_{\text{W}}$ to reduce this discrepancy. Overall, plasticity over multiple cycles of exploration yields the iterative dynamics

$$\theta^{\text{L}}_{\text{E}} \to \theta^{\text{L}}_{\text{W}} + k\text{L}, \qquad \theta^{\text{L}}_{\text{W}} \to \theta^{\text{L}}_{\text{E}} - k\text{L}.$$

1. Learning as an elastic relaxation between landmarks.

To gain further insight into the learning dynamics, it is useful to interpret the periodic attractor phase $\phi^{\text{A}}(t)$ as an internal estimate of position through the "unrolled" coordinate variable

$$\mathcal{X}^{\text{A}} = \phi^{\text{A}}/k. \tag{SI.V.33}$$

Likewise, we can replace the landmark phase θ^{L}_i with another linear variable

$$\mathcal{X}^{\text{L}}_i = \theta^{\text{L}}_i/k, \tag{SI.V.34}$$

denoting the internal representation of the position of landmark i (Fig. 4D). This enables us to associate physical positions to landmark cells, or more precisely their pinning phases, although these assigned positions are defined only up to shifts of the grid period. Plasticity over the long timescale T of exploration then yields the following learning dynamics for the physical positions in unrolled phase for the landmark cells:

$$d\mathcal{X}^{\text{L}}_{\text{E}}/d\text{T} = -\text{M}_{\text{EW}}\left[\mathcal{X}^{\text{L}}_{\text{E}} - \left(\mathcal{X}^{\text{L}}_{\text{W}} + \Delta\mathcal{X}^{\text{A}}_{\text{W}\to\text{E}}\right)\right] \tag{SI.V.35}$$

$$d\mathcal{X}^{\text{L}}_{\text{W}}/d\text{T} = -\text{M}_{\text{WE}}\left[\mathcal{X}^{\text{L}}_{\text{W}} - \left(\mathcal{X}^{\text{L}}_{\text{E}} + \Delta\mathcal{X}^{\text{A}}_{\text{E}\to\text{W}}\right)\right], \tag{SI.V.36}$$

where $\Delta\mathcal{X}^{\text{A}}_{\text{W}\to\text{E}} = -\Delta\mathcal{X}^{\text{A}}_{\text{E}\to\text{W}} = \text{L}$, and $\text{M}_{\text{EW}} = \text{M}_{\text{WE}}$.

These dynamics for the two landmark cell synapses in unrolled phase are equivalent to those of two particles at physical positions $\mathcal{X}^{\text{L}}_{\text{W}}$ and $\mathcal{X}^{\text{L}}_{\text{E}}$, connected by an overdamped spring with rest length L, and spring constant M_{WE} which sets the learning rate (Fig. 4E). If the separation $\mathcal{X}^{\text{L}}_{\text{E}} - \mathcal{X}^{\text{L}}_{\text{W}}$ between the particles is less (greater) than L, then the spring is compressed (extended) yielding a repulsive (attractive) force between the two particles. Learning stabilizes the two particle positions when their separation equals the spring rest length, so that $\mathcal{X}^{\text{L}}_{\text{E}} - \mathcal{X}^{\text{L}}_{\text{W}} = \text{L}$. This condition in unrolled phase is equivalent to the fundamental consistency condition for a well defined spatial map, namely that the phase advance due to path integration equals the phase difference between the pinning phases of landmark cells (Fig. 4C). However the utility of the unrolled phase representation lies in revealing a compelling picture for how a spatially consistent map arises from the combined neuronal and synaptic dynamics, through a simple, emergent first order relaxational dynamics of landmark particles connected by damped springs. As we see below, this simple effective particle-spring description of synaptic plasticity in response to spatial exploration generalizes to arbitrary landmarks in arbitrary two dimensional environments.

We note that if the environment has not been fully learned or has been recently deformed, the internal representation of landmarks in unrolled phase will lag behind the true geometry for a time, leading to "boundary-tethered" firing fields seen in [18, 19]. Additionally, we have solved the dynamics when the firing fields of the border cells have a finite extent $\mathrm{L}_{\mathrm{Wall}}$ and the landmark cells have a finite strength ω, and we find the dynamics obeys that of Eq. SI.V.35 and Eq. SI.V.36 (Eq. 8(Main) and Eq. 9(Main)). One notable difference is that the internal map will be contracted, and the rest length will be $\Delta\mathcal{X}^{\mathrm{A}}_{\mathrm{W}\to\mathrm{E}} = \mathrm{L}_{\mathrm{Int}} + 2v_0 \tanh\left(\omega \mathrm{L}_{\mathrm{Wall}}/2v_0\right)/\omega < \mathrm{L}$ (See Sec. XI 2). We note that this still produces a consistent, path-independent representation in the center of the track, far from the firing fields of either border cell.

VI. AN EXAMPLE OF AN EXACTLY SOLVABLE RING ATTRACTOR MODEL

To build intuition, and to derive explicit equations in a concrete setting, we turn to a modified and simplified version of the model of Ben-Yishai et al. [20] which is exactly solvable and yields analytic, effective force laws in the model reduced description.

Without external inputs, the model follows the dynamics:

$$\frac{ds\left(u\right)}{dt} = -s(u) + \mathcal{G}\left(\int_{u'} \mathrm{J}(u-u')s(u')\right), \tag{SI.VI.37}$$

where $\mathrm{J}(\Delta u) = \mathrm{J}_0\cos(\Delta u)$, and the nonlinearity is defined by:

$$\mathcal{G}(h) = \begin{cases} -1 & h \leq -1 \\ h & -1 < h < 1 \\ 1 & h \geq 1. \end{cases}$$

1. Steady state solution

Any steady state solution will have the form

$$s^{*}(u) = \mathcal{G}(h(u)),$$

where

$$h(u) = \int_{u'} \mathrm{J}(u-u')s^{*}(u') = \mathrm{J}_0\left(\int_{u'}\cos(u-u')s^{*}(u')\right)$$

To solve these equations, we note that regardless of $s^{*}(u)$, $h(u)$ will have the form $h_0\cos(u-\phi^{\mathrm{A}})$. Thus any attractor solution must have the form $s^{*}(u) = \mathcal{G}(h_0\cos(u-\phi^{\mathrm{A}}))$. When $\mathrm{J}_0 < 1/\pi$, the only steady state is the uniform state $s^{*}(u) = 0$; however, when J_0 approaches $1/\pi$ from above, the uniform state becomes unstable, and there is now a stable steady state solution of

$$s^{*}(u) = \cos(u),$$

and the nonlinearity $\mathcal{G}$ is barely triggered (Fig. SI.8).

2. Effective force function

When J_0 approaches $1/\pi$ from above, the non-linearity is barely in effect, i.e. $\mathcal{G}(s^{*}(u)) = s^{*}(u)$ for nearly all u [21]. Defining the dynamics Eq. SI.VI.37 by the functional $\mathcal{D}_{\mathrm{A}}[s]$ we see that the Jacobian around any steady state, $\mathrm{Jac}_{\phi^{\mathrm{A}}}$, is nearly symmetric, i.e.

$$\mathrm{Jac}_{\phi^{\mathrm{A}}}(u',u) = \frac{d\left(\mathcal{D}_{\mathrm{A}}[s](u')\right)}{ds(u)} = \frac{d\left(\mathcal{D}_{\mathrm{A}}[s](u)\right)}{ds(u')} = \mathrm{Jac}_{\phi^{\mathrm{A}}}(u,u') \approx \mathrm{J}_0 \cos(u'-u),$$

and so the eigenvectors of the Jacobian are orthogonal. Therefore, any perturbation applied will be projected along the sliding mode in the manner of Eq. SI.III.14. For example, if we apply a perturbation of $\epsilon\cos(u-\phi^{\mathrm{P}})$, the effective sliding dynamics of the attractor bump will be given by,

$$\frac{ds(u)}{dt} = \epsilon\mathcal{F}(\phi^{\mathrm{P}}-\phi^{\mathrm{A}})s^{*\prime}(u-\phi^{\mathrm{A}}),$$

where the attractor force law is $\mathcal{F}(\phi^{\mathrm{P}}-\phi^{\mathrm{A}}) = \sin(\phi^{\mathrm{P}}-\phi^{\mathrm{A}})$.

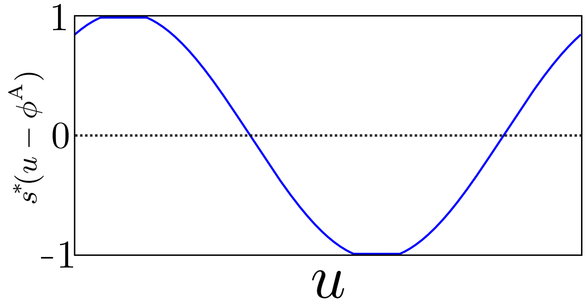

FIG. SI.8:

Steady attractor state $s^*(u-\phi^{\mathrm{A}})$ for Eq. SI.VI.37 $\mathrm{J}_0 = 1.01/\pi$, 1% above the threshold value. The network is in the linear regime in almost all parts of the neural sheet, except for the very top and bottom, which are truncated at ±1. As J_0 approaches $1/\pi$ from above, these nonlinear regions become infinitely small, and the Jacobian becomes symmetric.

Adding velocity conjunctive cells and landmark cells in the same manner as Sec. II, we obtain the explicit, effective reduced dynamics:

$$\frac{d\phi^{\mathrm{A}}}{dt} = \underbrace{\epsilon_{\mathrm{PI}}\sin(\Delta\phi_{\mathrm{PI}})v}_{kv} + \epsilon_{\mathrm{LM}}\sum\nolimits_i \mathrm{H}_i\left(x(t)\right)\tilde{\mathrm{W}}(\phi^{\mathrm{L}})\sin(\phi^{\mathrm{L}}-\phi^{\mathrm{A}}), \qquad \frac{d\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})}{d\mathrm{T}} = \Pr(\phi^{\mathrm{L}}|i\ \mathrm{Firing}) - \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}}). \qquad \text{(SI.VI.38)}$$

VII. GENERALIZATION TO TWO-DIMENSIONAL GRID CELLS

In order to make contact with experiments, we generalize all of the above to two dimensional space, and two-dimensional attractor models yielding grid cells. See Table V for a list of symbols and units.

Now attractor network grid cells live on a periodic *two-dimensional* neural sheet, where each neuron has position $\mathbf{u} = (u_1, u_2)$.

1. Velocity-conjunctive cells

When we generalize to 2D, there are now *four* kinds of velocity conjunctive cells: east-conjunctive, west conjunctive, north-conjunctive and south-conjunctive. Each one of these four cell-types live on their own distinct neural sheet with the same coordinate $\mathbf{u} = (u_1, u_2)$ as the sheet corresponding to the attractor network sheet that contains pure, non-conjunctive grid cells. The firing rates of these conjunctive cells at their own neural sheet position $\mathbf{u}$ depend instantaneously on the firing rate $s(\mathbf{u})$ of the non-conjunctive grid cells at their corresponding position $\mathbf{u}$, as well as

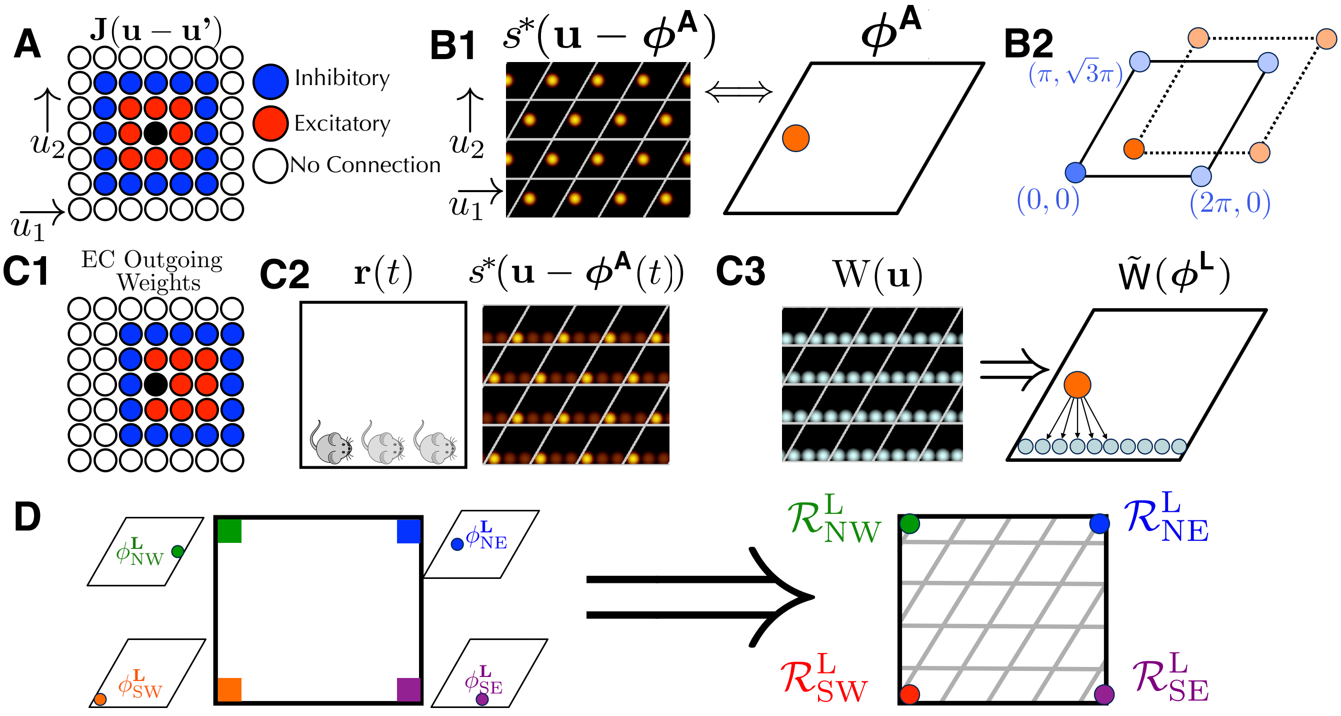
A
$\mathbf{J}(\mathbf{u}-\mathbf{u'})$
u_2
u_1
Inhibitory
Excitatory
No Connection
B1
$s^*(\mathbf{u}-\boldsymbol{\phi}^{\mathbf{A}})$
$\boldsymbol{\phi}^{\mathbf{A}}$
u_2
u_1
B2
$(\pi, \sqrt{3}\pi)$
$(0,0)$
$(2\pi,0)$
C1
EC Outgoing Weights
C2
$\mathbf{r}(t)$
$s^*(\mathbf{u}-\boldsymbol{\phi}^{\mathbf{A}}(t))$
C3
$\mathrm{W}(\mathbf{u})$
$\tilde{\mathrm{W}}(\boldsymbol{\phi}^{\mathbf{L}})$
D
$\phi^{\mathbf{L}}_{\mathrm{NW}}$
$\phi^{\mathbf{L}}_{\mathrm{NE}}$
$\phi^{\mathbf{L}}_{\mathrm{SW}}$
$\phi^{\mathbf{L}}_{\mathrm{SE}}$
$\mathcal{R}^{\mathrm{L}}_{\mathrm{NW}}$
$\mathcal{R}^{\mathrm{L}}_{\mathrm{NE}}$
$\mathcal{R}^{\mathrm{L}}_{\mathrm{SW}}$
$\mathcal{R}^{\mathrm{L}}_{\mathrm{SE}}$

FIG. SI.9:

A) A 2D neural sheet with short-range excitation and long-range inhibition, analogous to Fig. 1. Each neuron on the continuous sheet now has coordinates $\mathbf{u} = (u_1, u_2)$. **B1)** A 2D analogue of a single attractor pattern on the neural sheet, with high firing rates in red (compare to Fig. SI.1A). The set all unique stable attractor patterns is now indexed not by a single phase variable as in 1D, but a 2D phase variable $\boldsymbol{\phi}^{\mathbf{A}}$ ranging over a rhombus, the unit cell of the attractor bump pattern. Copies of the unit cell are shown via white lines. **B2**) Schematic of the discrete translational symmetry of the bump attractor state Eq. SI.VII.41. **C1)** Schematic of outgoing weights for east-conjunctive cell (while Eq. SI.VII.39 uses one-to-one conjunctive weights, this can be easily extended). North, south, and west-conjunctive weights can be constructed in the same way. **C2)** As the animal travels along the south wall, the average firing rates will form a "streak" across the neural sheet. **C3)** The landmark cell Hebbian weights will be a combination of 2D attractor states (Eq. SI.VII.43).This leads the Hebbian weights on the neural sheet to form the same streak; this learned state can be represented as a *distribution* over the periodic rhombus. Analogously, there is a force law, where the state of an attractor network $\boldsymbol{\phi}^{\mathbf{A}}$ will be pulled towards this distribution $\tilde{\mathrm{W}}_i(\boldsymbol{\phi}^{\mathbf{L}})$ (Eq. SI.VII.42). **D)** Here, we can unroll the two-dimensional attractor phase into a two-dimensional position variable, thereby associating landmark pinning phases to points in physical space. Given landmarks in all four corners, the landmark pinning phases correspond to different points on the phase rhombus, but through unrolling this rhombus, each can be associated to a physical corner of the environment.

on the animal running velocity through the formulas:

$$s_{\mathrm{NC}}(\mathbf{u}) = (\mathrm{v}_{\mathrm{North}}/v_{\mathrm{C0}})s(\mathbf{u}), \quad s_{\mathrm{EC}}(\mathbf{u}) = (\mathrm{v}_{\mathrm{East}}/v_{\mathrm{C0}})s(\mathbf{u}), \quad s_{\mathrm{SC}}(\mathbf{u}) = (\mathrm{v}_{\mathrm{South}}/v_{\mathrm{C0}})s(\mathbf{u}), \quad s_{\mathrm{WC}}(\mathbf{u}) = (\mathrm{v}_{\mathrm{West}}/v_{\mathrm{C0}})s(\mathbf{u}).$$

Here v_{C0} is some characteristic speed at which the velocity-conjunctive cells fire at the same rate as the non-conjunctive attractor cells, and $\mathrm{v}_{\mathrm{North}}, \mathrm{v}_{\mathrm{East}}, \mathrm{v}_{\mathrm{South}}, \mathrm{v}_{\mathrm{West}}$ are the north, east, south and west components of animal velocity respectively:

$$\mathrm{v}_{\mathrm{North}} = [\mathbf{v}\cdot\hat{\mathbf{y}}]_+, \quad \mathrm{v}_{\mathrm{East}} = [\mathbf{v}\cdot\hat{\mathbf{x}}]_+, \quad \mathrm{v}_{\mathrm{South}} = [-\mathbf{v}\cdot\hat{\mathbf{y}}]_+, \quad \mathrm{v}_{\mathrm{West}} = [-\mathbf{v}\cdot\hat{\mathbf{x}}]_+, \quad \mathbf{v} = (\mathrm{v}_{\mathrm{North}} - \mathrm{v}_{\mathrm{South}})\,\hat{\mathbf{y}} + (\mathrm{v}_{\mathrm{East}} - \mathrm{v}_{\mathrm{West}})\,\hat{\mathbf{x}},$$

where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ are unit vectors in the x and y directions respectively.

2. Full two-dimensional neural dynamics

The full dynamics of the attractor network (extended version of Eq. 10(Main)), analogous to Eq. SI.II.2, are

$$\frac{ds\left(\mathbf{u}\right)}{dt} = -\frac{s(\mathbf{u})}{\tau_m} + \mathcal{G}\left(\int_{\mathbf{u}'} \mathbf{J}(|\mathbf{u}-\mathbf{u}'|)s(\mathbf{u}')\right) + \epsilon_{\mathrm{LM}} \sum_i \underbrace{\left(\mathrm{W}_i(\mathbf{u})s_i^{\mathrm{L}}(t)\right)}_{\text{Landmark cell inputs}} +$$

$$\underbrace{\epsilon_{\mathrm{PI}}\left(s_{\mathrm{NC}}\left(\mathbf{u}-|\Delta\phi_{\mathrm{PI}}|\ \hat{\mathbf{u}}_2\right)\right)}_{\text{Input from north cells}} + \underbrace{\epsilon_{\mathrm{PI}}\left(s_{\mathrm{EC}}\left(\mathbf{u}-|\Delta\phi_{\mathrm{PI}}|\ \hat{\mathbf{u}}_1\right)\right)}_{\text{Input from east cells}} + \underbrace{\epsilon_{\mathrm{PI}}\left(s_{\mathrm{SC}}\left(\mathbf{u}+|\Delta\phi_{\mathrm{PI}}|\ \hat{\mathbf{u}}_2\right)\right)}_{\text{Input from south cells}} + \underbrace{\epsilon_{\mathrm{PI}}\left(s_{\mathrm{WC}}\left(\mathbf{u}+|\Delta\phi_{\mathrm{PI}}|\ \hat{\mathbf{u}}_1\right)\right)}_{\text{Input from west cells}}, \tag{SI.VII.39}$$

where the firing rates of the landmark cells are a function of 2D animal position:

$$s_i^{\mathrm{L}}(t) = \mathrm{H}_i(\mathbf{r}(t)).$$

3. Hebbian learning of landmark cells in two-dimensions

The long term learning is mediated by the updates of the Hebbian weights $\mathrm{W}_i(\mathbf{u})$ from the landmark cells to the attractor network in a manner analogous to Eq. SI.II.3:

$$\frac{d\mathrm{W}_i(\mathbf{u})}{d\mathrm{T}} = \langle s(\mathbf{u})|i \text{ Firing}\rangle - \mathrm{W}_i(\mathbf{u}) = \frac{\int_t s(\mathbf{u},t)s_i^{\mathrm{L}}(t)}{\int_t s_i^{\mathrm{L}}(t)} - \mathrm{W}_i(\mathbf{u}). \tag{SI.VII.40}$$

4. Two-dimensional model reduction

A. *Reduced two-dimensional attractor state*

Applying the same techniques as before, we see that attractor dynamics on a *two-dimensional* neural sheet can now yield a *two-dimensional* family of stable, or steady state, localized bump activity patterns $s^{*}(\mathbf{u}-\boldsymbol{\phi}^{\mathbf{A}})$. When 2D attractor dynamics yield a family of steady hexagonal bump patterns, this periodicity on the neural sheet has *hexagonal* symmetry(Fig. SI.9B2), and can be represented mathematically on the neural sheet as:

$$s^{*}(u_1,u_2) = s^{*}(u_1+2\pi,u_2) = s^{*}(u_1+\pi,u_2+\sqrt{3}\pi),$$

where we have defined units on the neural sheet in terms of this periodicity. Therefore, the coordinate $\boldsymbol{\phi}^{\mathbf{A}}$ specifying a point on the manifold of stable attractor patterns is a periodic variable defined modulo the periodicity of the steady state pattern:

$$\boldsymbol{\phi}^{\mathbf{A}} \equiv \boldsymbol{\phi}^{\mathbf{A}} + (2\pi,0) \equiv \boldsymbol{\phi}^{\mathbf{A}} + (\pi,\sqrt{3}\pi). \tag{SI.VII.41}$$

The attractor state is now a 2D phase $\boldsymbol{\phi}^{\mathbf{A}}$ on the periodic rhombus (Fig. SI.9B2).

B. *Reduced short-timescale exploration dynamics*

Likewise, we may obtain a 2D analogue (Eq. 15(Main)) to the dynamics of the attractor state:

$$d\boldsymbol{\phi}^{\mathbf{A}}/dt = \mathbf{K}\,d\mathbf{r}/dt\ +\ \sum_i \omega_i \mathrm{H}_i(\mathbf{r}(t)) \int_{\boldsymbol{\phi}^{\mathbf{L}}} \tilde{\mathrm{W}}_i\left(\boldsymbol{\phi}^{\mathbf{L}}\right)\ \boldsymbol{\mathcal{F}}\left(\boldsymbol{\phi}^{\mathbf{L}}-\boldsymbol{\phi}^{\mathbf{A}}\right). \tag{SI.VII.42}$$

Here we have replaced the 1D gain scalar, k, with $\mathbf{K}$, a 2×2 matrix that translates 2D animal velocity into phase advance in the 2D attractor network; $\mathbf{K}$ determines *both* grid spacing *and* orientation. When (north, east, south, west) cells have outgoing connections in the $\hat{\mathbf{u}}_2, \hat{\mathbf{u}}_1, -\hat{\mathbf{u}}_2, -\hat{\mathbf{u}}_1$ directions, $\mathbf{K}$ is a multiple of the identity and yields NESW oriented grid fields. When grid fields are at an angle, $\mathbf{K}$ will be some multiple of a rotation matrix.

C. Two-dimensional learning dynamics

We may likewise obtain the analogue (Eq. 13(Main)) of the learning dynamics of Eq. SI.III.26:

$$d\tilde{\mathrm{W}}_i(\boldsymbol{\phi}^{\mathbf{L}})/d\mathrm{T} = \Pr(\boldsymbol{\phi}^{\mathbf{A}}(t) = \boldsymbol{\phi}^{\mathbf{L}}|i \text{ Firing}) - \tilde{\mathrm{W}}_i(\boldsymbol{\phi}^{\mathbf{L}}), \tag{SI.VII.43}$$

where $\tilde{\mathrm{W}}_i(\boldsymbol{\phi}^{\mathbf{L}})$ is now a distribution over the periodic rhombus (Fig. SI.9C).

D. Two-dimensional linearized dynamics

Continuing further, we may make a small-angle approximation (Analogous to Eq. SI.V.33 but now using 2D variables with periodicity over the rhombus) to replace the attractor phase $\boldsymbol{\phi}^{\mathbf{A}}(t)$ with a *two-dimensional* unrolled linear variable:

$$\boldsymbol{\mathcal{R}}^{\mathbf{A}}(t) \equiv \mathbf{K}^{-1}\ \boldsymbol{\phi}^{\mathbf{A}}(t), \tag{SI.VII.44}$$

reflecting an internal estimate of instantaneous position in 2D physical space.

Likewise we may replace the distribution of Hebbian landmark weights $\tilde{\mathrm{W}}_i(\boldsymbol{\phi}^{\mathbf{L}})$ with a single 2D phase variable $\boldsymbol{\theta}_i^{\mathbf{L}}$ on the rhombus representing the weighted average (Analogous to Eq. SI.IV.30) of its synaptic weight distribution:

$$\boldsymbol{\theta}_i^{\mathbf{L}} = \iint \tilde{\mathrm{W}}_i(\boldsymbol{\phi}^{\mathbf{L}})\boldsymbol{\phi}^{\mathbf{L}}. \tag{SI.VII.45}$$

Analogously to Eq. SI.V.34, we can unroll the phase variable $\boldsymbol{\theta}_i^{\mathbf{L}}$ into a linear variable,

$$\boldsymbol{\mathcal{R}}_i^{\mathbf{L}} = \mathbf{K}^{-1}\boldsymbol{\theta}_i^{\mathbf{L}} \tag{SI.VII.46}$$

associated with a physical position in real space (Fig. SI.9D) up to the grid periodicity.

This reduction yields two-dimensional dynamics (Eq. 16(Main), Eq. 17(Main)) for internal estimates of position (i.e. unrolled attractor phase) and internal estimates of landmark position (i.e. unrolled mean phase of landmark cell synapses), given in analogy to Eqs. SI.IV.31, SI.IV.32 by:

$$d\boldsymbol{\mathcal{R}}^{\mathbf{A}}/dt = d\mathbf{r}/dt + \sum \omega_i\ \mathrm{H}_i(\mathbf{r}(t))\left(\boldsymbol{\mathcal{R}}_i^{\mathbf{L}} - \boldsymbol{\mathcal{R}}^{\mathbf{A}}\right), \tag{SI.VII.47}$$

$$d\boldsymbol{\mathcal{R}}_i^{\mathbf{L}}/d\mathrm{T} = \left\langle \boldsymbol{\mathcal{R}}^{\mathbf{A}}(t)|\text{Cell } i \text{ Firing}\right\rangle - \boldsymbol{\mathcal{R}}_i^{\mathbf{L}}. \tag{SI.VII.48}$$

VIII. REDUCING THE JOINT EXPLORATION AND LEARNING DYNAMICS TO A MECHANICAL MASS, SPRING SYSTEM.

A list of symbols and units can be found in Table IV.

We showed in Eq. SI.V.35 and Eq. SI.V.36, that the emergence of spatial consistency between path integration and landmarks through Hebbian learning dynamics, during exploration of a simple 1D environment, could be understood as the outcome of an elastic relaxation process between landmark cell synapses, viewed as particles in physical space connected by damped springs. Remarkably, this result generalizes far beyond this simple environment. As long as the exploration dynamics are time-reversible [22], the learning dynamics of *any* set of landmark cells in *any* geometry in 2D (and 1D) yields this particle-spring interpretation (Eq. 18(Main)):

$$d\boldsymbol{\mathcal{R}}_i^{\mathbf{L}}/d\mathrm{T} = -\sum_j \mathrm{M}_{ij}\left(\boldsymbol{\mathcal{R}}_i^{\mathbf{L}} - \left[\boldsymbol{\mathcal{R}}_j^{\mathbf{L}} + \Delta\boldsymbol{\mathcal{R}}_{j\to i}^{\mathbf{A}}\right]\right). \tag{SI.VIII.49}$$

The spring constant M_{ij} is related to the frequency with which the animal moves between each pair of landmark firing fields i, j, while the rest displacement $\Delta\boldsymbol{\mathcal{R}}_{j\to i}^{\mathbf{A}}$ is the average change in unrolled attractor phase as the animal moves from firing field j to field i, roughly related to the distance between the landmark firing fields. Below, we show how this is derived, as well as how precise expressions for the spring constants and rest lengths are determined by the statistics of exploration.

1. Animal position self-estimate as a function of position and path history

We first need to solve for an animal's internal estimate of position (i.e. its unrolled attractor phase) as a function of its path history. To do so we first make the bookkeeping substitution:

$$d\boldsymbol{\mathcal{R}}^{\mathbf{A}}/dt = \underbrace{d\mathbf{r}/dt}_{\text{Path Integration}} + \underbrace{\omega(\mathbf{r}(t))\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}(t)) - \boldsymbol{\mathcal{R}}^{\mathbf{A}}\right]}_{\text{Landmark Cells}} \tag{SI.VIII.50}$$

where $\omega(\mathbf{r}) = \sum \omega \mathrm{H}_i\left(\mathbf{r}\right)$ is the combined strength of all landmark cells that fire at $\mathbf{r}$, and $\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}) = \sum\left[\mathrm{H}_i\left(\mathbf{r}\right)\boldsymbol{\mathcal{R}}^{\mathbf{L}}_i\right]/\omega(\mathbf{r})$ is the average position estimate being reinforced at position $\mathbf{r}$. As the animal moves around the environment, the position self-estimate will get pushed to the learned reinforcing positions of landmarks the animal visits, path integrated as the animal moves, and eventually forgotten as the animal encounters new landmarks. We can take this basic intuition and turn it into a closed-form equation (Verified in Sec. XI 3); given any path history $\mathbf{r}(t)$ the solution for Eq. SI.VIII.50 is:

$$\boldsymbol{\mathcal{R}}^{\mathbf{A}}[\mathbf{r}(t), t] = \int_{-\infty}^{t} \underbrace{\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}\left(\mathbf{r}(t')\right) + \left(\mathbf{r}(t) - \mathbf{r}(t')\right)\right]}_{\text{Landmark Position Estimate + Path Integration from t'}} \times \underbrace{\left(\omega(\mathbf{r}(t'))e^{-\int_{t'}^{t}\omega(\mathbf{r}(t''))dt''}\right)}_{\text{Memory of time } t'} dt' \tag{SI.VIII.51}$$

a. Solving for learned position estimates as a function of current landmark position estimates. We now need to compute the mean position-self estimate seen by each landmark cell. We note that for any *individual* path, $\boldsymbol{\mathcal{R}}^{\mathbf{A}}[\mathbf{r}(t), t]$ is linear with respect to $\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}')$. Therefore, defining $\mathbf{r}_{\mathrm{A}}$ and $\mathbf{r}_{\mathrm{B}}$ as the starting and ending positions of a path, we can show that the average $\bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}(\mathbf{r}_{\mathrm{B}})$ is *also* linear with $\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}_{\mathrm{A}})$ by averaging over *all* paths starting at $\mathbf{r}_{\mathrm{A}}$ and ending at $\mathbf{r}_{\mathrm{B}}$. Therefore, we can construct a matrix equation:

$$\bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}(\mathbf{r}_{\mathrm{B}}) = \int_{\mathbf{r}_{\mathrm{A}}} \mathrm{S}(\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}})\left[\omega(\mathbf{r}_{\mathrm{A}})\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}_{\mathrm{A}}) + (\mathbf{r}_{\mathrm{B}} - \mathbf{r}_{\mathrm{A}})\right)\right],$$

where our matrix entries $\mathrm{S}(\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}})$ represent all possible ways the landmark position-estimates at position $\mathbf{r}_{\mathrm{A}}$ contribute to the mean position self-estimate at $\mathbf{r}_{\mathrm{B}}$. As long as the exploration dynamics are reversible, i.e., for any $\mathbf{r}(t)$, the reverse path $\mathbf{r}(-t)$ is equally likely, S is symmetric ($\mathrm{S}(\mathbf{r}_{\mathrm{A}}, \mathbf{r}_{\mathrm{B}}) = \mathrm{S}(\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}})$, proof in Sec. XI 4).

To solve for the learning dynamics, we expand $\omega(\mathbf{r}), \boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r})$ to understand the average position self-estimate as a function of position and the landmark position estimates of all landmark cells j:

$$\bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}(\mathbf{r}_{\mathrm{B}}) = \sum_j \int_{\mathbf{r}_{\mathrm{A}}} \mathrm{S}(\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}})\mathrm{H}_j(\mathbf{r}_{\mathrm{A}})\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}_j + (\mathbf{r}_{\mathrm{B}} - \mathbf{r}_{\mathrm{A}})\right)d\mathbf{r}'.$$

The mean position self-estimate seen by each landmark cell i is then:

$$\bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}_i = \sum_j \iint_{\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}}} \mathrm{H}_i(\mathbf{r}_{\mathrm{B}})\mathrm{S}(\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}})\mathrm{H}_j\left(\mathbf{r}_{\mathrm{A}}\right)\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}_j + (\mathbf{r}_{\mathrm{B}} - \mathbf{r}_{\mathrm{A}})\right).$$

Combining this with the landmark learning rule gives:

$$\frac{d\boldsymbol{\mathcal{R}}^{\mathbf{L}}_i}{d\mathrm{T}} = \bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}_i - \boldsymbol{\mathcal{R}}^{\mathbf{L}}_i = \sum_j \iint_{\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}}} \mathrm{H}_i(\mathbf{r}_{\mathrm{B}})\mathrm{S}(\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}})\mathrm{H}_j\left(\mathbf{r}_{\mathrm{A}}\right)\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}_j + (\mathbf{r}_{\mathrm{B}} - \mathbf{r}_{\mathrm{A}})\right] - \boldsymbol{\mathcal{R}}^{\mathbf{L}}_i.$$

$$= \sum_j \underbrace{\left[\iint_{\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}}} \mathrm{H}_i(\mathbf{r}_{\mathrm{B}})\mathrm{S}(\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}})\mathrm{H}_j\left(\mathbf{r}_{\mathrm{A}}\right)\right]}_{\mathrm{M}_{ij}\,\text{(Definition)}}\boldsymbol{\mathcal{R}}^{\mathbf{L}}_j + \sum_j \underbrace{\left[\iint_{\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}}} \mathrm{H}_i(\mathbf{r}_{\mathrm{B}})\mathrm{S}(\mathbf{r}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}})\mathrm{H}_j\left(\mathbf{r}_{\mathrm{A}}\right)(\mathbf{r}_{\mathrm{B}} - \mathbf{r}_{\mathrm{A}})\right]}_{\mathrm{M}_{ij}\Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{j\to i}\ \text{(Definition)}} - \boldsymbol{\mathcal{R}}^{\mathbf{L}}_i.$$

Where we have divided the contributions into symmetric components M_{ij} that depend on the landmark states $\boldsymbol{\mathcal{R}}^{\mathbf{L}}_i$ and antisymmetric components $\mathrm{M}_{ij}\Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{j\to i}$ which depend on path integration. We note that $\sum_j \mathrm{M}_{ij} = 1$ For all i [23]; therefore, we can rewrite the above equation as:

$$\frac{d\boldsymbol{\mathcal{R}}^{\mathbf{L}}_i}{d\mathrm{T}} = \sum_j \mathrm{M}_{ij}\left(\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}_j + \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{j\to i}\right] - \boldsymbol{\mathcal{R}}^{\mathbf{L}}_i\right).$$

Because $\mathrm{S}(\mathbf{r}_\mathrm{A}, \mathbf{r}_\mathrm{B}) = \mathrm{S}(\mathbf{r}_\mathrm{B}, \mathbf{r}_\mathrm{A})$, we can see that $\mathrm{M}_{ij} = \mathrm{M}_{ji}$ and $\Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{j\to i} = -\Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{i\to j}$. Therefore, the long term dynamics of mapping are equivalent to the first-order dynamics of a set of particles i, attached by damped springs of strength M_{ij}, each having a rest displacement vector of $\Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{j\to i}$. The spring constant is M_{ij} related to the frequency with which the animal moves between each landmark field, while the rest displacement $\Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{j\to i}$ is a weighted average of the distances between pairs of points in the two landmark fields.

While we have presented the 2D proof for convenience, this proof also works for navigation in one-dimensional geometries with one-dimensional attractor networks, where the rest displacement $\Delta\mathcal{X}^{\mathrm{A}}_{j\to i}$ is a scalar.

IX. CONNECTION TO EXPERIMENTS

We saw above that exploration in a simple 1D geometry lead to a consistent internal map in which the attractor network phase was mapped onto the current physical position alone, independent of path history (Fig. 4C). This consistency arises through the elastic relaxation process in Eq. 8(Main) and Eq. 9(Main), which makes the distance between the landmark cells in unrolled phase $\mathcal{X}^{\mathbf{L}}_{\mathrm{E}} - \mathcal{X}^{\mathbf{L}}_{\mathrm{W}}$ equal to the physical distance between their firing fields L, just like two particles connected by a spring with rest length L (Fig. 4E). Likewise, we have showed that navigation of an arbitrary environment will yield a "particles on springs" elastic relaxation process in 2D. While the 1D situation generalizes to two dimensions if there are only two landmarks, namely a west and east border cell (Fig. 6A1), yielding a rest length of $\mathrm{L}\hat{\mathbf{x}}$, adding more landmarks yields a more complex elastic relaxation process that we will build intuition about.

Consider the addition of a *south*-border landmark cell, in a 2D environment. How will the addition of this third landmark field affect the internal map?

In this case, east and west landmark particles will be connected by a spring of rest length $\Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{E}\to\mathrm{W}} = \mathrm{L}\hat{\mathbf{x}}$, as before, but they will each also be connected to the south landmark particle with springs. These springs have a rest length vector which is *smaller* than $\mathrm{L}\hat{\mathbf{x}}/2$, as contributions from the overlap between firing fields dominate the rest length. We may build some intuition about this process (See Sec. XI 2 C for more detail) by approximating:

$$\mathrm{M}_{\mathrm{WE}} = \mathrm{M}_{\mathrm{SE}} = \mathrm{M}_{\mathrm{SW}}, \qquad \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}} = \mathrm{L}\hat{\mathbf{x}}, \qquad \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{S}\to\mathrm{W}} = \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{E}\to\mathrm{S}} = 0,$$

which will yield a learned internal map of:

$$\boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{E}} = \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{S}} + \mathrm{L}\hat{\mathrm{x}}/3, \qquad \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{W}} = \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{S}} - \mathrm{L}\hat{\mathrm{x}}/3, \qquad \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{E}} - \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{W}} = 2\mathrm{L}\hat{\mathrm{x}}/3 < \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}.$$

Intuitively, as the animal travels from the east or west walls to the south walls, the landmark pinning phases of each of these three border cells will be attracted towards each other. In general, the combined three particle elastic system will settle into an equilibrium configuration in which the difference in unrolled phase between east and west landmarks will be *less* than the physical separation $\mathrm{L}\hat{\mathbf{x}}$, or equivalently the rest length $\Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{E}\to\mathrm{W}}$ of the spring connecting them. While we have presented a very simple case, we emphasize that more complex, non-overlapping distributions yield the same deformations.

This deformation and contraction of the internal map implies that the attractor phase assigned to any physical position in the interior will be relatively phase advanced (retarded) if the animal is on a trajectory leaving the west (east) wall. This path dependence in the attractor phase is entirely analogous to that seen in Fig. 4B. However, the reason is completely different. In Fig. 4B, the landmark particles are not separated by the rest length of the spring connecting them because the environment is not fully learned and so the particles are out of equilibrium, whereas in Fig. 6A2, the particles are not separated by the rest length, even in a *fully* learned environment, because additional springs from the south landmark create excess compression.

This theory makes a striking experimentally testable prediction, namely that even in a *fully learned* 2D environment, grid cell firing fields, when computed on subsets of animal trajectories conditioned on leaving a particular border, will be shifted *towards* that border (Fig. 6B). This shift occurs because at any given position, the attractor phase depends on the most recently encountered landmark. In particular, on a west to east (east to west) trajectory, the attractor phase will be advanced (retarded) relative to a east to west (west to east) trajectory. Thus on a west to east trajectory, the advanced phase will cause grid cells to fire earlier, yielding west shifted grid cell firing fields as a function of position. Similarly on an east to west trajectory, grid fields will be east shifted. In summary, the theory predicts grid cell firing patterns conditioned on trajectories leaving the west (east) border will be shifted west (east).

While we have derived this prediction qualitatively using the conceptual mass-spring picture in Fig. 6A2, we confirm

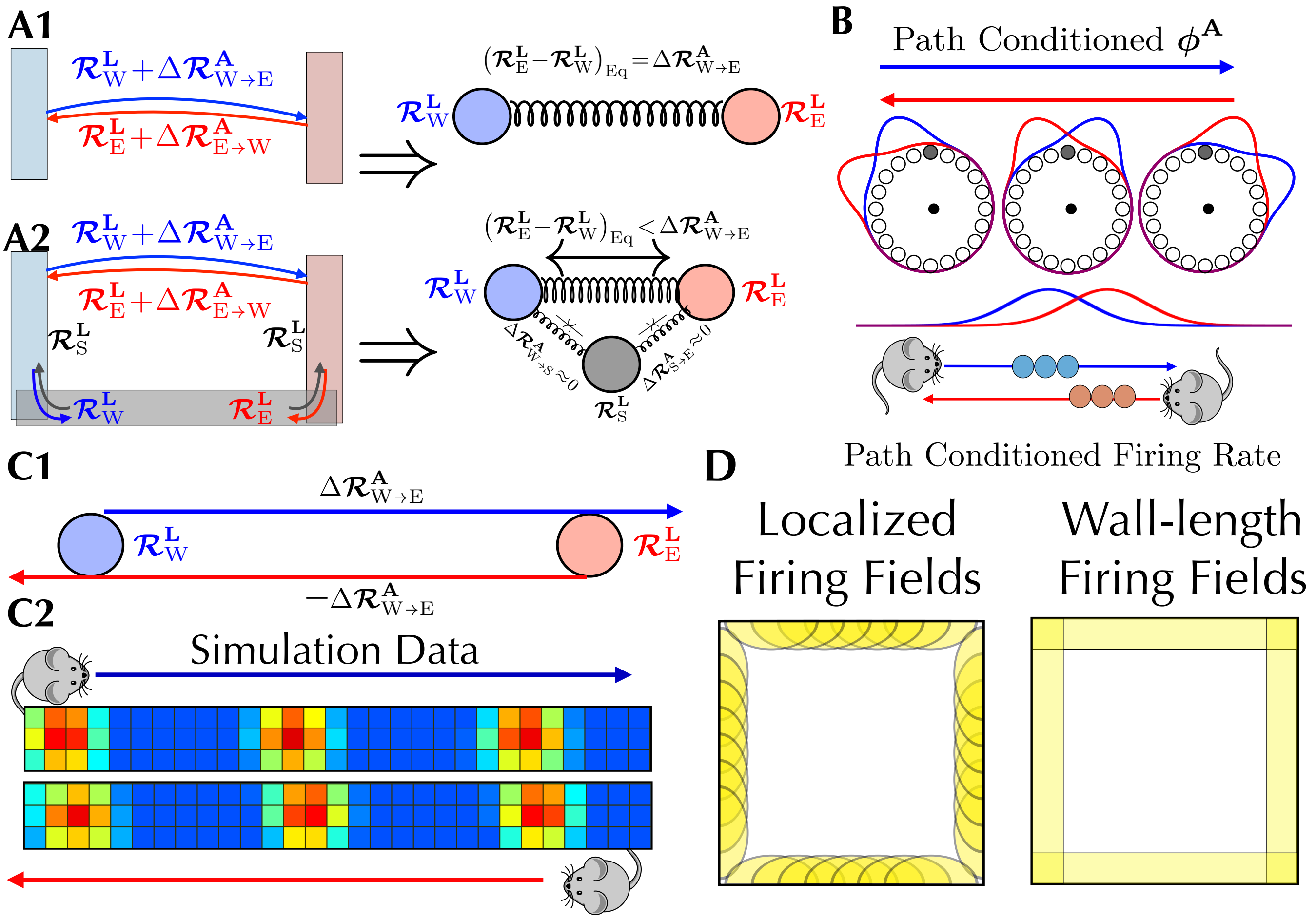
A1
$\mathcal{R}^{\mathbf{L}}_{\mathrm{W}}+\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{E}}+\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{E}\to\mathrm{W}}$
$(\mathcal{R}^{\mathbf{L}}_{\mathrm{E}}-\mathcal{R}^{\mathbf{L}}_{\mathrm{W}})_{\mathrm{Eq}}=\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{W}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{E}}$
A2
$\mathcal{R}^{\mathbf{L}}_{\mathrm{W}}+\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{E}}+\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{E}\to\mathrm{W}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{S}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{S}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{W}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{E}}$
$(\mathcal{R}^{\mathbf{L}}_{\mathrm{E}}-\mathcal{R}^{\mathbf{L}}_{\mathrm{W}})_{\mathrm{Eq}}<\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{W}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{E}}$
$\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{S}}\approx 0$
$\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{S}\to\mathrm{E}}\approx 0$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{S}}$
B
Path Conditioned $\phi^{\mathbf{A}}$
Path Conditioned Firing Rate
C1
$\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{W}}$
$\mathcal{R}^{\mathbf{L}}_{\mathrm{E}}$
$-\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}$
C2
Simulation Data
D
Localized
Firing Fields
Wall-length
Firing Fields

FIG. SI.10:

A1) For two landmark cells, the rest length $\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}$ of the spring connecting them equals the physical width L of the environment, and so the two landmark particles learn unrolled pinning phases $\mathcal{R}^{\mathbf{L}}_{\mathrm{E}}$ and $\mathcal{R}^{\mathbf{L}}_{\mathrm{W}}$ obeying the spatial consistency condition $(\mathcal{R}^{\mathbf{L}}_{\mathrm{E}}-\mathcal{R}^{\mathbf{L}}_{\mathrm{W}})_{\mathrm{Eq}}=\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}=\mathrm{L}\hat{\mathrm{x}}$ as in Fig. 4(Main) C. Blue and red arrows represent animal trajectories between the west and east walls, having equal and opposite path integration distance $\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}$, $\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{E}\to\mathrm{W}}$. **A2)** The addition of a southern landmark cell will cause a pinning effect which pulls $\mathcal{R}^{\mathbf{L}}_{\mathrm{W}},\mathcal{R}^{\mathbf{L}}_{\mathrm{E}}$ closer together. The animal can travel from the east and west landmark field to the southern landmark field with little path integration at all (blue/black and black/red arrow pairs), yielding $\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{S}}\approx 0$, $\Delta\mathcal{R}^{\mathbf{A}}_{\mathrm{S}\to\mathrm{E}}\approx 0$. **B)** If the attractor phase is advanced on a west to east trajectory (blue) relative to an east to west trajectory (red), then any particular grid cell (in this case the shaded grey cell) will fire earlier (later) on west-to-east (east-to-west) trajectory. Thus grid fields computed from trajectories leaving the west (east) border will shifted west (east). **C1)** When landmark pinning phases are pulled together closer than the path integration distance between them, then the attractor phase will shift *away* from whichever wall the animal last encountered. Therefore it will phase advance on west-to-east trajectories relative to east-to-west trajectories, as in Fig. 4B and Fig. 6B. **C2)** Thus simulations of Eq. 15 and Eq. 13 lead to grid cell firing patterns shifted *towards* whichever wall the animal last encountered. **D)** Schematic of the distribution of landmark cells for simulations of squares environments. To model a heterogeneous distribution of landmark cell degrees of localization, we include both landmark cells which fire uniformly along a boundary, as well as semi-elliptical landmark cells which are localized to a section of a boundary.

this intuition through direct numerical simulations of the full circuit dynamics in Eq. 15 and Eq. 13 (Fig. 6C2, D). Under reasonable parameters, our simulations can yield path-dependent shifts of up to ∼2 cm towards whichever wall the animal last touched (Sec. XII).

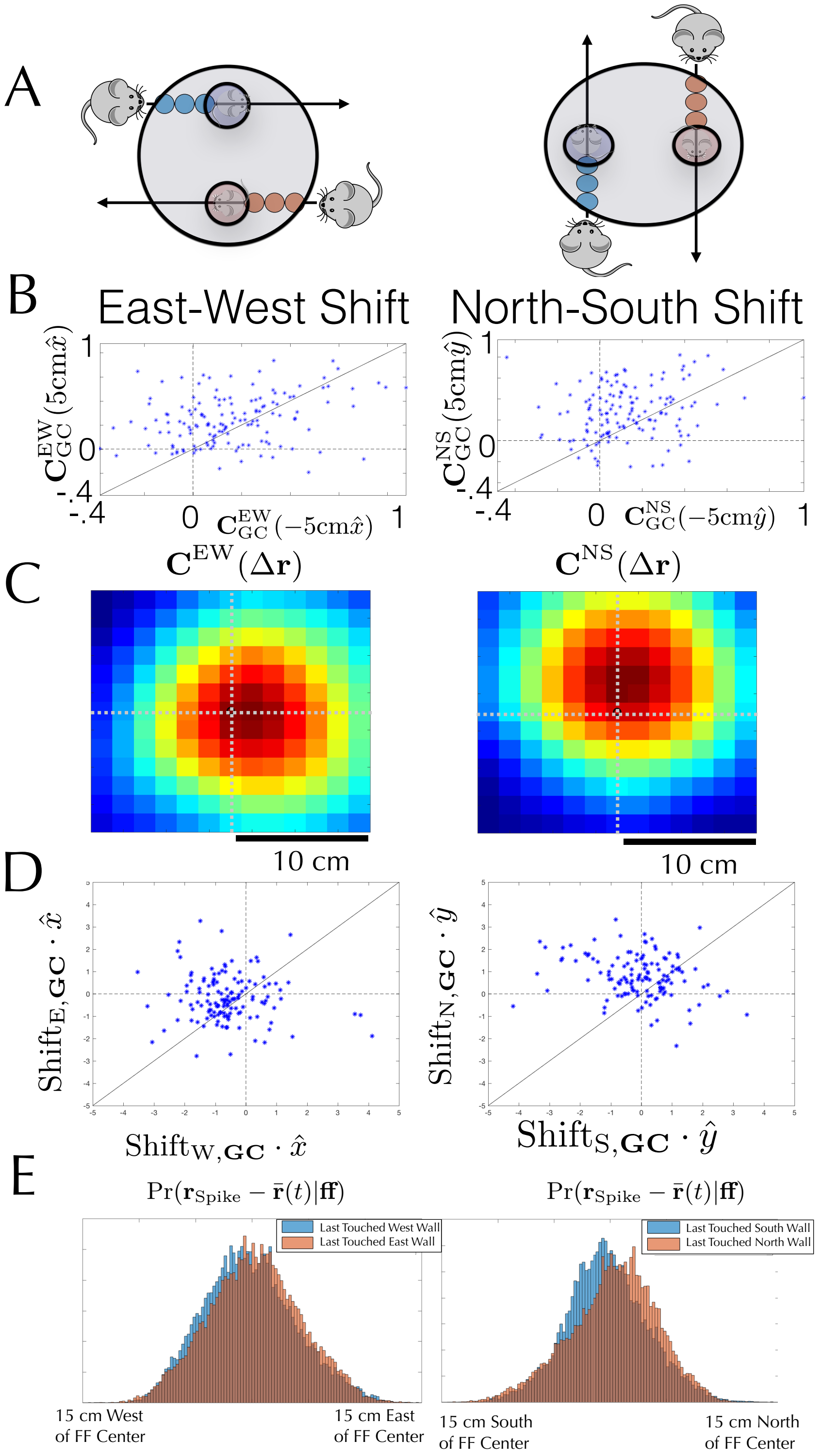

FIG. SI.11:

A) Schematic of observed shifts. An animal traveling from the east wall will have its firing patterns east-shifted; an animal traveling from the west wall will have its firing patterns west-shifted. Likewise, an animal traveling from the north (south) wall will have its firing patterns north (south) shifted. **B)** Path-dependent shifts demonstrated by Cross-Correlograms of individual grid cells. Most cells fall on the upper left of the plots, showing that the patterns tend to be shifted towards whichever wall the animal last touched for both the EW Walls ($P = 1.5 \cdot 10^{-5}$, Binomial Test, $P = 1.5 \cdot 10^{-5}$, Sign-Flip Test), and the NS walls ($P = 10^{-7}$, Binomial Test, $P = 10^{-7}$, Sign-Flip Test). **B)** The path-dependent shifts is best visualized through the Cross-Correlogram averaged over all grid cells. **C)** Path-dependent shifts demonstrated by Cross-Correlograms of individual grid cells. Most cells fall on the upper left of the plots, showing that the patterns are shifted towards whichever wall the animal last touched for both the EW Walls ($P = 3 \cdot 10^{-4}$ Binomial Test, $P = 2 \cdot 10^{-2}$ Sign-Flip Test), and the NS walls ($P = 10^{-5}$ Binomial Test, $P = 10^{-5}$ Sign-Flip Test). **D)** The path-dependent shifts is best visualized through a histogram of individual spike displacements.

1. Experimental observation of path-dependent shifts

We searched for such subtle shifts in a population of 143 grid cells from 14 different mice that had been exploring a familiar, well-learned, 1-meter open field (Sec. XII 2), using two separate analyses, based on cross-correlations and spike shifts with respect to field centers.

A. *Path conditioned rate maps*

One method for detecting a systematic firing field shift across many grid fields is to cross-correlate firing rate maps conditioned on trajectories leaving two different borders (Sec. IX 1 A). For example, for each cell, we can ask how much and in what direction we must shift its west border conditioned firing field to match, or correlate as much as possible with, the same cell's east conditioned firing field. We constructed maps of firing rate as a function of spatial position conditioned on the animal having last touched the north wall more recently than it touched the south wall, etc. An animal was defined to have "touched" a wall when the head-tracking diodes came within 10 cm of the wall (varying this distance did not significantly effect our results).

A type of cross-correlation was taken, using the cosine-angle between two path-conditioned rate maps.

$$\mathbf{C}^{\mathcal{C}1\mathcal{C}2}_{\mathrm{GC}}(\Delta\mathbf{r_C}) = \frac{\left|s^{\mathcal{C}1}_{\mathrm{GC}}(\mathbf{r}+\Delta\mathbf{r_C})s^{\mathcal{C}2}_{\mathrm{GC}}(\mathbf{r})\right|}{\left|s^{\mathcal{C}1}_{\mathrm{GC}}(\mathbf{r}+\Delta\mathbf{r_C})\right|\left|s^{\mathcal{C}2}_{\mathrm{GC}}(\mathbf{r})\right|},$$

where the mean firing rate is subtracted, and the inner product is only calculated using bins where there is data. To show significance, we calculate

$$\mathbf{C}^{\mathrm{EW}}_{\mathrm{GC}}(5\mathrm{cm}\hat{\mathbf{x}}) - \mathbf{C}^{\mathrm{EW}}_{\mathrm{GC}}(-5\mathrm{cm}\hat{\mathbf{x}}), \qquad \mathbf{C}^{\mathrm{NS}}_{\mathrm{GC}}(5\mathrm{cm}\hat{\mathbf{y}}) - \mathbf{C}^{\mathrm{NS}}_{\mathrm{GC}}(-5\mathrm{cm}\hat{\mathbf{y}})$$

And show that the patterns are shifted towards whichever wall the animal last touched for both the EW Walls ($\mathbf{C}^{\mathrm{EW}}_{\mathrm{GC}}(5\mathrm{cm}\ \hat{\mathbf{x}}) - \mathbf{C}^{\mathrm{EW}}_{\mathrm{GC}}(-5\mathrm{cm}\ \hat{\mathbf{x}}) > 0$, $\mathrm{P} = 1.5 \cdot 10^{-5}$, Binomial Test, $\mathrm{P} = 1.5 \cdot 10^{-5}$, Sign-Flip Test), and the NS walls ($\mathbf{C}^{\mathrm{NS}}_{\mathrm{GC}}(5\mathrm{cm}\ \hat{\mathbf{y}}) - \mathbf{C}^{\mathrm{NS}}_{\mathrm{GC}}(-5\mathrm{cm}\ \hat{\mathbf{y}})$, $\mathrm{P} = 10^{-7}$, Binomial Test, $\mathrm{P} = 10^{-7}$, Sign-Flip Test), in agreement with the theory.

Overall, this analysis shows that grid patterns are shifted towards the most recently encountered wall, both for the NS walls (3 cm, $\mathrm{P} = 1.5 \cdot 10^{-5}$, Binomial Test, $\mathrm{P} = 1.5 \cdot 10^{-5}$, Sign-Flip Test) and the EW Walls (1.5 cm, $\mathrm{P} = 10^{-7}$, Binomial Test, $\mathrm{P} = 10^{-7}$, Sign-Flip Test), matching the sign and magnitude seen in simulations. We avoid any sort of smoothing to prevent artifacts which might show up an experimental signature; as such, the bin size of the computed $s^{\mathcal{C}}_{\mathrm{GC}}(\mathbf{r})$ is 5cm× 5cm, and each individual trial leaves many bins for which $s^{\mathcal{C}}_{\mathrm{GC}}(\mathbf{r})$ is not defined; we create finer-grained cross-correlelograms with fewer undefined bins by choosing bin sizes of 5/3 cm, and smoothing in the manner of [24], but these maps are not used for showing statistical significance.

B. *Spike displacement*

Our results in path-conditioned rate maps can be corroborated by computing shifts in spikes relative to firing field centers, when conditioning spikes on the path history (Sec. IX 1 B). We used an adaptive smoothed rate map to identify firing fields [25]. Fields were detected as connected regions with a total area greater than 5 bins (~ 10 cm^2), where each bin had a firing rate above a threshold of binned firing rates for that rate map. For each firing field center, we gather spikes recorded in that neighborhood. Then, for each path condition $\mathcal{C}$ and each firing field center $\mathbf{r_{ff}}$, we calculate the average spike position $\mathbf{r}_{\mathrm{Spk}}$ within that firing field, and subtract the average *mouse* position $\mathbf{r}(t)$ within that firing field (See Sec. XII 2 B for explanation).

$$\mathbf{S}_{\mathcal{C},\mathrm{GC},\mathbf{ff}} = \langle \mathbf{r}_{\mathrm{Spk}} - \mathbf{r_{ff}} | \mathcal{C}, \mathbf{r}_{\mathrm{Spk}} \in \mathbf{ff}\ \rangle - \langle \mathbf{r}(t) - \mathbf{r_{ff}} | \mathcal{C}, \mathbf{r}(t) \in \mathbf{ff}\ \rangle \tag{SI.IX.52}$$

We calculate the path-dependent shift of an individual grid cell as the average shift of all firing fields in the center:

$$\mathbf{S}_{\mathcal{C},\mathrm{GC}} = \sum_{\mathbf{ff}} \mathbf{S}_{\mathcal{C},\mathrm{GC},\mathbf{ff}}$$

and examine how the shifts depend on which wall the animal last touched. To test for statistical significance, we calculate the relative shifts between path conditions for each cell:

$$\Delta\mathbf{S}_{\mathrm{EW,GC}} = (\mathbf{S}_{\mathrm{E,GC}} - \mathbf{S}_{\mathrm{W,GC}}) \cdot \hat{\mathbf{x}}, \qquad \Delta\mathbf{S}_{\mathrm{NS,GC}} = (\mathbf{S}_{\mathrm{N,GC}} - \mathbf{S}_{\mathrm{S,GC}}) \cdot \hat{\mathbf{y}}.$$

We test whether $\Delta\mathbf{S}_{\mathrm{EW,GC}}$, $\Delta\mathbf{S}_{\mathrm{NS,GC}}$ are significantly different from zero; for completeness, we perform both binomial tests, which only depend on the sign of $\Delta\mathbf{S}_{\mathrm{EW,GC}}$, $\Delta\mathbf{S}_{\mathrm{NS,GC}}$, as well as magnitude-weighted sign-flip tests.

Again, the patterns are shifted towards whichever wall the animal last touched (Fig. 7 C) for both the NS walls (.5 cm, $\mathrm{P} = 10^{-5}$ Binomial Test, $\mathrm{P} = 10^{-5}$ Sign-Flip Test) and the EW Walls (.5 cm, $\mathrm{P} = 3 \cdot 10^{-4}$ Binomial Test, $\mathrm{P} = 2 \cdot 10^{-2}$ Sign-Flip Test). The discrepancy in the estimated magnitude of the shift between the methods of analysis is likely due to poorly defined firing fields; a method based on firing field centers will give a lower signal-to-noise ratio, and thus a lower shift magnitude, than the cross-correlogram method.

2. Mechanical deformations in complex environments

Another experimental observation that can be reproduced by our theory is the distortion [26, 27] of grid cell patterns seen in an irregular environment (Fig. 8A). In our model, landmark cells with firing fields distributed across an entire wall will pull the attractor phase to its associated landmark pinning phases *regardless* of where along the wall the animal is. Here, we simulate Eq. SI.VII.42 , Eq. SI.VII.43 by discretizing the learning dynamics and using a random walk model for animal trajectories (Sec. XII). Experimental landmark cell firing fields associated with borders are heterogeneous; some are localized along a border, while others are distributed across an entire border. To replicate this distribution, we have two types of landmark cells in our model. (1) Landmark cells having uniform wall-length firing field, with a width of 10cm. (2) More localized, overlapping, firing fields along each wall.

The presence of a diagonal wall then causes the average attractor phase as a function of position to curve towards the wall, yielding spatial grid cell patterns that curve *away* from the wall (Fig. 8B, C). Previous theoretical accounts of this grid cell deformation have relied on purely phenomenological models that treated individual grid cell firing fields as particles with mostly repulsive interactions [28], without a clear mechanistic basis underlying this interaction. Here we provide, to our knowledge for the first time, a model with a clear mechanistic basis for such deformations, grounded in the interaction between attractor based path integration and landmark cells with plastic synapses. Such dynamics yields an emergent elasticity where the particles are landmark cell synapses rather than individual firing field centers.

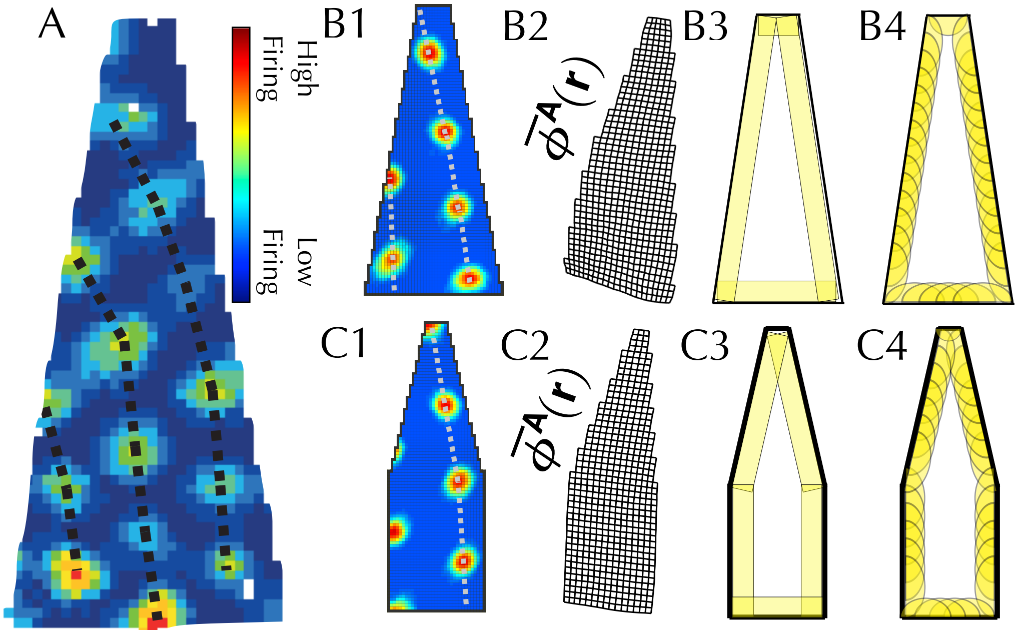

FIG. SI.12:

A) Experimental data of grid cell firing patterns deformed, curving *away* from a wall in an irregular geometry. **B1)** A full simulation of Eq. 15, Eq. 13 also yields grid firing patterns bent away from the wall. **B2)** Visualization of the average attractor state as a function of position (periodicity removed for visualization purposes). The reversal between the bending of the internal attractor phase and the bending of firing rate maps is similar to the reversal seen in Fig. 6B. **B3), B4)** Schematic of the distribution of landmark cells for simulations of trapezoidal environments. To model a heterogeneous distribution of landmark cell degrees of localization, we include both landmark cells which fire uniformly along a boundary, as well as semi-elliptical landmark cells which are localized to a section of a boundary. **C1-4)** Same as **B1-4)**, but for a slightly different geometry.

3. Topological defects in grid cells: a prediction

While the dynamics of the linearized Eq. SI.VII.48 will always flow to the same relative landmark representations $\boldsymbol{\mathcal{R}}_i^{\mathbf{L}}$, this is not the case for the full dynamics of Eq. SI.VII.42, Eq. SI.VII.43, which can learn multiple different stable landmark cell synaptic configurations. One striking example of this is the ability of the learning dynamics to generate "topological defects", where the number of firing fields traversed is not the same for two different paths (Fig. SI.13A, B and Sec. XII). An environmental geometry capable of supporting these defects will yield a set of firing patterns that depends not only on the final geometry, but also on the *history* of how this geometry was created (Fig. SI.13C).

Essential components to creation of topological defects are: (1) A "donut-shaped" environment, which can support the topological defect. (2) An environment rich in localized, strong landmark cues. (3) The larger the environment is, the less deformation it has to support per unit distance, i.e. if an environment is 3 firing fields wide, a topological defect must modify the grid spacing by 33%; if the environment was 5 firing fields wide, the grid spacing would only need to be modified by 20%. Therefore it is easier to create topological defects in a larger environment. (4) During the "winding" procedure, the animal cannot acclimate to the intermediate environment for too long; if the animal fully learns the intermediate environment, the winding procedure will not work (Fig. SI.13 E).

X. DETAILS OF THEORY

1. Path integration using a more realistic conjunctive model

In Eq. SI.II.2, we use a simplified model where we assume each velocity conjunctive cell at u has only *one* outgoing connection to $u \pm \Delta\phi_{\mathrm{PI}}$.

$$\frac{ds\left(u\right)}{dt} = -\frac{s(u)}{\tau_m} + \mathcal{G}\left(\int_{u'} \mathrm{J}(u-u')s(u')\right) + \underbrace{\epsilon_{\mathrm{PI}}\ s_{\mathrm{EC}}\left(u - \Delta\phi_{\mathrm{PI}}\right)}_{\text{Input from east cells}} + \underbrace{\epsilon_{\mathrm{PI}}\ s_{\mathrm{WC}}\left(u + \Delta\phi_{\mathrm{PI}}\right)}_{\text{Input from west cells}} + \sum_i \underbrace{\epsilon_{\mathrm{LM}}\left(\mathrm{W}_i(u)s_i^{\mathrm{L}}(t)\right)}_{\text{Landmark cell inputs}}$$

It might be more realistic to have a model where the outgoing conjunctive weights have the same form as that of the non-conjunctive cells and the landmark weights are also fed into the nonlinearity. Here, we show how to relax the assumption used in Eq. SI.II.2.

$$\frac{ds\left(u\right)}{dt} = -\frac{s(u)}{\tau_m} + \mathcal{G}\left(\int_{u'} \mathrm{J}(u-u') \times \left[s(u') + \underbrace{\epsilon_{\mathrm{PI}}\left(s_{\mathrm{EC}}\left(u' - \Delta\phi_{\mathrm{PI}}\right)\right)}_{\text{Input from east cells}} + \underbrace{\epsilon_{\mathrm{PI}} s_{\mathrm{WC}}\left(u' + \Delta\phi_{\mathrm{PI}}\right)}_{\text{Input from west cells}}\right] + \epsilon_{\mathrm{LM}} \sum_i \underbrace{\left(\mathrm{W}_i(u)s_i^{\mathrm{L}}(t)\right)}_{\text{Landmark cell inputs}}\right) \qquad \text{(SI.X.53)}$$

XI. EFFECT OF PERTURBATION SHAPE ON EFFECTIVE FORCE FUNCTION

Switching to the dynamics of Eq. SI.X.53 changes the shape of perturbations caused by landmark and velocity-conjuctive cells, such that we must use different perturbation functions for the force function and the landmark function. For example, when the attractor state is $s^*(u - \phi^{\mathrm{A}})$, the perturbation function for landmark cells will be:

$$\delta_s^{\mathrm{LM}}(u) = \mathcal{G}'\left(\int_{u'} \mathrm{J}(u-u')s^*(u' - \phi^{\mathrm{A}})\right)\epsilon_{\mathrm{LM}}\mathrm{W}(u)s_i^{\mathrm{L}}(t),$$

and the perturbation for east-conjunctive cells will be:

$$\delta_s^{\mathrm{East}}(u) = \epsilon_{\mathrm{PI}}\mathcal{G}'\left(\int_{u'} \mathrm{J}(u-u')s^*(u' - \phi^{\mathrm{A}})\right)\int_{u'} \mathrm{J}(u-u')s_{\mathrm{EC}}\left(u' - \Delta\phi_{\mathrm{PI}}\right),$$

and we can construct a similar perturbation for west-conjunctive cells.

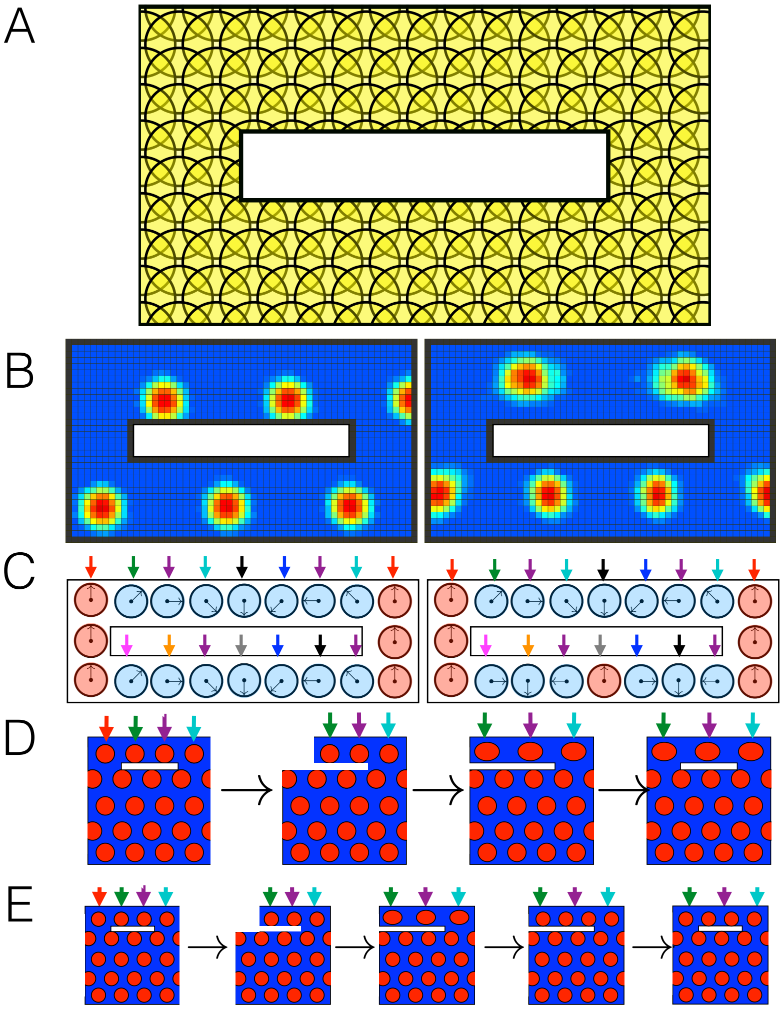

FIG. SI.13:

A) Schematic of the distribution of landmark cells for simulations the topological environment; cues are densely and uniformly localized throughout the arena. **B)** Two steady state grid cell patterns emerging from the same cue-rich environment. In the first firing pattern, the combination of landmark pinning and path integration yields a phase advance of four firing fields in traveling from west to east along either corridor. The second pattern has a topological defect; traveling from the west to east through the *north* corridor yields a phase increase of $\sim$ 1.5 firing fields; traveling east to west through the *south* corridor yields a phase decrease of $\sim$ 2.5 firing fields. This second pattern is stable nonetheless. **C)** Schematic of 1D underlying attractor state as a function of space. The two patterns in (A) correspond to two different landmark pinning phase patterns learned by the many landmarks. Both landmark pinning patterns are stable under Eq. 15, Eq. SI.VII.43. In the first pattern, the combination of landmark pinning and path integration yields the same phase advance in both the north and south corridors. The second pattern has a topological defect; the phase advance in the north corridor is *one full rotation less* than the phase advance through the south corridor. This is possible because many landmark cues (colored arrows) can yield many landmark cells with multiple stable synaptic configurations, or pinning phases under Eq. SI.VII.42, Eq. SI.VII.43. **D)** Schematic of proposed "deformation schedule" that could yield a topological defect in grid cell firing patterns. By separating/truncating the northern corridor, stretching it (along with spatial cues, denoted by colored arrows), then reconnecting it, it may be possible introduce one of these defects. Even though the initial geometry is identical to the final geometry, the deformation schedule has lead to a firing pattern which is three fields wide in the north and four fields wide in the south. **E)** Example of topological defect failing to form due to learning. If the winding procedure is done too slowly, the animal will learn the deformed geometry (Third box → Fourth Box), removing the topological effect.

Because the perturbations from path integration and landmarks have different functional forms, the force laws that we calculate (Sec. III 2) will have different functional forms. Before, when there was a single functional for perturbations, we were able to map:

$$\delta_s \to \mathcal{F}(\phi^{\mathrm{P}} - \phi^{\mathrm{A}}).$$

Now we must do this separately for landmark and self-motion input, where:

$$\delta_s^{\mathrm{East}} \to \mathcal{F}^{\mathrm{E}}(\phi^{\mathrm{P}} - \phi^{\mathrm{A}}), \qquad \delta_s^{\mathrm{LM}} \to \mathcal{F}^{\mathrm{LM}}(\phi^{\mathrm{P}} - \phi^{\mathrm{A}}).$$

Once this complication is taken into account, the rest of the calculation can proceed verbatim.

1. Verifying the Hebbian learning rule in the attractor basis

We can verify that in the attractor basis:

$$\mathrm{W}_i(u) = \int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}}) s^{*}(u - \phi^{\mathrm{L}}) \tag{SI.XI.54}$$

the learning rule of Eq. SI.III.26:

$$\frac{d\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})}{d\mathrm{T}} = \Pr(\phi^{\mathrm{L}}|i \text{ Firing}) - \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})$$

gives us the learning rule in the neural basis (Eq. SI.III.25):

$$\frac{d\mathrm{W}_i(u)}{d\mathrm{T}} = \langle s(u)|i \text{ Firing}\rangle - \mathrm{W}_i(u) = \int_{\phi} s^{*}(u - \phi^{\mathrm{L}}) \Pr(\phi^{\mathrm{L}}|i \text{ Firing}) - \mathrm{W}_i(u)$$

by inspection:

$$\begin{aligned} \frac{d\mathrm{W}_i(u)}{d\mathrm{T}} \underbrace{=}_{\text{Basis Switch}} \frac{d}{d\mathrm{T}}\left[\int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}}) s^{*}(u - \phi)\right] = \int_{\phi^{\mathrm{L}}} \frac{d\tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})}{d\mathrm{T}} s^{*}(u - \phi^{\mathrm{L}}) \\ \underbrace{=}_{\text{Eq.SI.III.26}} \int_{\phi^{\mathrm{L}}} \left[\Pr(\phi^{\mathrm{L}}|i \text{ Firing}) - \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}})\right] s^{*}(u - \phi^{\mathrm{L}}) = \int_{\phi^{\mathrm{L}}} \Pr(\phi^{\mathrm{L}}|i \text{ Firing}) s^{*}(u - \phi) - \underbrace{\int_{\phi^{\mathrm{L}}} \tilde{\mathrm{W}}_i(\phi^{\mathrm{L}}) s^{*}(u - \phi^{\mathrm{L}})}_{\mathrm{W}(u)} \\ = \underbrace{\int_{\phi^{\mathrm{L}}} s^{*}(u - \phi^{\mathrm{L}}) \Pr(\phi^{\mathrm{L}}|i \text{ Firing}) - \mathrm{W}_i(u)}_{\text{Eq.SI.III.25}}. \end{aligned}$$

2. Detailed calculations for learning a simple environmental geometry

This section is a more detailed version of Sec. V containing full calculations for the orientation and learning dynamics given arbitrary landmark strength (ω), animal running speed (v_0), and landmark field width ($\mathrm{L}_{\mathrm{Wall}}$). First, we solve for the steady-state position self-estimate given a set of landmarks, and use that to solve for equilibrium internal map. We then examine the path-dependent shifts that come from an unlearned environment, as well as the effects landmark strength and animal speed on the learning rate.

The position self-estimate will reach a steady cycle, so we start with the animal at $x(t = 0) = -\mathrm{L}/2$, having position self-estimate $\mathcal{X}_0^{\mathrm{A}}$. The position self-estimate will follow the linearized dynamics, which include terms for path integration as well as the east and west landmarks.

$$\begin{gathered} \frac{d\mathcal{X}^{\mathrm{A}}}{dt} = \frac{dx}{dt} \quad + \omega \mathrm{H}_{\mathrm{E}}(x) \left(\mathcal{X}_{\mathrm{E}}^{\mathrm{L}} - \mathcal{X}^{\mathrm{A}}\right) + \omega \mathrm{H}_{\mathrm{W}}(x) \left(\mathcal{X}_{\mathrm{W}}^{\mathrm{L}} - \mathcal{X}^{\mathrm{A}}\right) \\ \mathrm{H}_{\mathrm{E}}(x) = \left[x - (\mathrm{L}/2 - \mathrm{L}_{\mathrm{Wall}})\right]_{+}, \qquad \mathrm{H}_{\mathrm{W}}(x) = \left[(-\mathrm{L}/2 + \mathrm{L}_{\mathrm{Wall}}) - x\right]_{+} \end{gathered}$$

We can assume the position self-estimate will reach a steady cycle such that $\mathcal{X}^{\mathrm{A}}(t = 2\tau) = \mathcal{X}^{\mathrm{A}}(t = 0)$. Defining $\mathcal{X}_1^{\mathrm{A}} = \mathcal{X}^{\mathrm{A}}(\tau/2)$, $\mathcal{X}_2^{\mathrm{A}} = \mathcal{X}^{\mathrm{A}}(\tau)$, $\mathcal{X}_3^{\mathrm{A}} = \mathcal{X}^{\mathrm{A}}(3\tau/2)$, we can solve for the position self-estimate as a piecewise function:

$$\mathcal{X}^{\mathrm{A}}(t) = \begin{cases} \mathcal{X}_1^{\mathrm{A}} + v_0 t & \tau_{\mathrm{Wall}} < t < \tau - \tau_{\mathrm{Wall}} \\ \left(\mathcal{X}_1^{\mathrm{A}} + [\mathrm{L}_{\mathrm{Int}}/2]\right) e^{-\omega(t-\tau_{\mathrm{Wall}})} + \left[\mathcal{X}_{\mathrm{E}}^{\mathrm{L}} + \frac{v_0}{\omega}\right](1 - e^{-\omega t}) & \tau - \tau_{\mathrm{Wall}} \leq t \leq \tau \\ \mathcal{X}_2^{\mathrm{A}} e^{-\omega(t-\tau_{\mathrm{Wall}})} + \left[\mathcal{X}_{\mathrm{E}}^{\mathrm{L}} - \frac{v_0}{\omega}\right](1 - e^{-\omega t}) & \tau < t \leq \tau + \tau_{\mathrm{Wall}} \\ \dots \end{cases} \tag{SI.XI.55}$$

Where $\tau_{\mathrm{Wall}} = \mathrm{L}_{\mathrm{Wall}}/v_0$. This yields a set of linear equations:

$$\begin{aligned} \mathcal{X}_1^{\mathrm{A}} &= e^{-\omega\tau_{\mathrm{Wall}}}\mathcal{X}_0^{\mathrm{A}} + \left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)\left[\mathcal{X}_{\mathrm{W}}^{\mathrm{L}} + \frac{v_0}{\omega}\right] + [\mathrm{L}_{\mathrm{Int}}/2] \\ \mathcal{X}_2^{\mathrm{A}} &= e^{-\omega\tau_{\mathrm{Wall}}}\left(\mathcal{X}_1^{\mathrm{A}} + [\mathrm{L}_{\mathrm{Int}}/2]\right) + \left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)\left[\mathcal{X}_{\mathrm{E}}^{\mathrm{L}} + \frac{v_0}{\omega}\right] \\ \mathcal{X}_3^{\mathrm{A}} &= e^{-\omega\tau_{\mathrm{Wall}}}\mathcal{X}_2^{\mathrm{A}} + \left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)\left[\mathcal{X}_{\mathrm{E}}^{\mathrm{L}} - \frac{v_0}{\omega}\right] - [\mathrm{L}_{\mathrm{Int}}/2] \\ \mathcal{X}_4^{\mathrm{A}} = \mathcal{X}_0^{\mathrm{A}} &= e^{-\omega\tau_{\mathrm{Wall}}}\left(\mathcal{X}_3^{\mathrm{A}} - [\mathrm{L}_{\mathrm{Int}}/2]\right) + \left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)\left[\mathcal{X}_{\mathrm{W}}^{\mathrm{L}} - \frac{v_0}{\omega}\right]. \end{aligned}$$

The average position self-estimate seen by the east landmark comes from two components of piecewise function $\mathcal{X}^{\mathrm{A}}(t)$. The first is $\tau - \tau_{\mathrm{Wall}} < t < \tau$, the second is $\tau < t < \tau + \tau_{\mathrm{Wall}}$:

$$\bar{\mathcal{X}}_{\mathrm{E}}^{\mathrm{A}} = \Big\langle \mathcal{X}^{\mathrm{A}}(t) | \text{East Landmark Cell Firing} \Big\rangle = \frac{\int_0^{2\tau} \mathrm{H}_{\mathrm{E}}(x(t))\mathcal{X}^{\mathrm{A}}(t)}{\int_0^{2\tau} \mathrm{H}_{\mathrm{E}}(x(t))} = \frac{1}{2\tau_{\mathrm{Wall}}} \int_{\tau-\tau_{\mathrm{Wall}}}^{\tau+\tau_{\mathrm{Wall}}} \mathcal{X}^{\mathrm{A}}(t) =$$

$$\bar{\mathcal{X}}_{\mathrm{E}}^{\mathrm{A}} = \frac{1}{2\tau_{\mathrm{Wall}}} \times \left[\left(\int_0^{\tau_{\mathrm{Wall}}} \underbrace{\left(\mathcal{X}_1^{\mathrm{A}} + [\mathrm{L}_{\mathrm{Int}}/2]\right) e^{-\omega t} + \left[\mathcal{X}_{\mathrm{E}}^{\mathrm{L}} + v_0\right](1 - e^{-\omega t})}_{\tau-\tau_{\mathrm{Wall}}<t<\tau}\right) + \left(\int_0^{\tau_{\mathrm{Wall}}} \underbrace{\mathcal{X}_2^{\mathrm{A}} e^{-\omega t} + \left[\mathcal{X}_{\mathrm{E}}^{\mathrm{L}} - v_0\right](1 - e^{-\omega t})}_{\tau<t<\tau+\tau_{\mathrm{Wall}}}\right)\right] =$$

$$\frac{1}{2\tau_{\mathrm{Wall}}} \times \left[\left(\mathcal{X}_1^{\mathrm{A}} + [\mathrm{L}_{\mathrm{Int}}/2]\right)\left(\frac{1 - e^{-\omega\tau_{\mathrm{Wall}}}}{\omega}\right) + \left[\mathcal{X}_{\mathrm{E}}^{\mathrm{L}} + \overbrace{\cancel{v_0}}^{\text{Cancels}}\right]\left(\tau_{\mathrm{Wall}} - \frac{1 - e^{-\omega\tau_{\mathrm{Wall}}}}{\omega}\right)\right] +$$
$$\frac{1}{2\tau_{\mathrm{Wall}}}\left[\mathcal{X}_2^{\mathrm{A}}\left(\frac{1 - e^{-\omega\tau_{\mathrm{Wall}}}}{\omega}\right) + \left[\mathcal{X}_{\mathrm{E}}^{\mathrm{L}} - \overbrace{\cancel{v_0}}^{\text{Cancels}}\right]\left(\tau_{\mathrm{Wall}} - \left(\frac{1 - e^{-\omega\tau_{\mathrm{Wall}}}}{\omega}\right)\right)\right]$$

$$= \frac{1}{2\tau_{\mathrm{Wall}}} \times \left[\left(\mathcal{X}_1^{\mathrm{A}} + [\mathrm{L}_{\mathrm{Int}}/2]\right)\left(\frac{1 - e^{-\omega\tau_{\mathrm{Wall}}}}{\omega}\right) + \mathcal{X}_{\mathrm{E}}^{\mathrm{L}}\left(\tau_{\mathrm{Wall}} - \frac{1 - e^{-\omega\tau_{\mathrm{Wall}}}}{\omega}\right) + \mathcal{X}_2^{\mathrm{A}}\left(\frac{1 - e^{-\omega\tau_{\mathrm{Wall}}}}{\omega}\right) + \mathcal{X}_{\mathrm{E}}^{\mathrm{L}}\left[\tau_{\mathrm{Wall}} - \frac{1 - e^{-\omega\tau_{\mathrm{Wall}}}}{\omega}\right]\right]$$
$$= \mathcal{X}_{\mathrm{E}}^{\mathrm{L}} + \frac{\left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)}{2\omega\tau_{\mathrm{Wall}}}\left[\left(\mathcal{X}_1^{\mathrm{A}} + [\mathrm{L}_{\mathrm{Int}}/2] - \mathcal{X}_{\mathrm{E}}^{\mathrm{L}}\right) + \left(\mathcal{X}_2^{\mathrm{A}} - \mathcal{X}_{\mathrm{E}}^{\mathrm{L}}\right)\right]$$

Therefore, at equilibrium:

$$\mathcal{X}_{\mathrm{E}}^{\mathrm{L}} = \bar{\mathcal{X}}_{\mathrm{E}}^{\mathrm{A}} = \frac{\left(\mathcal{X}_1^{\mathrm{A}} + [\mathrm{L}_{\mathrm{Int}}/2]\right) + \mathcal{X}_2^{\mathrm{A}}}{2}$$

There is a translational symmetry to this problem, such that any shifted version of a solution is also a solution. We center around zero for simplicity, such that $\mathcal{X}_1^{\mathrm{A}} = -\mathcal{X}_3^{\mathrm{A}}$, $\mathcal{X}_2^{\mathrm{A}} = -\mathcal{X}_4^{\mathrm{A}}$, and $\mathcal{X}_{\mathrm{E}}^{\mathrm{L}} = -\mathcal{X}_{\mathrm{W}}^{\mathrm{L}}$. Combining the above equations and this symmetry gives the steady state solution:

$$\begin{aligned} \mathcal{X}_1^{\mathrm{A}} &= \mathcal{X}_3^{\mathrm{A}} = 0, \\ \mathcal{X}_2^{\mathrm{A}} &= \left([\mathrm{L}_{\mathrm{Int}}/2] + \frac{2[1 - e^{-\omega\tau_{\mathrm{Wall}}}]}{1 + e^{-\omega\tau_{\mathrm{Wall}}}}\left(\frac{v_0}{\omega}\right)\right) && = \left([\mathrm{L}_{\mathrm{Int}}/2] + 2\tanh\left(\omega\tau_{\mathrm{Wall}}/2\right)\left(\frac{v_0}{\omega}\right)\right) = -\mathcal{X}_0^{\mathrm{A}} \\ \mathcal{X}_{\mathrm{E}}^{\mathrm{L}} &= \left([\mathrm{L}_{\mathrm{Int}}/2] + \frac{[1 - e^{-\omega\tau_{\mathrm{Wall}}}]}{1 + e^{-\omega\tau_{\mathrm{Wall}}}}\left(\frac{v_0}{\omega}\right)\right) && = \left([\mathrm{L}_{\mathrm{Int}}/2] + \tanh\left(\omega\tau_{\mathrm{Wall}}/2\right)\left(\frac{v_0}{\omega}\right)\right) = -\mathcal{X}_{\mathrm{W}}^{\mathrm{L}}. \end{aligned}$$

A. *Out of equilibrium path-dependent shifts and learning dynamics*

When the system is out of equilibrium it is convenient to refer to the landmark representations in terms of their deviation from the equilibrium state.

$$\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} = \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \left(\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\right)_{\mathrm{Eq}}, \qquad \mathcal{X}^{\mathrm{L}}_{\mathrm{W}} = \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}} + \left(\mathcal{X}^{\mathrm{L}}_{\mathrm{W}}\right)_{\mathrm{Eq}}$$

We have the set of linear equations for how much the position self-estimates vary with the landmark position estimates, where we use the shorthand $\mathcal{X}^{\mathrm{A}}_{i} = \Delta\mathcal{X}^{\mathrm{A}}_{i} + \left(\mathcal{X}^{\mathrm{A}}_{i}\right)_{\mathrm{Eq.}}$:

$$\begin{aligned}
\Delta\mathcal{X}^{\mathrm{A}}_{1} &= e^{-\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{A}}_{0} + \left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}} \\
\Delta\mathcal{X}^{\mathrm{A}}_{2} &= e^{-\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{A}}_{1} + \left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} \\
\Delta\mathcal{X}^{\mathrm{A}}_{3} &= e^{-\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{A}}_{2} + \left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} \\
\Delta\mathcal{X}^{\mathrm{A}}_{4} = \Delta\mathcal{X}^{\mathrm{A}}_{0} &= e^{-\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{A}}_{3} + \left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}}
\end{aligned}$$

We can combine the equations for $\Delta\mathcal{X}^{\mathrm{A}}_{3}, \Delta\mathcal{X}^{\mathrm{A}}_{2}$ to get:

$$\begin{aligned}
\Delta\mathcal{X}^{\mathrm{A}}_{3} &= e^{-\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{A}}_{2} + \left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} \\
&= e^{-\omega\tau_{\mathrm{Wall}}}\left[e^{-\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{A}}_{1} + \left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\right] + \left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} \\
&= e^{-2\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{A}}_{1} + \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\left(1 - e^{-2\omega\tau_{\mathrm{Wall}}}\right)
\end{aligned}$$

We can express through $\mathcal{X}^{\mathrm{A}}_{1}$ in terms of $\mathcal{X}^{\mathrm{A}}_{3}$ through symmetry:

$$\begin{aligned}
\Delta\mathcal{X}^{\mathrm{A}}_{3} &= e^{-2\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{A}}_{1} + \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\left(1 - e^{-2\omega\tau_{\mathrm{Wall}}}\right) \\
\Delta\mathcal{X}^{\mathrm{A}}_{1} &= e^{-2\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{A}}_{3} + \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}}\left(1 - e^{-2\omega\tau_{\mathrm{Wall}}}\right)
\end{aligned}$$

Plugging one into the other:

$$\begin{aligned}
\Delta\mathcal{X}^{\mathrm{A}}_{1} &= e^{-2\omega\tau_{\mathrm{Wall}}}\left[e^{-2\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{A}}_{1} + \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\left(1 - e^{-2\omega\tau_{\mathrm{Wall}}}\right)\right] + \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}}\left(1 - e^{-2\omega\tau_{\mathrm{Wall}}}\right) && \Rightarrow \\
\Delta\mathcal{X}^{\mathrm{A}}_{1} &= \left(e^{-4\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{A}}_{1} + e^{-2\omega\tau_{\mathrm{Wall}}}\left(1 - e^{-2\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}}\left(1 - e^{-2\omega\tau_{\mathrm{Wall}}}\right) && \Rightarrow \\
\left(1 - e^{-4\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{A}}_{1} &= e^{-2\omega\tau_{\mathrm{Wall}}}\left(1 - e^{-2\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}}\left(1 - e^{-2\omega\tau_{\mathrm{Wall}}}\right) && \Rightarrow \\
\left(1 - e^{-4\omega\tau_{\mathrm{Wall}}}\right)\Delta\mathcal{X}^{\mathrm{A}}_{1} &= \left(1 - e^{-2\omega\tau_{\mathrm{Wall}}}\right)\left[e^{-2\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}}\right]
\end{aligned}$$

This yields the change in position self-estimate:

$$\Delta\mathcal{X}^{\mathrm{A}}_{1} = \frac{\left[e^{-2\omega\tau_{\mathrm{Wall}}}\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}}\right]}{1 + e^{-2\omega\tau_{\mathrm{Wall}}}} = \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \frac{\left(\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}} - \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\right)}{1 + e^{-2\omega\tau_{\mathrm{Wall}}}}$$

This allows us to recover the first coefficient related to path-dependent shift. When $\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} = -\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}}$,

$$\Delta\mathcal{X}^{\mathrm{A}}_{1} = \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \tfrac{\left(\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}} - \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\right)}{1 + e^{-2\omega\tau_{\mathrm{Wall}}}} = \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\left[\tfrac{1 + e^{-2\omega\tau_{\mathrm{Wall}}} - 2}{1 + e^{-2\omega\tau_{\mathrm{Wall}}}}\right] = -\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\left[\tfrac{1 - e^{-2\omega\tau_{\mathrm{Wall}}}}{1 + e^{-2\omega\tau_{\mathrm{Wall}}}}\right] = -\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\ \tanh\left(\omega\tau_{\mathrm{Wall}}\right). \qquad \text{(SI.XI.56)}$$

We note that when $\omega\tau_{\mathrm{Wall}} \to \infty$, the path-dependent shift is *exactly* the shift in the estimated position of the landmark last touched $\mathcal{X}^{\mathrm{L}}_{\mathrm{W}}$. When $\omega\tau_{\mathrm{Wall}} \to 0$, the shift goes to 0, as the memory of $\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}$ is nearly the same as that of $\mathcal{X}^{\mathrm{L}}_{\mathrm{W}}$.

B. *Learning timescale coefficient*

In order to understand the learning dynamics, we must calculate the effect of the *estimated* landmark position on the *estimated* position self-estimate that becomes associated with each landmark:

$$\Delta\bar{\mathcal{X}}^{\mathrm{A}}_{\mathrm{E}} = \mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \left[\frac{\Delta\mathcal{X}^{\mathrm{A}}_{1} + \Delta\mathcal{X}^{\mathrm{A}}_{2}}{2} - \mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\right]\frac{\left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)}{\omega\tau_{\mathrm{Wall}}}.$$

Plugging in:

$$\Delta\mathcal{X}^{\mathrm{A}}_{2} = \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \left(\Delta\mathcal{X}^{\mathrm{A}}_{1} - \mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\right)e^{-\omega\tau_{\mathrm{Wall}}}$$

gives $\Delta\bar{\mathcal{X}}^{\mathrm{A}}_{\mathrm{E}}$ in terms of $\mathcal{X}^{\mathrm{A}}_{1}$:

$$\Delta\bar{\mathcal{X}}^{\mathrm{A}}_{\mathrm{E}} = \mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \left(\mathcal{X}^{\mathrm{A}}_{1} - \mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\right)\left[\frac{1 + e^{-\omega\tau_{\mathrm{Wall}}}}{2}\right]\left[\frac{\left(1 - e^{-\omega\tau_{\mathrm{Wall}}}\right)}{\omega\tau_{\mathrm{Wall}}}\right] = \mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \left(\mathcal{X}^{\mathrm{A}}_{1} - \mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\right)\left[\frac{1 - e^{-2\omega\tau_{\mathrm{Wall}}}}{2\omega\tau_{\mathrm{Wall}}}\right].$$

Plugging in the value of $\mathcal{X}^{\mathrm{A}}_{1}$:

$$\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \frac{\left(\Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{W}} - \Delta\mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\right)}{1 + e^{-2\omega\tau_{\mathrm{Wall}}}}$$

gives:

$$\Delta\bar{\mathcal{X}}^{\mathrm{A}}_{\mathrm{E}} = \mathcal{X}^{\mathrm{L}}_{\mathrm{E}} + \left(\mathcal{X}^{\mathrm{L}}_{\mathrm{W}} - \mathcal{X}^{\mathrm{L}}_{\mathrm{E}}\right)\left[\frac{1 - e^{-\omega\tau_{\mathrm{Wall}}}}{2\omega\tau_{\mathrm{Wall}}\left(1 + e^{-2\omega\tau_{\mathrm{Wall}}}\right)}\right]$$

yielding a learning time of:

$$\mathrm{T}_{\mathrm{Learning}} = \frac{2\omega\tau_{\mathrm{Wall}}\left(1 + e^{-2\omega\tau_{\mathrm{Wall}}}\right)}{1 - e^{-\omega\tau_{\mathrm{Wall}}}}. \qquad \text{(SI.XI.57)}$$

From Eq. SI.XI.56, we can see that as the landmark cells become stronger, the shifts become stronger, as the animals position self-estimate becomes more heavily weighted toward whichever landmark it most recently saw. From Eq. SI.XI.57 we see that, as landmark cells become stronger, the learning rate slows down, as landmark cells mostly see their own self-estimates; the contribution to position self-estimate from spatially disjoint landmarks decays quickly after the animal moves into the landmark firing field.

C. *Simple case of internal map contraction*

The learning rule for landmark landmark position estimates will be

$$\frac{d}{d\mathrm{T}}\begin{pmatrix} \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{E}} \\ \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{W}} \\ \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{S}} \end{pmatrix} = \left(\begin{array}{ccc} & \mathrm{M}_{\mathrm{WE}}\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{W}} + \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}\right) & +\mathrm{M}_{\mathrm{SE}}\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{S}} + \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{S}\to\mathrm{E}}\right) \\ \mathrm{M}_{\mathrm{EW}}\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{E}} - \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}\right) & & +\mathrm{M}_{\mathrm{SW}}\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{S}} + \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{S}\to\mathrm{W}}\right) \\ \mathrm{M}_{\mathrm{ES}}\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{E}} - \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{E}\to\mathrm{S}}\right) & +\mathrm{M}_{\mathrm{WS}}\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{W}} - \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{S}}\right) & \end{array}\right) \qquad \text{(SI.XI.58)}$$

For simplicity, we approximate:

$$\mathrm{M}_{\mathrm{WE}} = \mathrm{M}_{\mathrm{SE}} = \mathrm{M}_{\mathrm{SW}}, \qquad \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}} = \mathrm{L}\hat{\mathbf{x}}, \qquad \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{S}\to\mathrm{W}} = \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{E}\to\mathrm{S}} = 0$$

and solve for the equilibrium state Eq. SI.XI.58 to find the learned landmark position estimates:

$$\boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{E}} = \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{S}} + \hat{\mathbf{x}}\mathrm{L}/3, \qquad \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{W}} = \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{S}} - \hat{\mathbf{x}}\mathrm{L}/3, \qquad \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{E}} - \boldsymbol{\mathcal{R}}^{\mathbf{L}}_{\mathrm{W}} = 2\mathrm{L}\hat{\mathbf{x}}/3 < \Delta\boldsymbol{\mathcal{R}}^{\mathbf{A}}_{\mathrm{W}\to\mathrm{E}}.$$

3. Proof of convolutional integral for position self-estimate as a functional of path history

We can check that the solution for the position self-estimate Eq. SI.VIII.51:

$$\boldsymbol{\mathcal{R}}^{\mathbf{A}}[\mathbf{r}(t), t] = \int_{-\infty}^{t} \left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}\left(\mathbf{r}(t')\right) + \left(\mathbf{r}(t) - \mathbf{r}(t')\right)\right]\omega(\mathbf{r}(t'))\left[e^{-\int_{t'}^{t}\omega(\mathbf{r}(t''))dt''}\right]dt'$$

satisfies the dynamics of Eq. SI.VIII.50 :

$$\frac{d\boldsymbol{\mathcal{R}}^{\mathbf{A}}(t)}{dt} = \frac{d\mathbf{r}(t)}{dt} + \omega(\mathbf{r})\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}(t)) - \boldsymbol{\mathcal{R}}^{\mathbf{A}}\right]$$

by inspection. We plug Eq. SI.VIII.51 into Eq. SI.VIII.50 to get:

$$\begin{aligned}\frac{d\boldsymbol{\mathcal{R}}^{\mathbf{A}}}{dt} = {} & \underbrace{\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}\left(\mathbf{r}(t)\right) + \left(\mathbf{r}(t) - \mathbf{r}(t)\right)\right]\left[\omega(\mathbf{r}(t))e^{-\int_{t}^{t}\omega(\mathbf{r}(t''))dt''}\right]}_{\omega\boldsymbol{\mathcal{R}}^{\mathbf{L}}} + \underbrace{\int_{-\infty}^{t}\frac{d\mathbf{r}(t)}{dt}\omega(\mathbf{r}(t'))e^{-\int_{t'}^{t}\omega(\mathbf{r}(t''))dt''}dt'}_{d\mathbf{r}/dt} \\ & + \underbrace{\int_{-\infty}^{t}\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}\left(\mathbf{r}(t')\right) + \left(\mathbf{r}(t) - \mathbf{r}(t')\right)\right] \times \left[-\omega(\mathbf{r}(t'))\omega(\mathbf{r}(t))e^{-\int_{t'}^{t}\omega(\mathbf{r}(t''))dt''}\right]dt'}_{-\omega\boldsymbol{\mathcal{R}}^{\mathbf{A}}}\end{aligned}$$

The underbraced identities are more easily seen by simplifying terms:

$$\begin{aligned}\frac{d}{dt}\boldsymbol{\mathcal{R}}^{\mathbf{A}} = {} & \underbrace{\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}\left(\mathbf{r}(t)\right)\overbrace{\cancel{+\left(\mathbf{r}(t) - \mathbf{r}(t)\right)}}^{0}\right]\left[\omega(\mathbf{r}(t))\overbrace{\cancel{e^{-\int_{t}^{t}\omega(\mathbf{r}(t''))dt''}}}^{1}\right]}_{\omega\boldsymbol{\mathcal{R}}^{\mathbf{L}}} + \underbrace{d\mathbf{r}/dt\overbrace{\cancel{\int_{-\infty}^{t}\omega(\mathbf{r}(t'))e^{-\int_{t'}^{t}\omega(\mathbf{r}(t''))dt''}}dt'}^{1}}_{d\mathbf{r}/dt} \\ & -\omega(\mathbf{r}(t))\underbrace{\int_{-\infty}^{t}\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}\left(\mathbf{r}(t')\right) + \left(\mathbf{r}(t) - \mathbf{r}(t')\right)\right] \times \left[\omega(\mathbf{r}(t'))\cdot e^{-\int_{t'}^{t}\omega(\mathbf{r}(t''))dt''}\right]dt'}_{\boldsymbol{\mathcal{R}}^{\mathbf{A}}}\end{aligned}$$

4. Proof that $\mathbf{S}(\mathbf{r_A}, \mathbf{r_B})$ is symmetric for time-symmetric path distributions

Here we prove that, as long as the distribution of animal trajectories is time-reversal symmetric (given any path, the reverse path $\mathbf{r}_{\text{rev}}(t) = \mathbf{r}(t)$ is equally likely), $\text{S}(\mathbf{r}_{\text{A}}, \mathbf{r}_{\text{B}})$, i.e., the effect of landmark forcing at one position on the mean position self-estimate at another position, will be symmetric. See Table IV for a list of symbols and units.

The mean position self-estimate of the animal at position $\mathbf{r}_{\text{B}}$ is the average self-estimate of all paths that pass $\mathbf{r}_{\text{B}}$ at time $t = 0$. (We pick t = 0 for mathematical convenience). $\bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}(\mathbf{r}_{\text{B}})$ is defined using a path integral over all possible $\mathbf{r}(t)$:

$$\bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}(\mathbf{r}_{\text{B}}) = \int \mathcal{D}\mathbf{r}(t)\text{Pr}[\mathbf{r}(t)]\ \delta(\mathbf{r}(0) - \mathbf{r}_{\text{B}})\boldsymbol{\mathcal{R}}^{\mathbf{A}}[\mathbf{r}(0), t = 0].$$

To avoid clutter, use the shorthand:

$$\boldsymbol{\mathcal{R}}^{\mathbf{A}}[\mathbf{r}(0), t = 0)] = \int_{t'}\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}\left(\mathbf{r}(t')\right) + \left(\mathbf{r}(t) - \mathbf{r}(t')\right)\right]\omega(\mathbf{r}(t'))e^{-\int_{t'}^{t}\omega(\mathbf{r}(t''))dt''} = \int_{t'}\text{F}[\mathbf{r}, t, t']\text{Mem}[\mathbf{r}, t, t']$$

Where:

$$\text{F}[\mathbf{r}, t, t'] = \left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}\left(\mathbf{r}(t')\right) + \left(\mathbf{r}(t) - \mathbf{r}(t')\right)\right]\omega(\mathbf{r}(t')), \quad \text{Mem}[\mathbf{r}, t, t'] = e^{-\int_{t'}^{t}\omega(\mathbf{r}(t''))dt''}.$$

Intuitively, $\text{F}[\mathbf{r}, t, t']$ is the contribution of landmark forcing at t' to the position self-estimate at t, and $\text{Mem}[\mathbf{r}, t, t']$ is the weighting, i.e., the "memory" of time t' at time t. We decompose this into contributions from different past

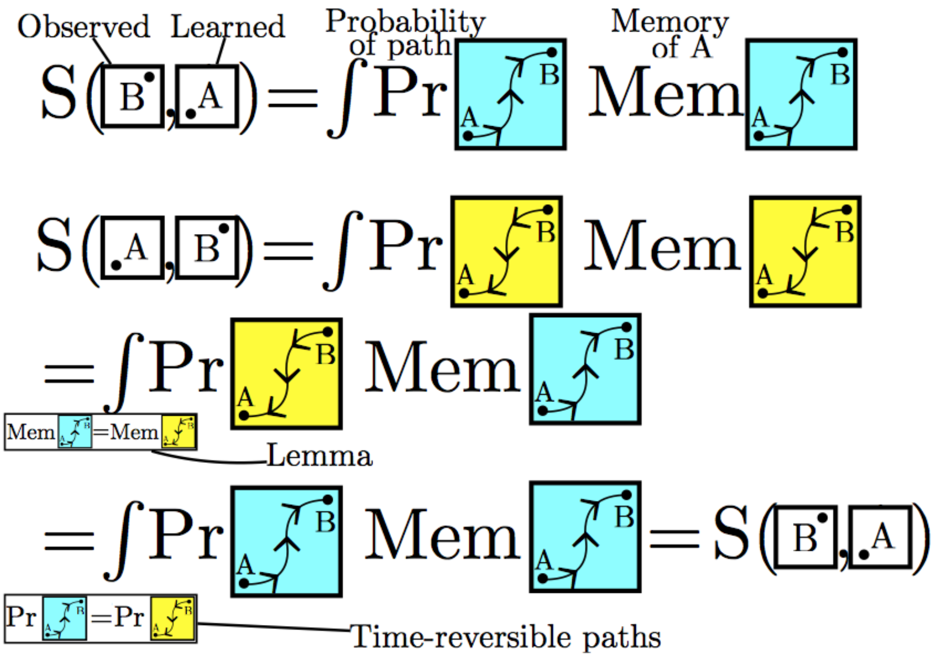

FIG. SI.14:

Sketch of proof in Sec. XI 4 that S is symmetric for time-symmetric path distributions. Our proof relies on two factors. (1) The probability of the reverse path is equal(time-reversal symmetry). (2) The contribution of the mean landmark state at position A to the mean attractor state at position B from the *forward* path is equal to the contribution of the mean landmark state at position B to the mean attractor state at position A from the *reverse* path (Eq. SI.XI.64).

times t':

$$\bar{\mathcal{R}}^{\mathbf{A}}(\mathbf{r}_{\mathrm{B}}) = \int \mathcal{D}\mathbf{r}(t)\mathrm{Pr}[\mathbf{r}(t)]\ \delta(\mathbf{r}(0)-\mathbf{r}_{\mathrm{B}}) \left(\int_{-\infty}^{0} \mathrm{F}[\mathbf{r}, t=0, t']\mathrm{Mem}[\mathbf{r}, t=0, t']dt'\right).$$

Reshuffling the order of integration and breaking terms down further into contributions of $\mathbf{r}_{\mathrm{A}} = \mathbf{r}(t')$

$$\bar{\mathcal{R}}^{\mathbf{A}}(\mathbf{r}_{\mathrm{B}}) = \int_{-\infty}^{0} dt' \int d\mathbf{r}_{\mathrm{A}} \int \mathcal{D}\mathbf{r}(t)\mathrm{Pr}[\mathbf{r}(t)] \ \times \underbrace{\delta(\mathbf{r}(0)-\mathbf{r}_{\mathrm{B}})\delta(\mathbf{r}(t')-\mathbf{r}_{\mathrm{A}})}_{\text{Ensures path from A to B}} \times (\mathrm{F}[\mathbf{r}, t=0, t']\mathrm{Mem}[\mathbf{r}, t=0, t']).$$

Because we have assumed the statistics of the animal trajectories $\mathbf{r}(t)$ will be time-reversal symmetric, the reverse, time shifted path $\mathbf{r}_{\mathrm{rev}}(t) = \mathbf{r}(t'-t)$ is equally likely. We therefore apply the symmetrization procedure:

$$2\bar{\mathcal{R}}^{\mathbf{A}}(\mathbf{r}_{\mathrm{B}}) = \int_{-\infty}^{0} dt' \int d\mathbf{r}_{\mathrm{A}} \int \mathcal{D}\mathbf{r}(t)\mathrm{Pr}[\mathbf{r}(t)] \ \times \left(\underbrace{\delta(\mathbf{r}(0)-\mathbf{r}_{\mathrm{B}})\delta(\mathbf{r}(t')-\mathbf{r}_{\mathrm{A}})}_{\text{Ensures path from A to B}}\mathrm{F}[\mathbf{r}, 0, t']\mathrm{Mem}[\mathbf{r}, 0, t']\right) + \left(\underbrace{\delta(\mathbf{r}_{\mathrm{rev}}(t')-\mathbf{r}_{\mathrm{A}})\delta(\mathbf{r}_{\mathrm{rev}}(0)-\mathbf{r}_{\mathrm{B}})}_{\text{Ensures rev. path from A to B}}\mathrm{F}[\mathbf{r}_{\mathrm{rev}}, 0, t']\mathrm{Mem}[\mathbf{r}_{\mathrm{rev}}, 0, t']\right). \tag{SI.XI.59}$$

We note that the total forgetting of the forward path from time t' to time 0 is the same as the degree along the reverse path, i.e. $\text{Mem}[\mathbf{r}_{\text{rev}}, 0, t'] = \text{Mem}[\mathbf{r}, 0, t']$ (Eq. SI.XI.64) Therefore, we can simplify Eq. SI.XI.59:

$$2\bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}(\mathbf{r}_\text{B}) = \int_{-\infty}^{0} dt' \int d\mathbf{r}_\text{A} \int \mathcal{D}\mathbf{r}(t)\text{Pr}[\mathbf{r}(t)] \times$$
$$\text{Mem}[\mathbf{r}, 0, t'] \left(\underbrace{\delta(\mathbf{r}(0) - \mathbf{r}_\text{B})\delta(\mathbf{r}(t') - \mathbf{r}_\text{A})}_{\text{Ensures path from A to B}} \text{F}[\mathbf{r}, 0, t'] + \underbrace{\delta(\mathbf{r}_{\text{rev}}(t') - \mathbf{r}_\text{A})\delta(\mathbf{r}_{\text{rev}}(0) - \mathbf{r}_\text{B})}_{\text{Ensures rev. path from A to B}} \text{F}[\mathbf{r}_{\text{rev}}, 0, t'] \right). \tag{SI.XI.60}$$

We now expand and simplify the contribution of landmark forcing from $\mathbf{r}_\text{A}$ to the position self-estimate at $\mathbf{r}_\text{B}$ for both the forward and reverse paths (Eq. SI.XI.65, Eq. SI.XI.66):

$$\underbrace{\delta(\mathbf{r}(t') - \mathbf{r}_\text{A})\delta(\mathbf{r}(0) - \mathbf{r}_\text{B})}_{\text{Ensures path from A to B}} \text{F}[\mathbf{r}, 0, t'] = \underbrace{\delta(\mathbf{r}(t') - \mathbf{r}_\text{A})\delta(\mathbf{r}(0) - \mathbf{r}_\text{B})}_{\text{Ensures path from A to B}} \big(\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}_\text{A}) + \mathbf{r}_\text{B} - \mathbf{r}_\text{A}\big)\,\omega(\mathbf{r}_\text{A}), \tag{SI.XI.61}$$

$$\underbrace{\delta(\mathbf{r}_{\text{rev}}(t') - \mathbf{r}_\text{A})\delta(\mathbf{r}_{\text{rev}}(0) - \mathbf{r}_\text{B})}_{\text{Ensures reverse path from A to B}} \text{F}[\mathbf{r}_{\text{rev}}, 0, t'] = \underbrace{\delta(\mathbf{r}(t') - \mathbf{r}_\text{B})\delta(\mathbf{r}(0) - \mathbf{r}_\text{A})}_{\text{Ensures path from B to A}} \big(\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}_\text{A}) + \mathbf{r}_\text{B} - \mathbf{r}_\text{A}\big)\,\omega(\mathbf{r}_\text{A}) \tag{SI.XI.62}$$

Taking advantage of this shared structure in Eq. SI.XI.61, Eq. SI.XI.62, we simplify Eq. SI.XI.60 to:

$$2\bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}(\mathbf{r}_\text{B}) = \int d\mathbf{r}_\text{A} \int_{-\infty}^{0} dt' \int \mathcal{D}\mathbf{r}(t)\text{Pr}[\mathbf{r}(t)] \times$$
$$\text{Mem}[\mathbf{r}, 0, t']\big[\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}_\text{A}) + (\mathbf{r}_\text{B} - \mathbf{r}_\text{A})\big]\omega(\mathbf{r}_\text{A}) \left[\underbrace{\delta(\mathbf{r}(0) - \mathbf{r}_\text{B})\delta(\mathbf{r}(t') - \mathbf{r}_\text{A})}_{\text{Ensures path from A to B}} + \underbrace{\delta(\mathbf{r}(0) - \mathbf{r}_\text{A})\delta(\mathbf{r}(t') - \mathbf{r}_\text{B})}_{\text{Ensures path from B to A}} \right] = \tag{SI.XI.63}$$
$$2 \int d\mathbf{r}_\text{A} \text{S}(\mathbf{r}_\text{B}, \mathbf{r}_\text{A})\big[\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}_\text{A}) + (\mathbf{r}_\text{B} - \mathbf{r}_\text{A})\big]\omega(\mathbf{r}_\text{A})$$

Where our matrix entries:

$$\text{S}(\mathbf{r}_\text{B}, \mathbf{r}_\text{A}) = \frac{1}{2} \times \int_{-\infty}^{0} dt' \int \mathcal{D}\mathbf{r}(t)\text{Pr}[\mathbf{r}(t)]\ \text{Mem}[\mathbf{r}, 0, t'] \left(\underbrace{\delta(\mathbf{r}(0) - \mathbf{r}_\text{B})\delta(\mathbf{r}(t') - \mathbf{r}_\text{A})}_{\text{Ensures path from A to B}} + \underbrace{\delta(\mathbf{r}(0) - \mathbf{r}_\text{A})\delta(\mathbf{r}(t') - \mathbf{r}_\text{B})}_{\text{Ensures path from B to A}} \right)$$

are symmetric with respect to the swapping of $\mathbf{r}_\text{B}, \mathbf{r}_\text{A}$.

This proof assumes uniform density of animal positions with uniform areas and equal strengths for each landmark cell. The proof can be generalized beyond these constraints by making effective particles corresponding to certain landmarks more "massive", but here we present the simpler proof in the interest of clarity.

A. Lemmas about functionals used in symmetry proof

We may show $\text{Mem}[\mathbf{r}_{\text{rev}}, 0, t'] = \text{Mem}[\mathbf{r}, 0, t']$ through:

$$\text{Mem}[\mathbf{r}_{\text{rev}}, 0, t'] = e^{-\int_{t'}^{0} \omega(\mathbf{r}_{\text{rev}}(t''))dt''} = e^{-\int_{t'}^{0} \omega(\mathbf{r}(t'-t''))} = e^{-\int_{t'}^{0} \omega(\mathbf{r}(t''))dt''} = \text{Mem}[\mathbf{r}, 0, t'] \tag{SI.XI.64}$$

We can simplify the effect that the landmark forcing at $\mathbf{r}_\text{A}$ has on the position self-estimate at $\mathbf{r}_\text{B}$ (Eq. SI.XI.61) as:

$$
\begin{aligned}
\underbrace{\delta(\mathbf{r}(t')-\mathbf{r}_{\mathrm{A}})\delta(\mathbf{r}(0)-\mathbf{r}_{\mathrm{B}})}_{\text{Ensures path from A to B}}\mathrm{F}[\mathbf{r},0,t'] &= \\
\underbrace{\delta(\mathbf{r}(t')-\mathbf{r}_{\mathrm{A}})\delta(\mathbf{r}(0)-\mathbf{r}_{\mathrm{B}})}_{\text{Ensures path from A to B}}\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}\left(\mathbf{r}(t')\right)+\left(\mathbf{r}(0)-\mathbf{r}(t')\right)\right]\omega(\mathbf{r}(t')) &= \\
\underbrace{\delta(\mathbf{r}(t')-\mathbf{r}_{\mathrm{A}})\delta(\mathbf{r}(0)-\mathbf{r}_{\mathrm{B}})}_{\text{Ensures path from A to B}}\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}_{\mathrm{A}})+\mathbf{r}_{\mathrm{B}}-\mathbf{r}_{\mathrm{A}}\right)\omega(\mathbf{r}_{\mathrm{A}}) &
\end{aligned}
\tag{SI.XI.65}
$$

and can likewise do this for the reverse path(Eq. SI.XI.62):

$$
\begin{aligned}
\underbrace{\delta(\mathbf{r}_{\mathrm{rev}}(t')-\mathbf{r}_{\mathrm{A}})\delta(\mathbf{r}_{\mathrm{rev}}(0)-\mathbf{r}_{\mathrm{B}})}_{\text{Ensures reverse path from A to B}}\mathrm{F}[\mathbf{r}_{\mathrm{rev}},0,t'] &= \\
\underbrace{\delta(\mathbf{r}_{\mathrm{rev}}(t')-\mathbf{r}_{\mathrm{A}})\delta(\mathbf{r}_{\mathrm{rev}}(0)-\mathbf{r}_{\mathrm{B}})}_{\text{Ensures reverse path from A to B}}\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}\left(\mathbf{r}_{\mathrm{rev}}(t')\right)+\left(\mathbf{r}_{\mathrm{rev}}(0)-\mathbf{r}_{\mathrm{rev}}(t')\right)\right]\omega(\mathbf{r}_{\mathrm{rev}}(t')) & \\
= \underbrace{\delta(\mathbf{r}_{\mathrm{rev}}(t')-\mathbf{r}_{\mathrm{A}})\delta(\mathbf{r}_{\mathrm{rev}}(0)-\mathbf{r}_{\mathrm{B}})}_{\text{Ensures reverse path from A to B}}\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}_{\mathrm{A}})+\mathbf{r}_{\mathrm{B}}-\mathbf{r}_{\mathrm{A}}\right)\omega(\mathbf{r}_{\mathrm{A}}) & \\
= \underbrace{\delta(\mathbf{r}(t')-\mathbf{r}_{\mathrm{B}})\delta(\mathbf{r}(0)-\mathbf{r}_{\mathrm{A}})}_{\text{Ensures path from B to A}}\left(\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r}_{\mathrm{A}})+\mathbf{r}_{\mathrm{B}}-\mathbf{r}_{\mathrm{A}}\right)\omega(\mathbf{r}_{\mathrm{A}}). &
\end{aligned}
\tag{SI.XI.66}
$$

XII. DETAILS OF SIMULATIONS AND DATA ANALYSIS

Here, we provide details of the simulations and analysis used for Sec. XII. See Table VI for a list of symbols and units. Code available at https://github.com/ganguli-lab/EmergentElasticityAnalysisAndSimulations

1. Simulations

A. Exploration

In our simulations, we discretize space onto a grid. For simplicity, we have the animal follow diffusive dynamics, implemented through a random walk; at every time step, the animal moves to one of four neighboring cells; any move which would take the animal outside the box is prohibited. The animal has a position self-estimate $\boldsymbol{\mathcal{R}}^{\mathbf{A}}(t)$ as well as an attractor state $\boldsymbol{\phi}^{\mathbf{A}}(t)$, which undergoes discrete path-integration at every time step:

$$
\begin{aligned}
\boldsymbol{\mathcal{R}}^{\mathbf{A}}(t+\Delta t) &\rightarrow \boldsymbol{\mathcal{R}}^{\mathbf{A}}(t)+\Delta\mathbf{r}_{\mathrm{Sim}}(t), \\
\boldsymbol{\phi}^{\mathbf{A}}(t+\Delta t) &\rightarrow \boldsymbol{\phi}^{\mathbf{A}}(t)+\mathbf{K}\ \cdot\ \Delta\mathbf{r}_{\mathrm{Sim}}(t)
\end{aligned}
$$

Afterwards, the position self-estimate is pulled towards the position estimates of any landmark cells which are firing:

$$
\begin{aligned}
\boldsymbol{\mathcal{R}}^{\mathbf{A}}(t+\Delta t) &\rightarrow \boldsymbol{\mathcal{R}}^{\mathbf{A}}(t+\Delta t)+\left(\omega(\mathbf{r})\left[\boldsymbol{\mathcal{R}}^{\mathbf{L}}(\mathbf{r})-\boldsymbol{\mathcal{R}}^{\mathbf{A}}(t+\Delta t)\right]\right)\times\Delta t \\
\boldsymbol{\phi}^{\mathbf{A}}(t+\Delta t) &\rightarrow \boldsymbol{\phi}^{\mathbf{A}}(t+\Delta t)+\left(\sum_i \omega_i \mathrm{H}_i(\mathbf{r}(t))\sum_{\boldsymbol{\phi}^{\mathbf{L}}}\tilde{\mathrm{W}}_i\left(\boldsymbol{\phi}^{\mathbf{L}}\right)\ \boldsymbol{\mathcal{F}}\left(\boldsymbol{\phi}^{\mathbf{A}}-\boldsymbol{\phi}^{\mathbf{L}}\right)\right)\times\Delta t,
\end{aligned}
$$

Where $\boldsymbol{\phi}^{\mathbf{L}}$ is discretized into a 15×15 grid so that $\tilde{\mathrm{W}}_i\left(\boldsymbol{\phi}^{\mathbf{L}}\right)$ can be represented as an array. We set the timescale of animal motion to be:

$$
\Delta t = \frac{\left|\Delta\mathbf{r}_{\mathrm{Sim}}\right|^2}{\mathrm{D}}
$$

where D is the “diffusion rate” of the animal; this scaling removes dependence on the discretization size.

B. Learning

The learned states are initialized to their firing field center of masses. At every learning epoch T, the simulated animal is placed in the box with an initial position and position self-estimate and explores to get good statistics. $\bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}(\mathbf{r})$, is logged, and at the end of each learning epoch, the position estimate of each landmark cell i is updated to be the average position self-estimate when the landmark cell is firing.

$$\boldsymbol{\mathcal{R}}^{\mathbf{L}}_{i,\mathrm{T}+1} \to \bar{\boldsymbol{\mathcal{R}}}^{\mathbf{A}}_{i,\mathrm{T}}, \qquad \tilde{\mathrm{W}}_i \left(\boldsymbol{\phi}^{\mathbf{L}}\right) \to \Pr(\boldsymbol{\phi}^{\mathbf{A}}(t) = \boldsymbol{\phi}^{\mathbf{L}} | i \text{ Firing}).$$

Each of these will converge after a handful of learning epochs; in practice, we use twenty.

C. Simulation of square and bent environments

Landmark cell firing fields are heterogeneous; while some are distributed across an entire border; to replicate this distribution we have two types of landmark cells in our model. (1) Landmark cells having uniform wall-length firing field, with a width of 10cm, for example $\mathrm{H}(x, y) = e^{-(\frac{x-x_{\mathrm{wall}}}{5\mathrm{cm}})^2}$ for a landmark cell on the west wall. (2) More localized, overlapping, firing fields along each wall. Each firing field is a 5 cm × 10 cm half-ellipse of along a particular wall; i.e. $\mathrm{H}(x, y) = e^{-(\frac{y-y_0}{10\mathrm{cm}})^2-(\frac{x-x_{\mathrm{wall}}}{5\mathrm{cm}})^2}$ for a landmark field along the EW wall with center y_0. Each type of landmark cell is evenly distributed along each wall, with the total strength and number set such that total firing strength of localized and non-localized cells is the same, and their combined strength leads to a forgetting time of $\omega = 8$Hz along each wall.

Grid spacing is chosen to be 30 cm for square environments (1 × 1 meter); We set the diffusive constant D to be $(10 \text{ cm})^2$/Second such that it takes an animal ∼100 seconds to traverse the width of the environment.

Grid spacing is 50 *cm* with a 7° offset for the first trapezoidal environment (1.9 × .8 meters, same geometry as [26]); We set the diffusive constant D to be $(20 \text{ cm})^2$/Second such that it takes an animal ∼100 seconds to traverse the length of the environment.

Grid spacing is 50 *cm* with a 7° offset for the second trapezoidal environment (1.9 meters long. Two straight walls with lengths of .12 meters, .6 meters, with diagonal walls starting 1 meter from the smaller straight wall (14° angle); We set the diffusive constant D to be $(20 \text{ cm})^2$/Second such that it takes an animal ∼100 seconds to traverse the length of the environment.

The angular offset breaks the symmetry of the trapezoidal environments, yielding bending, but is not required to yield path-dependent shifts.

D. Simulation of topological environments

In order for an environment to support topological defects, cues must be rich and localized, leading to uniformly distributed landmark firing fields. To model this, we have uniformly localized landmark fields, with $\mathrm{H}(x, y) = e^{-(\frac{y-y_0}{10\mathrm{cm}})^2-(\frac{x-x_0}{10\mathrm{cm}})^2}$, arranged at a density such that their combined strength leads to a forgetting time of 1Hz throughout the environment. The environment was 1.8 meters × 1 meter, with a center rectangular section of 1.3 × .8 meters removed. **K** is chosen to yield a grid spacing of 60cm. The first simulation(no topological defect) was initialized with no landmark weights, while the second simulation (topological defects) was initialized with landmark weights corresponding to a topological defect; both initial conditions relaxed into different (meta)stable internal maps.

E. Force law and visualization

The simulated grid cell patterns are visualized by using a truncated parabolic firing rate:

$$\left[1 - \left|\frac{\boldsymbol{\phi}^{\mathbf{A}} - \mathbf{u}}{\mathrm{B}}\right|^2\right]_+$$

where $\mathbf{u}$ is the position of the "recorded" cell on the neural sheet and the field width B is chosen to be $2\pi/5$.

The force law chosen is a truncated sin function:

$$\mathcal{F}(\boldsymbol{\phi}^{\mathbf{L}} - \boldsymbol{\phi}^{\mathbf{A}}) = \begin{cases} \left(\boldsymbol{\phi}^{\mathbf{L}} - \boldsymbol{\phi}^{\mathbf{A}}\right) \times \frac{\sin\left(\left|\boldsymbol{\phi}^{\mathbf{L}} - \boldsymbol{\phi}^{\mathbf{A}}\right|\right)}{\left|\boldsymbol{\phi}^{\mathbf{L}} - \boldsymbol{\phi}^{\mathbf{A}}\right|} & \left|\left(\boldsymbol{\phi}^{\mathbf{L}} - \boldsymbol{\phi}^{\mathbf{A}}\right)\right| < \pi \\ 0 & \left|\left(\boldsymbol{\phi}^{\mathbf{L}} - \boldsymbol{\phi}^{\mathbf{A}}\right)\right| \geq \pi \end{cases} \qquad \text{(SI.XII.67)}$$

We choose this function because it has the correct qualitative features. In addition, in experimental data, the width of a firing rate peak is on the order of the spacing between two firing peaks; this prohibits a force law which is much more short-ranged than this (Sec. III 2).

2. Experimental methods

Data included a subset of published neural recordings previously presented in Hardcastle et al., 2017, Hardcastle et al., 2015. Briefly, mice explored a square box while foraging for chocolate cheerios sprinkled on the floor. During each recording, neural signals from medial entorhinal cortex were recorded and subsequently clustered into distinct neurons. A grid score was computed for each cell following Langston et al., 2010. Cells above a threshold of .4 were considered grid cells. Each grid cell in the dataset was recorded after an average of 28 (data selected from Hardcastle et al., 2015) or 20 (data selected from Hardcastle et al., 2017) exposures to the recording environment.

A. Pre-processing of trajectories

To control for the effect of head direction and running speed, we preprocessed the data by translating

$$\mathbf{r}(t) \to \mathbf{r}(t) + 1\mathrm{cm} \;\times\; \hat{\mathbf{HD}}(t),$$

where $\hat{\mathbf{HD}}(t)$ is a unit vector representing the animal's head direction as a function of time. This is to avoid artifacts related to tracking; a purely position-dependent firing rate model depends on *some* part of the animal's body, which unlikely to be exactly the position of the tracking diode. Because head direction is correlated with the last border touched, head direction-depdendent shifts from this artifact would yield path-dependent shifts; our preprocessing removes this possibility.

B. Subtraction of average animal position for shifts in patterns around firing fields

We define the path conditioned shift (Eq. SI.IX.52) as the *difference* between the average spike position within a firing field and the mean *animal* position within that firing field.

$$\mathbf{S}_{\mathcal{C},\mathrm{GC},\mathbf{ff}} = \langle \mathbf{r}_{\mathrm{Spk}} - \mathbf{r}_{\mathbf{ff}} | \mathcal{C}, \mathbf{r}_{\mathrm{Spk}} \in \mathbf{ff} \;\rangle - \langle \mathbf{r}(t) - \mathbf{r}_{\mathbf{ff}} | \mathcal{C}, \mathbf{r}(t) \in \mathbf{ff} \;\rangle$$

The animal's position within the firing field is subtracted to eliminate any systematic biases that might come from the animal trajectory rather than the actual neural activity (Fig. SI.15).

[1] G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. Friston, PLOS Computational Biology **4**, 1 (2008).

[2] We have chosen for the path integration and landmark inputs to not be fed into the nonlinearity for mathematical simplicity; these can be fed in arbitrarily, although doing so yields different force functions for path integration and landmark input.

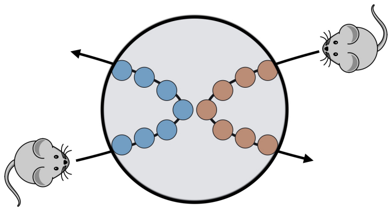

FIG. SI.15:

Schematic of the motivation for subtracting mouse position in Eq. SI.IX.52. An animal is most likely to be closest to the last wall it touched; if the mean animal position was *not* subtracted from the mean spike position, this would yield a path-dependent shift in spike positions purely dependent on animal trajectory rather than neural activity.

[3] $s_{\rm EC}(u)$, $s_{\rm WC}(u)$ could follow their own differential equations [7, 29]; as long as they yield a perturbation which is linear with animal velocity the mathematical techniques developed would still hold. We choose not to do so here for simplicity.

[4] A training session is roughly the same as a learning epoch.

[5] We have neglected to add a multiplier representing the base learning rate. In principle, Eq. SI.II.3 should read $d/d{\rm T} = {\rm T}_0^{-1}\dots$, but we leave this base learning rate ${\rm T}_0^{-1}$ out for simplicity.

[6] A. Samsonovich and B. L. McNaughton, Journal of Neuroscience **17**, 5900 (1997), http://www.jneurosci.org/content/17/15/5900.full.pdf.

[7] Y. Burak and I. R. Fiete, PLoS Comput Biol **5**, e1000291 (2009).

[8] ${\rm Jac}_{\phi^{\rm A}}$ is shorthand for ${\rm Jac}_{s^*(u-\phi^{\rm A})}$.

[9] The matrix ${\rm Jac}_{\phi^{\rm A}}$ is a function of pairs of neural sheet positions, so its eigenvectors are functions of neural sheet position.

[10] We can show this by contradiction; if ${\rm Jac}_{\phi^{\rm A}}$ had any other eigenvector with a non-negative eigenvalue, the family of steady states would be larger.

[11] This is true as long as the Jacobian is symmetric, i.e., all eigenvectors are orthogonal. When this is not the case, Eq. SI.III.11 must include a *non-orthogonal* projection onto the sliding mode, which yields (Eq. SI.III.17) in the continuous case.

[12] The eigenvectors of any symmetric matrix are orthogonal.

[13] The integration bounds are over the ring so do not change.

[14] $\mathcal{N} = \int_u \left(s^{*\prime}(u)\right)^2$.

[15] This second variable is needed to account for the fact that a landmark with uniform $\tilde{\rm W}(\phi^{\rm L})$ will not exert any force.

[16] In simulations, we observe localization of $\tilde{\rm W}(\phi^{\rm L})$ al long as the landmark fields themselves are localized.

[17] Slope magnitude can be packed into a multiplier of $\mathcal{F}$.

[18] A. T. Keinath, R. A. Epstein, and V. Balasubramanian, bioRxiv (2017), 10.1101/174367, http://www.biorxiv.org/content/early/2017/08/10/174367.full.pdf.

[19] K. Gothard, W. Skaggs, and B. L, J. Neurosci. **16**, 8027 (1996).

[20] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, Proceedings of the National Academy of Sciences **92**, 3844 (1995), http://www.pnas.org/content/92/9/3844.full.pdf.

[21] Only u very close to $0, \pi$ will have $\mathcal{G}(s^*(u)) \neq s^*(u)$ (See Fig. SI.8).

[22] Time-reversible means that for any $\mathbf{r}(t)$, the reverse path $\mathbf{r}(-t)$ is equally likely.

[23] This is due to the fact that shifting *all* position estimates and landmark position estimates will not change the dynamics.

[24] T. Stensola, H. Stensola, M.-B. Moser, and E. I. Moser, Nature **518** (2015).

[25] W. E. Skaggs, B. L. McNaughton, M. A. Wilson, and C. A. Barnes, Hippocampus **6**, 149 (1996).

[26] J. Krupic, M. Bauza, S. Burton, C. Barry, and J. O'Keefe, Nature **518** (2015), 10.1038/nature14153.

[27] J. Krupic, M. Bauza, S. Burton, and J. O'Keefe, Science **359**, 1143 (2018), http://science.sciencemag.org/content/359/6380/1143.full.pdf.

[28] J. Krupic, M. Bauza, S. Burton, C. Lever, and J. O'Keefe, … of the Royal … **369**, 20130188 (2014).

[29] B. Si, S. Romani, and M. Tsodyks, Plos Comput Biol **10**, e1003558 (2014).