Planning as Dynamics Relaxation: Hippocampal Recurrent Network Realizes Optimal Goal-Directed Navigation

Yuhang He[1,2,3,4,5] **Junfeng Zuo**[1,2,3,4,5] **Tianhao Chu**[1,3] **Si Wu**[1,2,3,4,5]*
[1]School of Psychological and Cognitive Sciences, Peking University
[2]IDG/McGovern Institute for Brain Research, Peking University
[3]Peking-Tsinghua Center for Life Sciences,
Academy for Advanced Interdisciplinary Studies, Peking University
[4]Center of Quantitative Biology, Peking University
[5]Key Laboratory of Machine Perception (Ministry of Education), Peking University

Abstract

Neural correlates of the cognitive map are well documented, yet how a neural circuit turns a stored map into an obstacle-avoiding route remains unclear. Here we show that a hippocampal network with appropriate recurrent connections achieves optimal goal-directed navigation through its own relaxation dynamics. The recurrent weights represent the transition probabilities between the locations the neurons encode, walls are reflected by vanishing weights, and behavioral-timescale synaptic plasticity (BTSP) writes this connectivity during goal-free exploration. When a goal is recalled, a top-down current clamps the goal neuron and the population relaxes into a goal-locked activity field. We prove that this field is the desirability field of a linearly-solvable Markov decision process: the relaxed state is the exact optimum of a stated control problem, the KL-regularized objective of Todorov, and the optimal next-step distribution is carried by the branch synaptic currents. The identity extends to directed transitions, and a weak-cost limit returns the Green's function of the recurrent Laplacian. A place-tuned readout converts the field's log-gradient into a direction of travel, and the same synapses answer a predictive query when the source is placed at the current location. Simulations on mazes with complex obstacles show near-shortest routes, an identity that survives a thousand-fold range of recall amplitude, the loss of 80% of place cells, and tens of thousands of states, low-rank updates of every plan after a local change, and a transient that is anytime: the local action is ready long before the field converges. We hope this offers a general circuit principle for planning in abstract cognitive maps beyond spatial navigation.

1 Introduction

An animal in a familiar environment can recall a destination and plan a route that avoids walls and takes available detours. The neural correlates of this cognitive map are well established: hippocampal place cells signaling location (O'Keefe and Dostrovsky, 1971), entorhinal grid cells tiling the explored space (Hafting et al., 2005), and experience-dependent plasticity that binds them into a stored representation (McHugh et al., 1996; Mehta et al., 2000; Lever et al., 2002; Bittner et al., 2017). How these circuits produce an obstacle-avoiding route from the stored map is much less understood.

*Corresponding author: siwu@pku.edu.cn.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Related work (App. A). Predictive-map and successor-representation models propose that the hippocampus stores long-horizon transition statistics from which a value function for any goal can be derived (Dayan, 1993; Stachenfeld et al., 2017; Piray and Daw, 2021; Bono et al., 2023; Fang et al., 2023; George et al., 2023), a view that extends to abstract task structures (Behrens et al., 2018; Whittington et al., 2020, 2022; George et al., 2021; Garvert et al., 2023; Courellis et al., 2024; El-Gaby et al., 2024; Tacikowski et al., 2024; Bakermans et al., 2025). Linearly-solvable Markov decision processes (LMDPs) reduce planning to a linear equation whose solution, the desirability field, gives the navigation direction through its log-gradient (Kappen, 2005; Todorov, 2006, 2009; Levine, 2018), with harmonic and Laplacian methods as a related geometric picture (Connolly et al., 1990; Mahadevan and Maggioni, 2007; Machado et al., 2017); other biological models simulate prospective trajectories during theta sequences and replay (Johnson and Redish, 2007; Pfeiffer and Foster, 2013; Mattar and Daw, 2018; Liu et al., 2019; Schwartenbeck et al., 2023; Jensen et al., 2024). A lineage of recurrent models lets a network settle toward a plan: the cognitive graph of Muller et al. (1996), wavefront propagation (Ponulak and Hopfield, 2013), attractor preplay (Corneil and Gerstner, 2015), recurrent successor representations (Fang et al., 2023), Endotaxis (Zhang et al., 2024), and the spacetime attractor of Jensen et al. (2026); their relaxed states support planning as an arrival time, a trajectory, or a predictive field. Whether the network that stores the map can itself relax into the optimal plan, with the optimum proven rather than approximated, has remained open.

Here we show that a hippocampal recurrent network performs goal-directed navigation through its own relaxation dynamics (Fig. 1). The recurrent weights represent the transition probabilities between the locations the neurons encode, written during goal-free exploration by BTSP (Bittner et al., 2017; Li et al., 2024), with walls reflected by vanishing weights. When the animal recalls a destination, a top-down input clamps the corresponding neuron and the population relaxes into a goal-locked field, which we prove to be the desirability field of an LMDP: the relaxed state is the exact optimum of a stated control problem, and the optimal next-step distribution is carried by the circuit's own branch currents. A downstream place-tuned population combines the field with its local place kernel by dendritic multiplication (Mitchell and Silver, 2003), and its population vector points along the field's log-gradient toward the goal.

Our specific contributions are as follows:

- **LMDP equivalence.** The goal-clamped steady state equals the LMDP desirability field, with the optimal policy in the branch synaptic currents (Theorem 1); the identity extends to directed transitions (Corollary 1) and, in a weak-cost limit, to the Green's function of the recurrent connectivity, a circuit basis for the spectral account of Zuo et al. (2026) (Sec. 2.4).
- **Place-tuned readout.** A place-tuned population multiplies the field by its local place kernel; its population vector points toward the goal, is invariant to the recall amplitude, and reads the raw firing rate (Sec. 2.3).
- **Learning by BTSP.** BTSP writes the connectivity during goal-free exploration, and a local change updates every plan through a low-rank correction (Sec. 3).
- **Scope and comparisons.** Grid mazes, continuous arenas, directed maps, and maps with up to $38{,}556$ states confirm the identity to machine precision, show that the transient is anytime, and compare the circuit with learned goal-conditioned models at matched experience (Sec. 4).

2 The model

2.1 Recurrent network model

The hippocampus (e.g. CA3) is densely interconnected by recurrent collaterals (Amaral and Witter, 1989; Ishizuka et al., 1990; Sammons et al., 2024; Mishra et al., 2016; Li et al., 2024), and after exploration its pyramidal cells form a place code (O'Keefe and Burgess, 1996). We model this population as N recurrently connected units in continuous time; a unit is an assembly of cells that codes one location (Sec. 2.5). Unit i codes location i with firing rate $r_i(t)$, and the recurrent synapse from unit j to unit i has weight S_{ij}, collected in $\boldsymbol{S} \in \mathbb{R}^{N\times N}$ (Sec. 3 describes how plasticity writes $\boldsymbol{S}$). Each unit receives a background I_i^{bg}, an entorhinal input $I_i^{\mathrm{EC}}(\boldsymbol{y}_t)$ tuned to the current position $\boldsymbol{y}_t$ (Hafting et al., 2005), and a top-down goal current $I_i^{\mathrm{goal}}(g) \propto \delta_{ig}$ delivered to the goal unit during

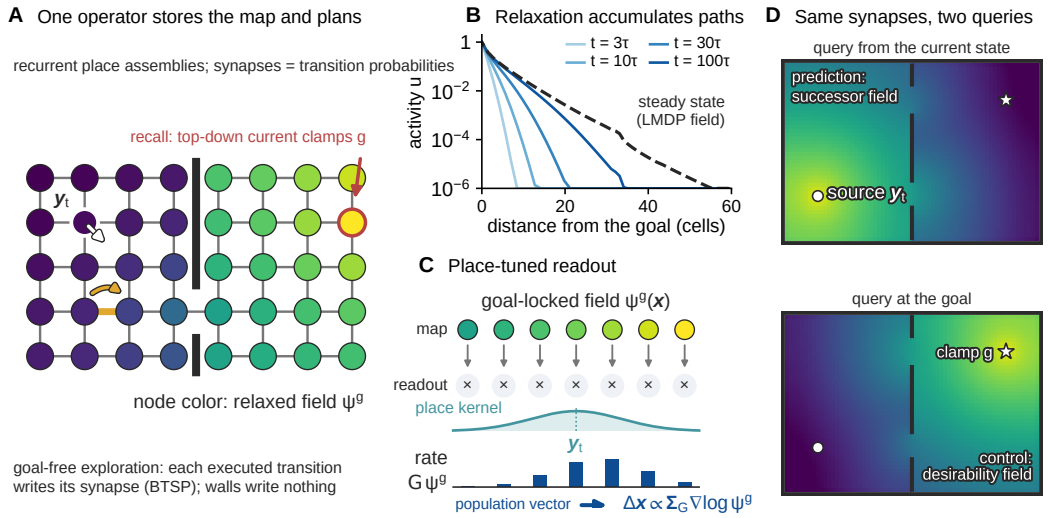


Figure 1: **One recurrent operator stores the map and computes the plan.** (**A**) Recurrent place assemblies; a synapse exists only where BTSP wrote an executed transition (gold), so walls write nothing. Recalling a goal clamps one assembly (red); the population relaxes into the goal-locked field ψ^g (node color, log scale), and a readout at the current position $\boldsymbol{y}_t$ points up its log-gradient (white arrow). (**B**) The activity profile along the route grows out from the clamped goal and converges to the LMDP desirability field (two-room maze, $\alpha = 0.985$). (**C**) Place-tuned readout cells multiply the field by a local place kernel $\mathcal{G}$ centered at $\boldsymbol{y}_t$; their population vector is $\boldsymbol{\Sigma}_G \nabla \log \psi^g$ (Eq. 7). (**D**) Two queries of one operator: a source at the current state yields the successor field (prediction), a clamp at the goal the desirability field (control).

recall (Ito et al., 2015):

$$\tau_u \, \dot{r}_i = -r_i + g_s \sum_j S_{ij} \, \phi(r_j) - g_{\text{inh}} \, \phi(r_i) + I_i^{\text{bg}} + I_i^{\text{EC}}(\boldsymbol{y}_t) + I_i^{\text{goal}}(g), \tag{1}$$

where ϕ is a smooth saturating activation function (we use $\tanh$), τ_u is the membrane time constant, and $g_s, g_{\text{inh}} > 0$ scale the recurrent excitation and the local inhibition. Without recall the population rests at $\boldsymbol{r}^0(\boldsymbol{y}_t)$, set by the background and entorhinal drives. Recalling goal g clamps the goal unit at $r_g = r_g^0 + \bar{u}_g$, with $\bar{u}_g > 0$ the recall amplitude, and the remaining units evolve toward a goal-driven steady state. Writing $u_i := r_i - r_i^0$ and $\delta I_i := I_i^{\text{goal}}(g)$ for the only input that changes, the expansion of ϕ around r_i^0 gives

$$\tau \, \dot{\boldsymbol{u}} = -(\boldsymbol{I}_N - \alpha \, \boldsymbol{S}) \, \boldsymbol{u} + \delta \boldsymbol{I} + \mathcal{O}(\|\boldsymbol{u}\|^2), \qquad \alpha := \frac{g_s \, \rho}{1 + g_{\text{inh}} \, \rho} \in (0, 1), \qquad \tau := \frac{\tau_u}{1 + g_{\text{inh}} \, \rho}, \tag{2}$$

where $\rho := \phi'(r_i^0)$ is the slope of the activation function at rest (App. C.2). Local inhibition reduces both the effective gain α and the effective time constant τ, and $\alpha < 1$ keeps the network stable. A length-L path through $\boldsymbol{S}$ contributes $\alpha^L = e^{-\kappa L}$ with $\kappa := -\log \alpha$, so $1/\kappa$ is the number of synaptic steps activity travels before it decays.

2.2 Goal-clamped relaxation produces the desirability field

Relaxation carries its ordinary meaning in physics and numerical analysis: after the goal clamp perturbs the resting state, activity spreads outward through the recurrent connections, about one synaptic hop per time constant τ, until the population settles into a new steady field (Fig. 1B). With symmetric $\boldsymbol{S}$ the dynamics are a gradient flow on a convex energy, and one Euler step of Eq. (2) is one Jacobi sweep of the equation the field solves. Planning is this settling, and Theorem 1 characterizes its result.

With the goal unit clamped at $u_g = \bar{u}_g$, let $\mathcal{I}_g := \{1, \ldots, N\} \setminus \{g\}$ denote the non-goal units. Setting $\dot{\boldsymbol{u}} = 0$ in Eq. (2) on $\mathcal{I}_g$ and dropping the quadratic remainder gives

$$(\boldsymbol{I}_{\mathcal{I}_g} - \alpha \, \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g}) \, \boldsymbol{u}_{\mathcal{I}_g} = \boldsymbol{b}_g, \qquad (\boldsymbol{b}_g)_i := \alpha \, S_{ig} \, \bar{u}_g \quad \text{for } i \in \mathcal{I}_g, \tag{3}$$

where $\boldsymbol{b}_g$ is the current the clamped goal sends into unit i through the synapse S_{ig}. Since $\alpha < 1$, the inverse expands as $\sum_{L\geq 0} \alpha^L \boldsymbol{S}^L_{\mathcal{I}_g\mathcal{I}_g}$: u_i sums every legal path from i to the goal, weighted by the product of its synaptic weights and damped by $e^{-\kappa L}$, so shorter routes weigh more and walls, whose synapses never form, contribute nothing.

To interpret this field, we state a planning problem and match its solution to the relaxation. An animal at location i chooses a distribution $\mu(\cdot \mid i)$ over its next location and pays a per-step cost with two parts: a fixed amount $q_0 > 0$ for the step, plus a deviation penalty $\lambda\, D_{\mathrm{KL}}\big(\mu(\cdot \mid i) \,\|\, P_0(\cdot \mid i)\big)$, where $P_0(\cdot \mid i)$ describes how the animal moves with no goal in mind; the plasticity that writes $\boldsymbol{S}$ also writes $\boldsymbol{P}_0$ as the spontaneous next-step distribution among the accessible neighbors of i (Sec. 3.1). "Optimal" throughout the paper means the exact optimum of this KL-regularized objective, the linearly-solvable class of Todorov (2006). With $V(i)$ the expected total future cost from i under the best plan and $V(g) = 0$, the substitution $z^g(i) := \exp(-V(i)/\lambda)$ turns the Bellman equation into a linear one (Kappen, 2005; Todorov, 2006, 2009):

$$(\boldsymbol{I}_{\mathcal{I}_g} - \gamma\, \boldsymbol{P}_{0,\,\mathcal{I}_g\mathcal{I}_g})\, \boldsymbol{z}^g_{\mathcal{I}_g} = \gamma\, \boldsymbol{p}_g, \qquad (\boldsymbol{p}_g)_i := P_0(g \mid i), \qquad \gamma := e^{-q_0/\lambda} \in (0,1), \tag{4}$$

with $z^g(g) = 1$, and the optimal next-step distribution is $\mu^*(j \mid i) \propto P_0(j \mid i)\, z^g(j)$. Eqs. (3) and (4) have the same form: each holds the goal at a fixed value and spreads it through the population, through $\alpha\boldsymbol{S}$ or through $\gamma\boldsymbol{P}_0$, and $\alpha = \gamma$ matches the recurrent decay κ per synaptic step to the planning cost q_0/λ per step. The plasticity rule samples each accessible transition forward and backward equally often (Sec. 3.1), so $\boldsymbol{P}_0$ satisfies detailed balance with respect to a positive measure $\boldsymbol{\pi}$, $\pi_i P_0(j \mid i) = \pi_j P_0(i \mid j)$, and the same rule writes

$$\boldsymbol{S} = \boldsymbol{\Pi}^{1/2}\, \boldsymbol{P}_0\, \boldsymbol{\Pi}^{-1/2}, \qquad \boldsymbol{\Pi} := \mathrm{diag}(\pi_1, \ldots, \pi_N), \tag{5}$$

which detailed balance makes symmetric (Apps. E.1 and D.3).

Theorem 1 (Equivalence between goal-clamped relaxation and the LMDP desirability)**.** *Suppose $\boldsymbol{P}_0$ is reversible with respect to a positive measure $\boldsymbol{\pi}$, the recurrent weights are $\boldsymbol{S} = \boldsymbol{\Pi}^{1/2}\boldsymbol{P}_0\boldsymbol{\Pi}^{-1/2}$, $\alpha = \gamma$, and the network operates in the linear regime of Eq.* (2)*, so that the goal-clamped steady state solves Eq.* (3)*. Then at every unit i,*

$$\frac{u_i}{\bar{u}_g} \;=\; \sqrt{\frac{\pi_i}{\pi_g}}\; z^g(i), \tag{6}$$

where z^g is the LMDP desirability field with discount γ and $z^g(g) = 1$, and the optimal policy is carried by the branch synaptic currents, $\mu^*(j \mid i) \propto S_{ij}\, u_j$.

The remainder dropped in Eq. (3) has an explicit bound (App. C.2). The step cost q_0 sets the recurrent decay and penalizes long trajectories; the deviation cost measures how far each planned step bends the spontaneous flow away from rest; the relaxation stores the value of this trade-off at every unit. When units cover the environment evenly, bidirectional sampling makes $\boldsymbol{\pi}$ uniform (App. E.3), $\boldsymbol{S} = \boldsymbol{P}_0$, and the normalized field $\psi^g(i) := u_i/\bar{u}_g$ coincides with $z^g(i)$ (general case: App. F.4).

Strong recall leaves the linear regime at the clamped unit, where ϕ saturates. The leading effect is a uniform rescaling of the whole field: the drive the clamped unit sends into the network shrinks to an effective amplitude $A_{\mathrm{eff}}(\bar{u}_g) < \bar{u}_g$ (App. I.1), every non-goal unit, still in its linear range, is rescaled by the same factor, and the log-gradient readout of Sec. 2.3 cancels it. A residual shape distortion remains in a shell around the goal where neighboring units also saturate; it decays exponentially with distance, so the shell radius grows only as $\log \bar{u}_g$ and stays bounded for any saturating ϕ (App. I.4; Sec. 4.1).

Reversibility is the condition for a symmetric realization, not for the computation; the identity extends to genuinely one-way transitions:

Corollary 1 (Directed transitions)**.** *Let $\boldsymbol{P}_0$ be any transition kernel for which Eq.* (4) *has a unique solution, and let $\boldsymbol{S} = \boldsymbol{D}\boldsymbol{P}_0\boldsymbol{D}^{-1}$ for any positive diagonal $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_N)$. Under the remaining assumptions of Theorem 1, $u_i/\bar{u}_g = (d_i/d_g)\, z^g(i)$ and $\mu^*(j \mid i) \propto S_{ij}u_j$ (App. D.4).*

In the directed case the legal move $i \to j$ is stored in the synapse from unit j to unit i, $S_{ij} \propto P_0(j \mid i)$, so activity flows from the goal backward along legal routes. This is the synapse that reverse replay writes under a causal plasticity kernel: a sequence that plays j before i potentiates $j \to i$, so reverse

replay of the experienced trajectory writes $\boldsymbol{S}$, while forward replay writes its transpose, the predictive operator of Sec. 2.3. Reverse replay is prominent at reward (Foster and Wilson, 2006; Ambrose et al., 2016), and the BTSP rule inferred from CA3 place-field dynamics is strongly asymmetric in a new environment and close to symmetric once it is familiar (Madar et al., 2025); in a familiar reversible environment the distinction disappears, because $\boldsymbol{S}$ is symmetric.

2.3 A place-tuned readout of the goal-locked field

To choose a step, the animal needs a single direction at its current position $\boldsymbol{y}_t$. It is supplied by a downstream population of place-tuned readout cells driven by the entorhinal and self-motion inputs that set place tuning at rest (O'Keefe and Burgess, 1996; McNaughton et al., 2006; Aoki et al., 2019; Qian et al., 2025). Each readout cell has a preferred location $\boldsymbol{x}$ and receives a place-tuned input $\mathcal{G}_{\boldsymbol{y}_t}(\boldsymbol{x})$, largest at $\boldsymbol{x} = \boldsymbol{y}_t$ and falling off over the place-field width, together with the value $\psi^g(\boldsymbol{x})$ from the upstream unit coding the same location; the two multiply on the dendrite (Mitchell and Silver, 2003; Silver, 2010), so the cell fires at a rate proportional to $\mathcal{G}_{\boldsymbol{y}_t}(\boldsymbol{x})\,\psi^g(\boldsymbol{x})$ (Fig. 1C). The next-step direction is the population vector, which weights each cell's preferred displacement $\boldsymbol{x} - \boldsymbol{y}_t$ by its rate:

$$\Delta\boldsymbol{x}_{\mathrm{pop}}(\boldsymbol{y}_t, g) = \frac{\int (\boldsymbol{x} - \boldsymbol{y}_t)\,\mathcal{G}_{\boldsymbol{y}_t}(\boldsymbol{x})\,\psi^g(\boldsymbol{x})\,d\boldsymbol{x}}{\int \mathcal{G}_{\boldsymbol{y}_t}(\boldsymbol{x})\,\psi^g(\boldsymbol{x})\,d\boldsymbol{x}} \;\approx\; \boldsymbol{\Sigma}_G(\boldsymbol{y}_t)\,\nabla \log \psi^g(\boldsymbol{y}_t), \tag{7}$$

where $\boldsymbol{\Sigma}_G(\boldsymbol{y}_t)$ is the covariance of $\mathcal{G}_{\boldsymbol{y}_t}$ around $\boldsymbol{y}_t$. The approximation is a local Taylor expansion of $\log \psi^g$, valid when the place field is narrower than the scale on which ψ^g varies (App. G.2), and it matches the mean displacement of the optimal next-step distribution μ^* at $\boldsymbol{y}_t$ (App. H). The circuit never computes $\boldsymbol{x} - \boldsymbol{y}_t$ online: each readout cell's offset from the current position is a fixed property of that cell, so the population vector is a projection through offset-weighted synapses, with divisive normalization supplying the denominator (Silver, 2010).

Two invariances make the readout robust. The direction is unchanged by any overall rescaling of ψ^g, so uniform gain changes in the recurrent population, including changes of the recall amplitude $\bar{u}_g$, drop out (Sec. 2.2). And the readout can take the raw firing rate $r = r^0 + u$ rather than the deviation u:

Proposition 1 (Reading the raw rate)**.** *Let the kernel be centered, $\int (\boldsymbol{x} - \boldsymbol{y}_t)\mathcal{G}_{\boldsymbol{y}_t}(\boldsymbol{x})\,d\boldsymbol{x} = \boldsymbol{0}$, and the resting rate homogeneous, $r^0(\boldsymbol{x}) = c$. Then $\Delta\boldsymbol{x}_{\mathrm{pop}}$ computed from $c + u$ equals $\Delta\boldsymbol{x}_{\mathrm{pop}}$ computed from u times the positive factor $\int \mathcal{G}u/(cM_0 + \int \mathcal{G}u)$, $M_0 = \int \mathcal{G}$; the direction is exactly invariant (App. G.3).*

A position-dependent resting profile adds only a goal-independent term, which the same centering removes; at rest the readout returns the place code itself, because $\Delta\boldsymbol{x}_{\mathrm{pop}}$ vanishes for a symmetric kernel.

Prediction and control are two queries of the same synapses (Fig. 1D): a source at the current state, $(\boldsymbol{I} - \alpha\boldsymbol{S})^{-1}\boldsymbol{e}_{\boldsymbol{y}_t}$, is the resolvent column of a successor representation (Dayan, 1993; Stachenfeld et al., 2017) and predicts where the animal will be; a clamp at the goal is the Dirichlet problem of Eq. (3) and tells it which way to go.

2.4 The weak-cost limit recovers Green's-function navigation

A recent account of hippocampal–entorhinal coding (Zuo et al., 2026) assigned two roles to a discrete Laplacian on the explored environment: its eigenfunctions reproduce the firing maps of grid, band, and boundary cells, and its Green's function is a potential whose gradient guides goal-directed paths. The recurrent relaxation supplies the circuit. Let $\boldsymbol{L} := \boldsymbol{I} - \boldsymbol{S}$ be the Laplacian of the recurrent connectivity; it equals the Laplacian of Zuo et al. (2026) up to a positive scalar (App. J), and $\boldsymbol{L}$, $\boldsymbol{S}$, and that Laplacian share one set of eigenvectors, so those spatial codes emerge as firing patterns of a network whose connectivity BTSP writes. When $q_0 \to 0$, $\alpha \to 1$ and Eq. (3) reduces to $\boldsymbol{L}_{\mathcal{I}_g\mathcal{I}_g}\boldsymbol{u}_{\mathcal{I}_g} = \boldsymbol{S}_{\mathcal{I}_g g}\bar{u}_g$, whose solution is the Green's function of $\boldsymbol{L}_{\mathcal{I}_g\mathcal{I}_g}$, the navigation potential of Zuo et al. (2026) (App. J.2). In the planning picture, $\boldsymbol{z}^g$ flattens as $q_0 \to 0$ and its structure moves into the leading correction $\log z^g(i) \approx -(q_0/\lambda)\,\mathbb{E}_i[\tau_g]$, the expected number of spontaneous steps to the goal; the two cost regimes are two ends of one axis, both arising from the same weights (App. L.4).

Assumption (what it gives)	Where it is relaxed or tested
Reversible $\boldsymbol{P}_0$ (symmetric $\boldsymbol{S}$; Theorem 1; gradient flow)	directed transitions: Corollary 1, Fig. 4A; sampling asymmetry: App. L.5
Linear regime of Eq. (2) (steady state solves the desirability equation)	saturation rescales the field, shell near the goal: Sec. 2.2, Fig. 2C–F
$\alpha = \gamma < 1$ (recurrent decay equals the step cost; stability)	horizon and latency across α: Fig. 4D, App. M.1
Transition-gated BTSP, symmetric kernel ($\boldsymbol{S}$ vanishes on walls; $\boldsymbol{S} = \boldsymbol{P}_0$)	asymmetric kernel: Sec. 2.2, App. L.5; partial coverage: Fig. 4C
Centered local place kernel (readout follows $\nabla \log \psi^g$; raw rate suffices)	anisotropic kernels near walls: App. M.4
Fixed lazy rate ϵ, Eq. (9) (one operator for all goals and edits)	ϵ collapse; 38,556-state maps: Fig. 4F

Table 1: Assumptions, what each provides, and where each is relaxed or tested.

2.5 Assumptions and scope

Table 1 lists the assumptions behind the results and where each is relaxed or tested. The units of the theory are assemblies: the computation is unchanged under any full-rank recoding of the population, so overlapping non-negative codes, place-cell-like in form, preserve the field and the policy (App. M.6); one unit per location is a coordinate choice.

3 Learning the recurrent connectivity

3.1 Transition-gated symmetric plasticity

Theorem 1 requires that $\boldsymbol{S}$ vanishes on inaccessible transitions and that $\boldsymbol{P}_0$ satisfies detailed balance; both arise from BTSP during goal-free exploration. BTSP is triggered by dendritic plateau potentials and potentiates synapses whose presynaptic input fired within a few seconds of the plateau (Bittner et al., 2017; Magee and Grienberger, 2020); CA3 place fields emerge through a symmetric BTSP-like process at recurrent CA3–CA3 synapses (Li et al., 2024), consistent with broad symmetric spike-timing-dependent plasticity there (Mishra et al., 2016), and symmetric rules support stable predictive spatial codes in models (Keck et al., 2025). We model the induction kernel as symmetric, $k(\tau) = k(-\tau) \geq 0$, with τ the lag between pre- and postsynaptic firing within the BTSP window.

The seconds-long window lets BTSP bind firing at successively visited locations, in sequences that arise from physical exploration (McNaughton et al., 2006) and from theta sequences and replay (Johnson and Redish, 2007; Davidson et al., 2009; Wikenheiser and Redish, 2015; Pfeiffer and Foster, 2013; Widloski and Foster, 2022). Let $a_i(t) \geq 0$ be the firing rate of unit i and $\Xi_{ij}(t, \tau) \in \{0, 1\}$ a transition gate that equals one when units i and j fire in sequence at lag τ within the BTSP window centered at t, marking an accessible transition. The accumulated weight is

$$W_{ij}^{\mathrm{tr}} = \sum_{\tau} k(\tau)\, \mathbb{E}\big[a_i(t)\, a_j(t+\tau)\, \Xi_{ij}(t, \tau)\big]. \tag{8}$$

Walls permit no traversal, and replay respects the same barriers (Gupta et al., 2010; Wu and Foster, 2014; Widloski and Foster, 2022), so $W_{ij}^{\mathrm{tr}} = 0$ on these pairs. Traversals and replay sample accessible transitions in either direction (Foster and Wilson, 2006; Diba and Buzsáki, 2007; Gupta et al., 2010), so with the symmetric kernel $\boldsymbol{W}^{\mathrm{tr}}$ is symmetric in expectation; the asymmetric residue left by finite sampling and goal-directed behavior is controlled by a single scale, $\chi = \alpha \|\boldsymbol{E}_{\mathrm{asym}}\|/(1-\alpha)$ (Apps. F.2 and L.5). We turn $\boldsymbol{W}^{\mathrm{tr}}$ into the transition kernel by distributing each row's outgoing weight among accessible neighbors and keeping the remainder as the probability of staying put:

$$P_0(j \mid i) = \epsilon\, W_{ij}^{\mathrm{tr}}\ (j \neq i), \qquad P_0(i \mid i) = 1 - \epsilon \sum_{j \neq i} W_{ij}^{\mathrm{tr}}, \qquad 0 < \epsilon \leq 1/\max_i \textstyle\sum_{j \neq i} W_{ij}^{\mathrm{tr}}, \tag{9}$$

with the lazy rate ϵ fixed once per environment and kept across later edits. When $\boldsymbol{W}^{\mathrm{tr}}$ is symmetric, so is $\boldsymbol{P}_0$; detailed balance holds with respect to the uniform measure, $\boldsymbol{S} = \boldsymbol{P}_0$, and Theorem 1 applies (nonuniform measures: App. F.4). One set of synapses thus carries the map and generates the network's spontaneous flow through it.

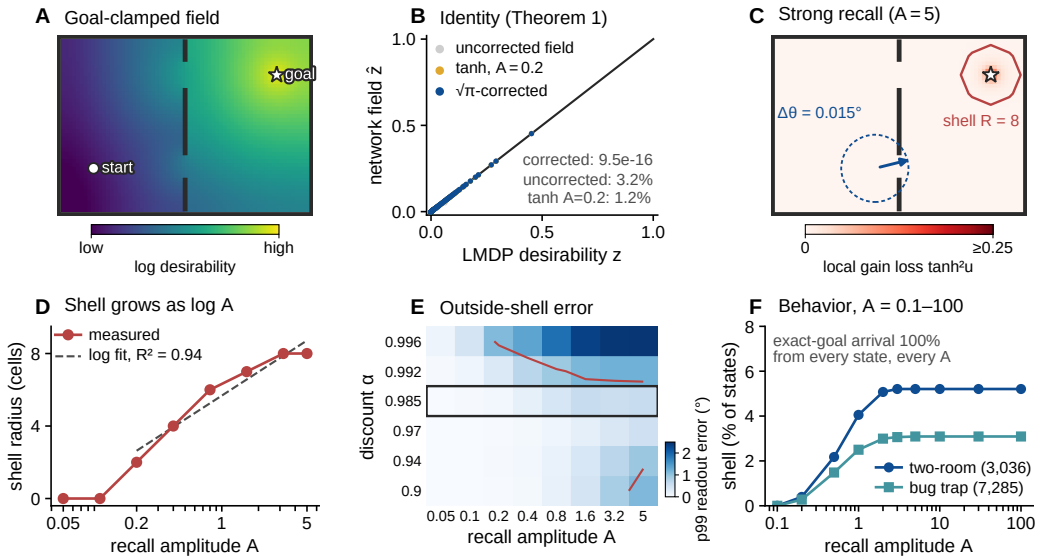


Figure 2: **The relaxation reproduces the LMDP field, and the readout is amplitude-robust.** (**A**) Two-room maze; color: log desirability at $\alpha = 0.985$. (**B**) Network steady state in the LMDP coordinate against the LMDP solution $\boldsymbol{z}$: the $\sqrt{\boldsymbol{\pi}}$-corrected field lies on the identity (9.5×10^{-16}); uncorrected, it deviates by 3.2%; weak nonlinear recall ($A = 0.2$) by 1.2%. (**C**) Strong recall ($A = 5$) confines the loss of local gain to a shell of radius 8 cells around the goal; at a probe outside the shell the readout error is $0.015°$. (**D**) Shell radius grows as $\log A$ ($R^2 = 0.94$). (**E**) Outside-shell 99th-percentile readout error across α and A; the row $\alpha = 0.985$ (box) stays below $1°$ (contour). (**F**) For $A = 0.1$–100 the shell covers at most 5.2% of the two-room maze and 3.1% of the bug trap, and every rollout from every start reaches the goal.

3.2 Local environmental change as a low-rank update

The plasticity that wrote the map remains active after the environment changes. When a corridor opens or a doorway closes, transition events at the affected synapses resume or cease and the corresponding entries of $\boldsymbol{W}^{\mathrm{tr}}$ update, while entries away from the change keep their values; a few traversals through the changed region update the connectivity for all future plans. Editing k synapses changes k entries of $\boldsymbol{S}$, and the change in goal-locked fields decomposes (App. K.4) into a rank-$\mathcal{O}(k)$ propagation pattern plus a per-goal scaling that keeps each clamp at its amplitude. The readout is invariant to this scaling, so a few directions of synaptic change explain the rerouting toward every goal (Sec. 4.2).

4 Experiments

We tested the theory on grid mazes, each open cell coded by one unit with eight-neighbor connectivity, and on irregular graphs of place centers in continuous arenas (App. L.1; code in App. N).

4.1 The relaxation reproduces the LMDP field, and the readout is amplitude-robust

Theorem 1 predicts that $\boldsymbol{\psi}^g$ equals the desirability $\boldsymbol{z}$ up to the factor $\sqrt{\pi_i/\pi_g}$. To make this factor visible we used a two-room maze whose narrow doorways make the spontaneous occupancy mildly non-uniform (Fig. 2A). At $\alpha = 0.985$ the corrected field $\widehat{z}_i := \sqrt{\pi_g/\pi_i}\,\psi_i^g$ matched $\boldsymbol{z}$ to machine precision, and the uncorrected field differed by 3.2%, the size of the correction (Fig. 2B). The agreement held across all α tested and on a 7,285-state bug-trap maze, where the branch-current policy matched μ^* to 2.8×10^{-16} in total variation (App. L.2).

With $\phi = \tanh$ and recall amplitude $A := \bar{u}_g$ from 0.05 to 5, the strongest recall deviated from the linear prediction inside a shell of about eight units around the goal (Fig. 2C); outside it the readout direction matched the linear prediction to $0.15°$ on average and $0.60°$ at the 99th percentile. The shell radius grew as $\log A$ (Fig. 2D), and a joint sweep over α and A kept the outside-shell error

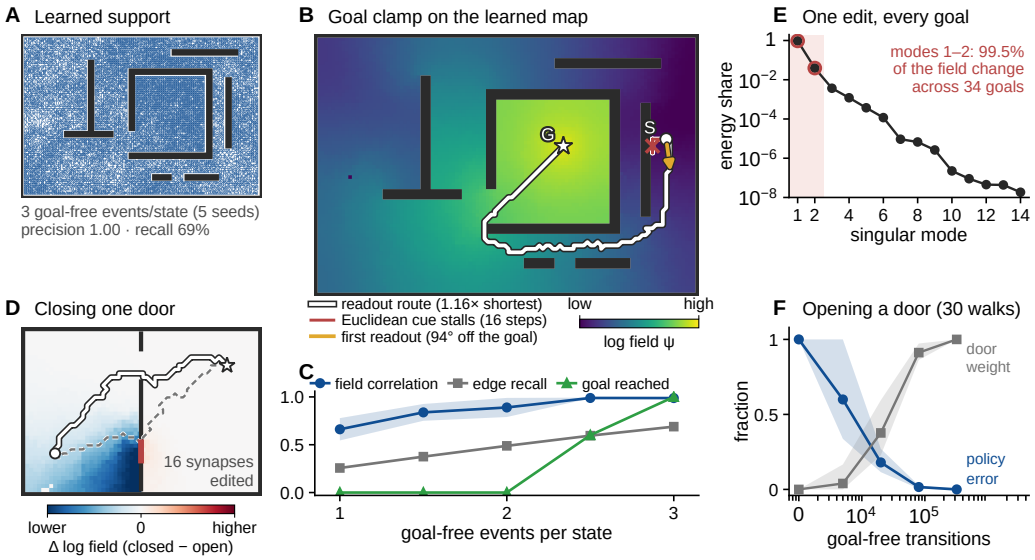


Figure 3: **Goal-free BTSP solves the bug trap, and a local edit reroutes plans globally.** (**A**) Connectivity learned by goal-free BTSP after 3 transitions per unit (precision 1.00, recall 69%; 5 runs). (**B**) Goal-clamped field (color) and population-vector route (white, 1.16× the shortest legal route); the first readout points 94° away from the Euclidean direction, and an agent following the Euclidean cue stalls at the wall after 16 steps. (**C**) Field correlation with the true-layout field, edge recall, and goal arrival against transitions per unit. (**D**) Two-door maze: closing the near door edits the 16 crossing synapses, and the next relaxation reroutes through the far door (white); color: field difference, closed minus open. (**E**) Singular spectrum of the field differences across 34 recalled goals: the first two modes carry 99.5% of the energy. (**F**) A newly opened doorway is learned by goal-free exploration alone: the door weight is written and the policy error relative to the true open map falls to zero in all 30 trajectories.

below 1° along the main row (Fig. 2E). From $A = 0.1$ to 100 on both mazes the shell never exceeded 5.2% of the states, greedy rollouts from every state reached the goal in all 103,190 cases (Fig. 2F), and stochastic rollouts started inside the shell, where the goal is adjacent, were captured by it in all 33,792 cases (App. M.2).

4.2 Goal-free BTSP solves the bug trap, and a local edit reroutes plans globally

In the bug-trap maze the goal sits a short Euclidean distance from the start, yet the only legal route detours through a distant entrance (Fig. 3B). Each sampled transition strengthened the forward and reverse synapses equally (App. L.3). After 3 transitions per unit the learned connectivity recovered 69% of the accessible connections with no false positives (Fig. 3A), and its goal-clamped field correlated at 0.988 ± 0.003 (mean ± SEM, 5 runs) with the field of the true layout. Every run reached the goal by following the population vector, with median route length 1.16× the shortest legal route; the first readout already pointed 94° away from the Euclidean direction, toward the distant entrance (Fig. 3B,C). On the same map a Dijkstra oracle gives route ratio 1.00 and the raw-rate readout of Proposition 1 gives 1.03 (App. M.11).

To test the low-rank rerouting of Sec. 3.2, we trained BTSP on a two-door maze (2 transitions per unit). Closing the near door updated only the 16 synapses that cross it, and the next relaxation rerouted plans through the far door without further learning (Fig. 3D). Stacking the field changes for 34 recalled goals into a matrix, the first two singular modes carried 99.5% of the energy (Fig. 3E): one local edit reroutes plans toward every goal through a rank-2 correction (App. M.7). The converse edit is learned online: when a doorway opens, goal-free exploration writes the door synapses and the policy error relative to the true open map falls to zero in all 30 trajectories (Fig. 3F); a closed

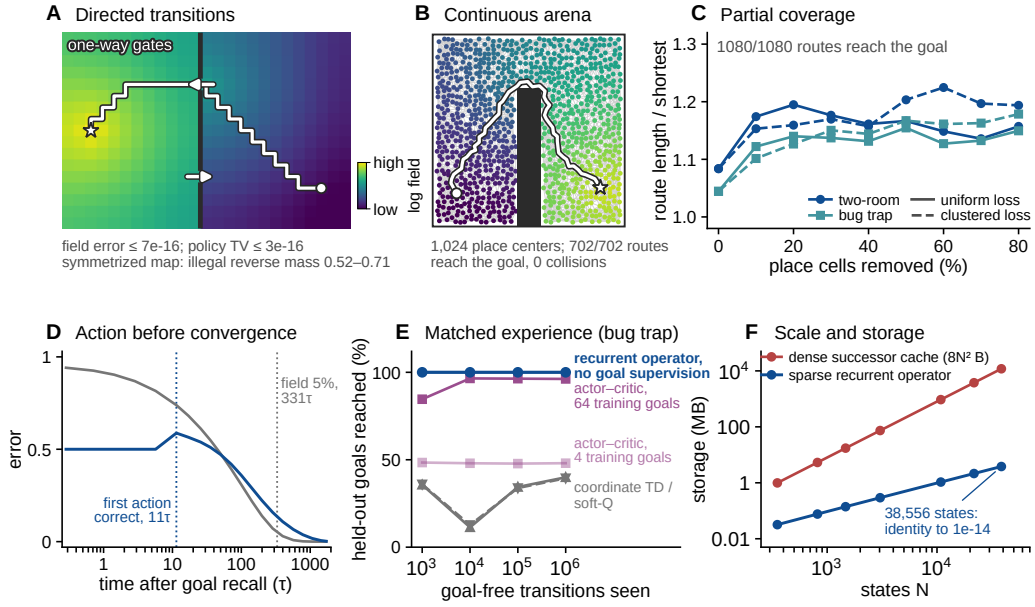


Figure 4: **Scope: directed maps, continuous arenas, partial coverage, anytime readiness, matched experience, and scale.** (**A**) A map with two one-way gates: the directed realization of Corollary 1 reproduces the directed LMDP field to machine precision; a symmetrized map assigns 0.52–0.71 of its mass to illegal reverse crossings. (**B**) Irregular graph of $1{,}024$ place centers in a continuous arena with barriers: $702/702$ routes reach the goal without collision. (**C**) Random or clustered loss of place cells: all $1{,}080$ routes reach the goal through 80% dropout on both mazes. (**D**) Anytime readiness (bug trap, $\alpha = 0.993$): the current action is correct after 11τ; the whole field converges to 5% after 331τ. (**E**) Matched experience on the bug trap, held-out goals: the recurrent operator reaches 100% of stochastic arrivals from 10^3 transitions on; soft-Q and goal-coordinate TD learners about 40% at 10^6; a goal-conditioned actor–critic 96% with 64 training goals, 48% with 4. (**F**) Storage grows linearly with the number of states through $38{,}556$ states, where the field identity holds to 10^{-14}; a dense all-goal cache grows quadratically.

doorway is detected by blocked transitions, and a gate on blocked attempts rerouted $8/8$ streaming episodes (App. M.7).

4.3 Scope: directed maps, continuous arenas, partial coverage, latency, and baselines

Figure 4 collects tests of the assumptions in Table 1. On a 425-state map with two one-way gates, the directed realization of Corollary 1 matched the directed LMDP field to 6.8×10^{-16} and its policy to 2.8×10^{-16} over 18 conditions, whereas a symmetrized map assigned up to 0.71 of its probability mass to an illegal reverse crossing (Fig. 4A; acquisition from random walks in App. M.5). On irregular graphs of place centers in continuous arenas, $702/702$ routes reached the goal without touching a barrier (Fig. 4B), and refining the discretization converged monotonically to the screened-Poisson continuum problem (App. M.9). Removing up to 80% of place cells, uniformly or in clusters, left $1{,}080/1{,}080$ routes successful (Fig. 4C), and an operator written from scratch after dropout reached the goal in $689/720$ routes at 50% loss (App. M.10). Heterogeneous unit gains, activity noise, and synaptic jitter are covered in App. M.3.

The transient is anytime: the local action depends on the local gradient of a partially relaxed field, so it is correct long before the whole field converges. On the bug trap at $\alpha = 0.993$ the current action is ready after 11τ while the field converges to 5% after 331τ (Fig. 4D), and across 24 conditions with $\alpha \geq 0.985$ the median speedup was $82\times$ (App. M.1). Latency is set by the spectral gap of $\boldsymbol{I} - \alpha \boldsymbol{S}$ (measured slow rate within 10^{-12} of the spectral prediction), and the field of a nearby goal or the pre-edit field serves as a warm start.

At matched experience, the operator was compared with goal-conditioned learners trained on the same transition streams (Fig. 4E; App. M.11). On the bug trap with held-out goals it reached $5{,}120/5{,}120$ stochastic arrivals at every budget from 10^3 to 10^6 transitions; a tabular soft-Q learner and goal-coordinate TD value learners reached 39–40% at 10^6, and a goal-conditioned soft actor–critic 96% with 64 training goals and 48% with 4. Successor representations built from the same resolvent coordinates agree with our fields and behave the same; the difference is storage: the operator keeps $\mathcal{O}(N)$ synapses and answers any goal by one relaxation, whereas a dense all-goal cache grows as N^2 ($252\times$ at $3{,}036$ states; Fig. 4F), and the identities held through $38{,}556$ states (App. M.12).

5 Discussion

We presented a hippocampal recurrent network that plans obstacle-avoiding routes through its own relaxation dynamics: BTSP writes the recurrent weights during goal-free exploration, goal recall relaxes the same circuit into the LMDP desirability field with the optimal policy in its branch currents, and a place-tuned readout converts the field into a direction of travel.

Position in the literature. Models that let a recurrent network settle toward a plan divide by where the goal enters. When the source sits at the current state, as in recurrent successor representations (Fang et al., 2023) and Endotaxis (Zhang et al., 2024), the relaxed state is a predictive field and the goal enters through a readout. When the goal enters as a boundary condition, as in wavefront planning (Ponulak and Hopfield, 2013) and the spacetime attractor of Jensen et al. (2026), the relaxed state supports control as an arrival time or a trajectory over states and future times. Our claim is the identity: one recurrent operator stores the map and relaxes into the provably optimal plan, with the policy in its own currents and the operator written by an observed plasticity rule. The two boundary-driven models are complementary: the desirability field is an exponentially weighted compression of the spacetime attractor's future time slices, and prefrontal circuits could supply the goal boundary condition that the hippocampal network evaluates on its fine-grained graph (Ito et al., 2015).

Alternative goal mechanisms. Goal-modulated connectivity would need goal-specific synapses; the model has none, and the first recall of a new goal is already correct. A learned goal-conditioned readout needs per-goal supervision; ours is zero-shot, and the learned models of Fig. 4E need tens of training goals to approach it. Decision-time replay predicts latency that grows with path length; here it is set by the spectral gap.

Predictions. Four consequences are falsifiable in recordings during goal recall. (i) Place-field centers stay stable during recall and acquire a multiplicative goal-dependent gain proportional to $\psi^g(\boldsymbol{x})$, larger for cells with shorter legal routes to the goal, consistent with the stable CA1 fields of Duvelle et al. (2021); successor accounts predict fields reshaped toward reward. (ii) The first recall of a novel goal is already correct. (iii) Decision latency is set by the spectral gap of the recurrent operator and does not grow with path length. (iv) A local change of the environment produces a low-rank change of the goal-locked fields, shared across goals and concentrated near the edit, detectable within a few traversals in barrier-change paradigms (Widloski and Foster, 2022).

Limitations. The model describes a hippocampal population without committing to a subfield; BTSP-like signatures have been reported in several (Bittner et al., 2017; Magee and Grienberger, 2020; Qian et al., 2025; Li et al., 2024), and intracellular recordings during planning pauses could test which one hosts it. The transition gate Ξ_{ij} is idealized: self-motion signals and internally generated sequences likely cooperate to restrict BTSP to accessible transitions, and identifying this mechanism, with the BTSP kernel and the inhibitory gain in vivo, is a concrete next step. The exact branch-current policy needs action-labelled access to individual synaptic currents; the population vector is the implemented approximation (App. M.4).

Outlook. The same idea extends beyond spatial navigation. Hippocampal–entorhinal maps have been reported for sensory continua, social hierarchies, and abstract task structures (Behrens et al., 2018; Garvert et al., 2023; Courellis et al., 2024; Tacikowski et al., 2024), where a "step" is a movement in space, a shift of attention, or an imagined step along a chain of associations (state-dependent costs: App. M.8). Behind the technical equivalence lies a simple picture of cognition. Each synaptic weight S_{ij} records one transition, and the agent's actions decide which transitions exist. States are then defined by the transitions they participate in: the transitions come first, and the states emerge from them. Goal-directed cognition becomes the relaxation of this learned web of transitions, and the planning is the relaxation itself.

Acknowledgments and Disclosure of Funding

This work received no third-party funding. The authors declare no competing interests.

References

Amaral, D.G. & Witter, M.P. (1989). The three-dimensional organization of the hippocampal formation: A review of anatomical data. Neuroscience 31, 571–591.

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. Biological Cybernetics 27, 77–87.

Ambrose, R.E., Pfeiffer, B.E. & Foster, D.J. (2016). Reverse replay of hippocampal place cells is uniquely modulated by changing reward. Neuron 91, 1124–1136.

Aoki, Y., Igata, H., Ikegaya, Y. & Sasaki, T. (2019). The integration of goal-directed signals onto spatial maps of hippocampal place cells. Cell Reports 27, 1516–1527.e5.

Bakermans, J.J.W., Warren, J., Whittington, J.C.R. & Behrens, T.E.J. (2025). Constructing future behavior in the hippocampal formation through composition and replay. Nature Neuroscience 28, 1061–1072.

Banino, A. et al. (2018). Vector-based navigation using grid-like representations in artificial agents. Nature 557, 429–433.

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

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

Bono, J., Zannone, S., Pedrosa, V. & Clopath, C. (2023). Learning predictive cognitive maps with spiking neurons during behavior and replays. eLife 12, e80671.

Burak, Y. & Fiete, I.R. (2009). Accurate path integration in continuous attractor network models of grid cells. PLOS Computational Biology 5, e1000291.

Buzsáki, G. (2015). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. Hippocampus 25, 1073–1188.

Connolly, C.I., Burns, J.B. & Weiss, R. (1990). Path planning using Laplace's equation. Proceedings of the IEEE International Conference on Robotics and Automation, 2102–2106.

Connolly, C.I. & Grupen, R.A. (1993). The applications of harmonic functions to robotics. Journal of Robotic Systems 10, 931–946.

Connolly, C.I. (1997). Harmonic functions and collision probabilities. The International Journal of Robotics Research 16, 497–507.

Corneil, D.S. & Gerstner, W. (2015). Attractor network dynamics enable preplay and rapid path planning in maze-like environments. Advances in Neural Information Processing Systems 28, 1684–1692.

de Cothi, W. & Barry, C. (2020). Neurobiological successor features for spatial navigation. Hippocampus 30, 1347–1355.

de Cothi, W., Nyberg, N., Griesbauer, E.-M., Ghanamé, C., Zisch, F., Lefort, J.M., Fletcher, L., Newton, C., Renaudineau, S., Bendor, D., Grieves, R., Duvelle, É., Barry, C. & Spiers, H.J. (2022). Predictive maps in rats and humans for spatial navigation. Current Biology 32, 3676–3689.e5.

Courellis, H.S. et al. (2024). Abstract representations emerge in human hippocampal neurons during inference. Nature 632, 841–849.

Cueva, C.J. & Wei, X.-X. (2018). Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. International Conference on Learning Representations.

Davidson, T.J., Kloosterman, F. & Wilson, M.A. (2009). Hippocampal replay of extended experience. Neuron 63, 497–507.

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. Neural Computation 5, 613–624.

Diba, K. & Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. Nature Neuroscience 10, 1241–1242.

Dordek, Y., Soudry, D., Meir, R. & Derdikman, D. (2016). Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. eLife 5, e10094.

Duvelle, É. et al. (2021). Hippocampal place cells encode global location but not connectivity in a complex space. Current Biology 31, 1221–1233.e9.

El-Gaby, M. et al. (2024). A cellular basis for mapping behavioural structure. Nature 636, 671–680.

Erdem, U.M. & Hasselmo, M. (2012). A goal-directed spatial navigation model using forward trajectory planning based on grid cells. European Journal of Neuroscience 35, 916–931.

Fang, C., Aronov, D., Abbott, L.F. & Mackevicius, E.L. (2023). Neural learning rules for generating flexible predictions and computing the successor representation. eLife 12, e80680.

Forli, A. et al. (2025). Replay and representation dynamics in the hippocampus of freely flying bats. Nature 645, 974–980.

Foster, D.J. & Wilson, M.A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature 440, 680–683.

Garvert, M.M., Saanum, T., Schulz, E., Schuck, N.W. & Doeller, C.F. (2023). Hippocampal spatio-predictive cognitive maps adaptively guide reward generalization. Nature Neuroscience 26, 615–626.

Gauthier, J.L. & Tank, D.W. (2018). A dedicated population for reward coding in the hippocampus. Neuron 99, 179–193.e7.

George, D., Rikhye, R.V., Gothoskar, N., Guntupalli, J.S., Dedieu, A. & Lázaro-Gredilla, M. (2021). Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. Nature Communications 12, 2392.

George, T.M., de Cothi, W., Stachenfeld, K.L. & Barry, C. (2023). Rapid learning of predictive maps with STDP and theta phase precession. eLife 12, e80663.

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

Gupta, A.S., van der Meer, M.A.A., Touretzky, D.S. & Redish, A.D. (2010). Hippocampal replay is not a simple function of experience. Neuron 65, 695–705.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E.I. (2005). Microstructure of a spatial map in the entorhinal cortex. Nature 436, 801–806.

Haga, T. & Fukai, T. (2018). Recurrent network model for learning goal-directed sequences through reverse replay. eLife 7, e34171.

He, L., Wang, X., Zhang, J. et al. (2026). Human hippocampal ripples coordinate planning sequences and compositional representations in neocortex. Nature Neuroscience 29, 1711–1721.

Hok, V., Lenck-Santini, P.-P., Roux, S., Save, E., Muller, R.U. & Poucet, B. (2007). Goal-related activity in hippocampal place cells. Journal of Neuroscience 27, 472–482.

Hopfield, J.J. & Tank, D.W. (1985). "Neural" computation of decisions in optimization problems. Biological Cybernetics 52, 141–152.

Ishizuka, N., Weber, J. & Amaral, D.G. (1990). Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. Journal of Comparative Neurology 295, 580–623.

Ito, H.T. et al. (2015). A prefrontal-thalamo-hippocampal circuit for goal-directed spatial navigation. Nature 522, 50–55.

Jadhav, S.P., Kemere, C., German, P.W. & Frank, L.M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. Science 336, 1454–1458.

Jensen, K.T., Hennequin, G. & Mattar, M.G. (2024). A recurrent network model of planning explains hippocampal replay and human behavior. Nature Neuroscience 27, 1340–1348.

Jensen, K.T., Doohan, P.T., Sablé-Meyer, M., Reinert, S., Baram, A.B., Sahani, M., Akam, T. & Behrens, T.E.J. (2026). A mechanistic theory of planning in prefrontal cortex. eLife 15, RP109757.

Johnson, A. & Redish, A.D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. Journal of Neuroscience 27, 12176–12189.

Joo, H.R. & Frank, L.M. (2018). The hippocampal sharp wave-ripple in memory retrieval for immediate use and consolidation. Nature Reviews Neuroscience 19, 744–757.

Kappen, H.J. (2005). Linear theory for control of nonlinear stochastic systems. Physical Review Letters 95, 200201.

Keck, J., Barry, C., Doeller, C.F. & Jost, J. (2025). Impact of symmetry in local learning rules on predictive neural representations and generalization in spatial navigation. PLOS Computational Biology 21, e1013056.

Khatib, O. (1986). Real-time obstacle avoidance for manipulators and mobile robots. The International Journal of Robotics Research 5, 90–98.

Lee, L., Parisotto, E., Chaplot, D.S., Xing, E.P. & Salakhutdinov, R. (2018). Gated path planning networks. Proceedings of Machine Learning Research 80, 2947–2955.

Lever, C., Wills, T., Cacucci, F., Burgess, N. & O'Keefe, J. (2002). Long-term plasticity in hippocampal place-cell representation of environmental geometry. Nature 416, 90–94.

Levine, S. (2018). Reinforcement learning and control as probabilistic inference: Tutorial and review. arXiv preprint arXiv:1805.00909.

Li, Y., Briguglio, J.J., Romani, S. & Magee, J.C. (2024). Mechanisms of memory-supporting neuronal dynamics in hippocampal area CA3. Cell 187, 6804–6819.e21.

Liu, Y., Dolan, R.J., Kurth-Nelson, Z. & Behrens, T.E.J. (2019). Human replay spontaneously reorganizes experience. Cell 178, 640–652.e14.

Liu, Y., Mattar, M.G., Behrens, T.E.J., Daw, N.D. & Dolan, R.J. (2021). Experience replay is associated with efficient nonlocal learning. Science 372, eabf1357.

London, M. & Häusser, M. (2005). Dendritic computation. Annual Review of Neuroscience 28, 503–532.

Machado, M.C., Bellemare, M.G. & Bowling, M. (2017). A Laplacian framework for option discovery in reinforcement learning. Proceedings of Machine Learning Research 70, 2295–2304.

Madar, A.D., Jiang, A., Dong, C. & Sheffield, M.E.J. (2025). Synaptic plasticity rules driving representational shifting in the hippocampus. Nature Neuroscience 28, 848–860.

Magee, J.C. & Grienberger, C. (2020). Synaptic plasticity forms and functions. Annual Review of Neuroscience 43, 95–117.

Mahadevan, S. & Maggioni, M. (2007). Proto-value functions: A Laplacian framework for learning representation and control in Markov decision processes. Journal of Machine Learning Research 8, 2169–2231.

Marr, D. (1971). Simple memory: A theory for archicortex. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 262, 23–81.

Mattar, M.G. & Daw, N.D. (2018). Prioritized memory access explains planning and hippocampal replay. Nature Neuroscience 21, 1609–1617.

McHugh, T.J., Blum, K.I., Tsien, J.Z., Tonegawa, S. & Wilson, M.A. (1996). Impaired hippocampal representation of space in CA1-specific NMDAR1 knockout mice. Cell 87, 1339–1349.

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

van der Meer, M.A.A. & Bendor, D. (2025). Awake replay: Off the clock but on the job. Trends in Neurosciences 48, 257–267.

Mehta, M.R., Quirk, M.C. & Wilson, M.A. (2000). Experience-dependent asymmetric shape of hippocampal receptive fields. Neuron 25, 707–715.

Milstein, A.D., Li, Y., Bittner, K.C., Grienberger, C., Soltesz, I., Magee, J.C. & Romani, S. (2021). Bidirectional synaptic plasticity rapidly modifies hippocampal representations. eLife 10, e73046.

Mishra, R.K., Kim, S., Guzman, S.J. & Jonas, P. (2016). Symmetric spike timing-dependent plasticity at CA3–CA3 synapses optimizes storage and recall in autoassociative networks. Nature Communications 7, 11552.

Mitchell, S.J. & Silver, R.A. (2003). Shunting inhibition modulates neuronal gain during synaptic excitation. Neuron 38, 433–445.

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

Muller, R.U., Stead, M. & Pach, J. (1996). The hippocampus as a cognitive graph. Journal of General Physiology 107, 663–694.

Nakazawa, K. et al. (2002). Requirement for hippocampal CA3 NMDA receptors in associative memory recall. Science 297, 211–218.

Nyberg, N., Duvelle, É., Barry, C. & Spiers, H.J. (2022). Spatial goal coding in the hippocampal formation. Neuron 110, 394–422.

O'Keefe, J. & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. Brain Research 34, 171–175.

O'Keefe, J. & Nadel, L. (1978). The Hippocampus as a Cognitive Map. Oxford: Clarendon Press.

O'Keefe, J. & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. Nature 381, 425–428.

Ormond, J. & O'Keefe, J. (2022). Hippocampal place cells have goal-oriented vector fields during navigation. Nature 607, 741–746.

Pfeiffer, B.E. & Foster, D.J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. Nature 497, 74–79.

Piray, P. & Daw, N.D. (2021). Linear reinforcement learning in planning, grid fields, and cognitive control. Nature Communications 12, 4942.

Poirazi, P., Brannon, T. & Mel, B.W. (2003). Pyramidal neuron as two-layer neural network. Neuron 37, 989–999.

Ponulak, F. & Hopfield, J.J. (2013). Rapid, parallel path planning by propagating wavefronts of spiking neural activity. Frontiers in Computational Neuroscience 7, 98.

Qian, F.K., Li, Y. & Magee, J.C. (2025). Mechanisms of experience-dependent place-cell referencing in hippocampal area CA1. Nature Neuroscience 28, 1486–1496.

Rimon, E. & Koditschek, D.E. (1992). Exact robot navigation using artificial potential functions. IEEE Transactions on Robotics and Automation 8, 501–518.

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

Sagiv, Y., Akam, T., Witten, I.B. & Daw, N.D. (2025). Between planning and map building: Prioritizing replay when future goals are uncertain. Neuron 113, 4278–4292.e5.

Sammons, R.P. et al. (2024). Structure and function of the hippocampal CA3 module. Proceedings of the National Academy of Sciences of the United States of America 121, e2312281120.

Samsonovich, A. & McNaughton, B.L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. Journal of Neuroscience 17, 5900–5920.

Sarel, A., Finkelstein, A., Las, L. & Ulanovsky, N. (2017). Vectorial representation of spatial goals in the hippocampus of bats. Science 355, 176–180.

Schuck, N.W. & Niv, Y. (2019). Sequential replay of nonspatial task states in the human hippocampus. Science 364, eaaw5181.

Schwartenbeck, P. et al. (2023). Generative replay underlies compositional inference in the hippocampal-prefrontal circuit. Cell 186, 4885–4897.e14.

Silver, R.A. (2010). Neuronal arithmetic. Nature Reviews Neuroscience 11, 474–489.

Sosa, M., Plitt, M.H. & Giocomo, L.M. (2025). A flexible hippocampal population code for experience relative to reward. Nature Neuroscience 28, 1497–1509.

Stachenfeld, K.L., Botvinick, M.M. & Gershman, S.J. (2017). The hippocampus as a predictive map. Nature Neuroscience 20, 1643–1653.

Tacikowski, P., Kalender, G., Ciliberti, D. & Fried, I. (2024). Human hippocampal and entorhinal neurons encode the temporal structure of experience. Nature 635, 160–167.

Tamar, A., Wu, Y., Thomas, G., Levine, S. & Abbeel, P. (2016). Value iteration networks. Advances in Neural Information Processing Systems 29, 2154–2162.

Todorov, E. (2006). Linearly-solvable Markov decision problems. Advances in Neural Information Processing Systems 19, 1369–1376.

Todorov, E. (2009). Efficient computation of optimal actions. Proceedings of the National Academy of Sciences of the United States of America 106, 11478–11483.

Tolman, E.C. (1948). Cognitive maps in rats and men. Psychological Review 55, 189–208.

Treves, A. & Rolls, E.T. (1994). Computational analysis of the role of the hippocampus in memory. Hippocampus 4, 374–391.

Tsodyks, M. (1999). Attractor neural network models of spatial maps in hippocampus. Hippocampus 9, 481–489.

Whittington, J.C.R., Muller, T.H., Mark, S., Chen, G., Barry, C., Burgess, N. & Behrens, T.E.J. (2020). The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalization in the hippocampal formation. Cell 183, 1249–1263.e23.

Whittington, J.C.R., McCaffary, D., Bakermans, J.J.W. & Behrens, T.E.J. (2022). How to build a cognitive map. Nature Neuroscience 25, 1257–1272.

Widloski, J. & Foster, D.J. (2022). Flexible rerouting of hippocampal replay sequences around changing barriers in the absence of global place field remapping. Neuron 110, 1547–1558.e8.

Wikenheiser, A.M. & Redish, A.D. (2015). Hippocampal theta sequences reflect current goals. Nature Neuroscience 18, 289–294.

Wu, X. & Foster, D.J. (2014). Hippocampal replay captures the unique topological structure of a novel environment. Journal of Neuroscience 34, 6459–6469.

Zhang, T., Rosenberg, M., Jing, Z., Perona, P. & Meister, M. (2024). Endotaxis: A neuromorphic algorithm for mapping, goal-learning, navigation, and patrolling. eLife 12, RP84141.

Zhou, X. et al. (2026). Human hippocampal ripples prioritize model-based learning. Neuron 114, 3111–3125.e7.

Zuo, J., He, Y., Zhang, W., Fang, F. & Wu, S. (2026). From representation to action: A unified Laplacian framework for spatial representation and path planning. Proceedings of Machine Learning Research 306, 169030–169045.

Appendix Contents

A Related work **19**

B Notation and standing assumptions **21**

C Linearization of the recurrent population **23**
C.1 Expansion around the goal-free fixed point 23
C.2 Homogeneous-gain scalar-discount reduction 24
C.3 Diagonal-gain operator 24

D LMDP desirability and exact linear network equivalence **25**
D.1 Bellman equation with KL control cost 25
D.2 Linear desirability equation 26
D.3 Exact linear network–LMDP theorem 26
D.4 Directed transitions: diagonal similarity 27

E Reversibility, learned transition kernels, and normalization **28**
E.1 Symmetric network coordinate 28
E.2 Transition-gated symmetric accumulation 28
E.3 Lazy normalization 29
E.4 Standard random-walk normalization 29

F Finite sampling, nonuniform π, and nonreversible perturbations **30**
F.1 Reversible projection of empirical transition counts 30
F.2 Perturbation by irreversible sampling error 30
F.3 Perturbation of a reversible stationary measure 31
F.4 Readout bias from nonuniform stationary measure 31

G Local place-kernel readout **32**
G.1 Continuous local moment expansion 32
G.2 Centered-kernel correction 33
G.3 Reading the raw rate 33
G.4 Accessible directions 34

H Relation between the readout and the LMDP optimal policy **34**

I Nonlinear perturbation under strong goal recall **35**
I.1 Nonlinear steady-state equation 35
I.2 Linear exponential envelope 35
I.3 Small-gain control of the nonlinear solution 36
I.4 Far-field amplitude projection 36

J Weak-cost limit and Green functions **37**
J.1 Screened Dirichlet operator 38
J.2 Boundary-driven weak-cost limit 38
J.3 Dirichlet source-response Green function 39
J.4 Full-graph mean-zero Green function 39

K Local environmental edits and low-rank structure **40**
K.1 Rank of a local connectivity edit 40
K.2 Fixed-goal Dirichlet update 41
K.3 Shared full-resolvent update 41
K.4 Unit-clamped all-goal fields 42

L Numerical experiment details **43**
L.1 Common graph, field, and readout conventions 43
L.2 Equivalence and strong-recall robustness 44
L.3 Goal-free BTSP learning and local low-rank remapping 45
L.4 Weak-cost Green-function limit 46

L.5 Asymmetric BTSP perturbation and bilateral replay 47

M Additional experiments: scope, robustness, and comparisons **50**
M.1 Realized relaxation dynamics, latency, and anytime readiness 51
M.2 Strong recall: distance-resolved shells and behavior 52
M.3 Gain heterogeneity, activity noise, and operator perturbations 52
M.4 Exact policy readout and anisotropic population kernels 52
M.5 Directed transitions 53
M.6 Overlapping population codes 53
M.7 Local opening and closure events and transition-support errors 53
M.8 State-dependent costs 54
M.9 Irregular place-center graphs and the continuum limit 54
M.10 Partial coverage by place cells 54
M.11 Comparisons with cached and learned goal-conditioned models 55
M.12 Scale and the lazy rate 55

N Reproduction **55**

A Related work

Cognitive maps as learned transition structure. The cognitive map hypothesis began as a behavioral account of flexible navigation (Tolman, 1948; O'Keefe and Nadel, 1978). Neural correlates were later identified in hippocampal place cells, entorhinal grid cells, and experience-dependent spatial coding (O'Keefe and Dostrovsky, 1971; O'Keefe and Burgess, 1996; Hafting et al., 2005; McNaughton et al., 2006; McHugh et al., 1996; Mehta et al., 2000; Lever et al., 2002). The same principle has since been extended beyond physical space, to relational, temporal, and abstract task structure (Behrens et al., 2018; Whittington et al., 2020, 2022; George et al., 2021; El-Gaby et al., 2024; Courellis et al., 2024; Tacikowski et al., 2024; Garvert et al., 2023; Bakermans et al., 2025). Our model adopts this transition-centric view and goes one step further: the transition structure is encoded directly in the recurrent connectivity. Each recurrent synapse carries the probability of a locally accessible move; walls and blocked corridors leave the corresponding synapses at zero; and goal recall clamps a single neuron in the same population. Planning then emerges as the relaxation of the very circuit that stores the map.

Successor representations and predictive maps. Successor representations (SRs) store long-horizon expected occupancies under a reference policy and convert any reward or goal vector into a value function (Dayan, 1993; Russek et al., 2017; Momennejad et al., 2017; Gershman, 2018). This idea has been developed into hippocampal predictive-map accounts, in which place- and grid-like codes reflect the eigenstructure of a learned transition matrix (Stachenfeld et al., 2017; de Cothi and Barry, 2020; de Cothi et al., 2022; Piray and Daw, 2021; Bono et al., 2023; Fang et al., 2023; George et al., 2023). These models explain why hippocampal representations should carry predictive structure, while leaving open how the predictive computation is realized in a recurrent circuit. Our framework supplies this circuit: the predictive field is generated at recall time by a goal-clamped relaxation, and we prove it equals the desirability field of an LMDP (Theorem 1). The same recurrent connectivity serves all goals, with the choice of goal entering only as a boundary clamp at recall.

Linearly-solvable planning, harmonic navigation, and Laplacian methods. Linearly-solvable Markov decision processes (LMDPs) reformulate KL-regularized planning as a linear desirability equation (Kappen, 2005; Todorov, 2006, 2009; Levine, 2018). In robotics, potential fields, harmonic functions, and Laplace equations have long served as path-planning tools that generate obstacle-avoiding routes (Khatib, 1986; Connolly et al., 1990; Connolly and Grupen, 1993; Connolly, 1997; Rimon and Koditschek, 1992). In reinforcement learning, Laplacian eigenfunctions and proto-value functions provide geometry-aware bases for representation and option discovery (Mahadevan and Maggioni, 2007; Machado et al., 2017). In neuroscience and machine learning, Laplacian and spectral accounts have been linked to grid-cell coding and spatial abstraction (Dordek et al., 2016; Stachenfeld et al., 2017; Banino et al., 2018; Cueva and Wei, 2018; Zuo et al., 2026). These four lines treat the desirability equation, the harmonic potential, the Laplacian basis, and the cognitive map as four separate objects. Our framework joins them in one recurrent circuit: the same connectivity that stores transitions, under a goal clamp, produces a solution to the LMDP desirability equation; in the weak-cost limit, this solution converges to a Green's function of the Laplacian formed from the same connectivity.

Planning as network dynamics. Recurrent neural networks have long been proposed as computational substrates for pattern completion, attractor dynamics, and optimization (Amari, 1977; Hopfield and Tank, 1985; Tsodyks, 1999). In hippocampal-entorhinal models, continuous attractors support path integration and spatial codes (Samsonovich and McNaughton, 1997; Burak and Fiete, 2009), and spiking and recurrent network models can generate wavefronts, preplay events, and goal-directed sequences (Erdem and Hasselmo, 2012; Ponulak and Hopfield, 2013; Corneil and Gerstner, 2015; Haga and Fukai, 2018; Jensen et al., 2024). In machine learning, differentiable planning modules embed dynamic programming inside a neural network (Tamar et al., 2016; Lee et al., 2018). Our framework specifies a different role for the recurrent connectivity: the recurrent weights store the passive transition probabilities of the environment, the goal is delivered as a clamping current at a single neuron, and the steady state of the clamped circuit equals the LMDP desirability field. Planning therefore takes the form of a physical relaxation of the stored map, performed by the very circuit that holds the map; at decision time, the computation reduces to a single equilibration of the recurrent dynamics.

Replay, theta sequences, and internally generated trajectories. Hippocampal theta sequences and sharp-wave ripple replay express forward, reverse, and nonlocal trajectories at compressed timescales (Johnson and Redish, 2007; Foster and Wilson, 2006; Diba and Buzsáki, 2007; Davidson et al., 2009; Gupta et al., 2010; Pfeiffer and Foster, 2013; Wikenheiser and Redish, 2015; Buzsáki, 2015; Joo and Frank, 2018). These internally generated sequences have been linked to memory updating, planning, and reinforcement learning (Jadhav et al., 2012; Mattar and Daw, 2018; Liu et al., 2019, 2021; Schuck and Niv, 2019; Schwartenbeck et al., 2023; Sagiv et al., 2025; He et al., 2026; Zhou et al., 2026). Recent work also emphasizes that replay can serve offline learning, map construction, and experience restructuring, with its behavioral role varying across species and tasks (van der Meer and Bendor, 2025; Forli et al., 2025). Our framework assigns replay a mechanistic role compatible with these observations. Replay and theta sequences activate pairs of assemblies within the BTSP gating window, thereby writing or updating the corresponding recurrent transition probabilities. Once the connectivity is in place, the online action signal at each step is the goal-clamped relaxation field and its local log-gradient. The offline sampling that builds the map is therefore separated from the online relaxation that reads the map for action.

CA3 recurrence, BTSP, and local transition learning. CA3 has long been modeled as a recurrent associative network supporting pattern completion and memory retrieval (Marr, 1971; Treves and Rolls, 1994; Nakazawa et al., 2002). Anatomically, CA3 pyramidal cells are densely connected by recurrent collaterals (Amaral and Witter, 1989; Ishizuka et al., 1990; Sammons et al., 2024). At the plasticity level, behavioral-timescale synaptic plasticity (BTSP) creates new place fields within seconds of behavioral input (Bittner et al., 2017; Magee and Grienberger, 2020; Qian et al., 2025); recent CA3 recordings reveal BTSP-like dynamics in memory-supporting circuits (Li et al., 2024); and CA3-CA3 synapses can express broad, symmetric timing-dependent plasticity (Mishra et al., 2016). Our learning rule abstracts these observations into a transition-gated symmetric accumulation: when an accessible move or a replayed transition co-activates two assemblies within the BTSP window, the corresponding recurrent connection is strengthened symmetrically. Walls and blocked corridors leave the corresponding connections at zero, since they yield no transition events. Exact reversibility is an idealization that simplifies the equivalence theorem; the appendix treats finite sampling and direction-biased replay as perturbations, and bidirectional sampling projects the empirical transition statistics onto their reversible component.

Goal coding and a local population-vector readout. The hippocampal formation contains goal-, reward-, and vector-related signals that coexist with stable spatial coding (Hok et al., 2007; Gauthier and Tank, 2018; Aoki et al., 2019; Sarel et al., 2017; Ormond and O'Keefe, 2022; Nyberg et al., 2022; Sosa et al., 2025), and stable place fields can persist across changes in replay structure or task-dependent sequence content after environmental edits (Duvelle et al., 2021; Wu and Foster, 2014; Widloski and Foster, 2022). Our readout mechanism explains this coexistence by keeping location tuning and goal-locked modulation in separate components of the same firing rate. A downstream place-tuned cell receives two inputs: a local place-kernel input centered on the animal's current position, and a recurrent input from the upstream desirability field evaluated at the cell's preferred location. The two inputs combine multiplicatively on the dendrite, so the cell's place-field center is preserved while a goal-dependent gain shapes its rate. This dendritic gain interaction is supported by classical work on dendritic nonlinear integration and shunting gain control (Mitchell and Silver, 2003; Poirazi et al., 2003; London and Häusser, 2005; Silver, 2010). The model therefore predicts that, during recall, hippocampal cells retain their resting place-field centers while acquiring a subthreshold or rate-level modulation proportional to the desirability of those centers.

Local remapping and global rerouting. A computational consequence of representing the environment as a recurrent transition matrix is that local environmental changes correspond to local synaptic edits. When a doorway closes or a corridor opens, only the affected transition weights are revised; at the next goal recall, the relaxation propagates this local edit and updates the field for whichever goal is recalled. This provides a mechanistic interpretation for the flexible replay rerouting around barriers reported in Widloski and Foster (2022), and for the adaptive planning that coexists with stable place fields in Duvelle et al. (2021). Algebraically, a small synaptic edit is a low-rank perturbation of the recurrent connectivity, so the induced changes in goal-locked fields across all goals are captured by a small number of correction patterns. The map update is local, while its planning consequence is immediately global; this locality-globality coupling is a hallmark of the recurrent-relaxation account.

Planning as relaxation. A lineage of recurrent models lets a network settle toward a plan, and the models divide by where the goal enters. Muller et al. (1996) proposed that the hippocampus stores a cognitive graph whose synaptic resistances encode the distances between place fields; the graph holds sufficient information for navigation, and the optimal path is found by standard graph-search algorithms. Ponulak and Hopfield (2013) propagate a wavefront of spiking activity from the goal and read the plan from the synaptic changes it leaves behind; Corneil and Gerstner (2015) use attractor dynamics for preplay and rapid path planning; Haga and Fukai (2018) learn goal-directed sequences through reverse replay. Successor representations can be learned by spike-timing-dependent plasticity (George et al., 2023; Bono et al., 2023) and computed by recurrent dynamics (Fang et al., 2023); recurrent successor models and Endotaxis (Zhang et al., 2024) place the source at the current state, relax into a predictive field, and let the goal enter through a learned readout. The spacetime attractor of Jensen et al. (2026), proposed for prefrontal cortex, has one subpopulation per future time step, receives the current location in its present subspace and reward inputs in its future subspaces, and embeds copies of the environment's adjacency matrix between consecutive subspaces; its fixed point is a trajectory over states and future times that resembles an approximate posterior of a planning-as-inference problem. These relaxed states support planning as an arrival time, a trajectory, or a predictive field. The present model differs in the status of its relaxed state: a stationary field over one population, reused across all goals and horizons, that is identical to the optimal control field of a stated problem (Theorem 1), with the optimal policy in the branch synaptic currents, with the operator written by BTSP during goal-free exploration, and with local edits rerouting all plans at low rank. The two kinds of query, source at the current state and clamp at the goal, are two uses of the same operator (Sec. 2.3); the desirability field is an exponentially weighted compression of the future time slices of the spacetime attractor, and prefrontal circuits could supply the goal boundary condition that the hippocampal network evaluates on its fine-grained graph (Ito et al., 2015; Jensen et al., 2026).

B Notation and standing assumptions

The explored environment is represented by a finite set of assemblies

$$\mathcal{X} = \{1, \dots, N\}.$$

Assembly i codes location i, with physical coordinate $\boldsymbol{x}_i$ when a coordinate is needed. The recalled goal is denoted by g. The non-goal set is

$$\mathcal{I}_g = \mathcal{X} \setminus \{g\}.$$

For a matrix $\boldsymbol{M}$, $\boldsymbol{M}_{AB}$ denotes the submatrix with rows indexed by A and columns indexed by B. Dirichlet equations are written on $\mathcal{I}_g$, with the goal assembly held fixed.

If the reversible support graph has several connected components, the theory below is applied to the closed irreducible component containing the recalled goal. States in components not containing g have no passive path to g, and therefore have zero Dirichlet response and zero LMDP desirability for that goal. To keep notation light, $\mathcal{X}$ denotes the goal-connected component throughout the appendix.

The passive transition kernel is denoted by $\boldsymbol{P}_0$. The row convention is

$$(\boldsymbol{P}_0 \boldsymbol{f})_i = \sum_j P_0(j \mid i) f_j,$$

so $P_0(j \mid i)$ is the probability of moving from i to j under the spontaneous goal-free dynamics. The recurrent matrix $\boldsymbol{S}$ acts with the same convention:

$$(\boldsymbol{S}\boldsymbol{f})_i = \sum_j S_{ij} f_j.$$

The exact symmetric theory assumes that $\boldsymbol{P}_0$ is finite, row-stochastic, nonnegative, irreducible on the goal-connected component, and reversible with respect to a positive stationary measure $\boldsymbol{\pi}$, normalized by $\sum_i \pi_i = 1$:

$$\pi_i P_0(j \mid i) = \pi_j P_0(i \mid j). \tag{10}$$

Let

$$\boldsymbol{\Pi} = \operatorname{diag}(\pi_1, \dots, \pi_N), \qquad \boldsymbol{S} = \boldsymbol{\Pi}^{1/2} \boldsymbol{P}_0 \boldsymbol{\Pi}^{-1/2}.$$

Then

$$S_{ij} = \sqrt{\frac{\pi_i}{\pi_j}}\, P_0(j \mid i), \tag{11}$$

and $\boldsymbol{S}$ is symmetric. Since $\boldsymbol{S}$ is similar to $\boldsymbol{P}_0$, the two matrices have the same eigenvalues. On the goal-connected irreducible component, the Perron vector of $\boldsymbol{S}$ is

$$\boldsymbol{\varphi}_0 = \sqrt{\boldsymbol{\pi}}, \qquad (\sqrt{\boldsymbol{\pi}})_i = \sqrt{\pi_i},$$

with

$$\boldsymbol{S}\boldsymbol{\varphi}_0 = \boldsymbol{\varphi}_0, \qquad \|\boldsymbol{\varphi}_0\|_2 = 1.$$

The graph Laplacian in the symmetric network coordinate is

$$\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{S}.$$

On a connected component,

$$\ker \boldsymbol{L} = \operatorname{span}\{\sqrt{\boldsymbol{\pi}}\}.$$

If the full graph has several closed components, the nullspace is the span of the corresponding component-wise $\sqrt{\boldsymbol{\pi}}$ vectors.

The Dirichlet Laplacian for a clamped goal is

$$\boldsymbol{L}_g^{\mathrm{Dir}} = \boldsymbol{I}_{\mathcal{I}_g} - \boldsymbol{S}_{\mathcal{I}_g\mathcal{I}_g}.$$

Because the killed passive chain on $\mathcal{I}_g$ is transient when every state in the component can hit g,

$$r_{\mathrm{sp}}(\boldsymbol{P}_{0,\mathcal{I}_g\mathcal{I}_g}) < 1.$$

Since $\boldsymbol{S}_{\mathcal{I}_g\mathcal{I}_g}$ is similar to $\boldsymbol{P}_{0,\mathcal{I}_g\mathcal{I}_g}$ and is symmetric, all eigenvalues of $\boldsymbol{S}_{\mathcal{I}_g\mathcal{I}_g}$ are strictly smaller than one in absolute value, and

$$\boldsymbol{L}_g^{\mathrm{Dir}} \succ 0.$$

The recurrent discount is

$$\alpha \in (0,1), \qquad \kappa = -\log \alpha.$$

The LMDP discount is

$$\gamma = \exp(-q_0/\lambda).$$

The exact network–LMDP equivalence uses

$$\alpha = \gamma.$$

The corresponding screening mass is

$$m = \alpha^{-1} - 1 = \exp(q_0/\lambda) - 1.$$

All fields denoted by $\boldsymbol{u}$ in the linear, nonlinear, and readout analyses are background-subtracted goal-induced perturbations above the goal-free fixed point. Raw firing rates have the form

$$\boldsymbol{r} = \boldsymbol{r}^0 + \boldsymbol{u}, \qquad r_i = r_i^0 + u_i.$$

The LMDP-equivalent field is $\boldsymbol{u}$, not the raw rate $\boldsymbol{r}$. The normalized goal-locked activity field is

$$\boldsymbol{\psi}^g = (\psi_i^g)_{i\in\mathcal{X}}, \qquad \psi_i^g = \frac{u_i}{\bar{u}_g},$$

where $\bar{u}_g$ is the clamped goal perturbation. The LMDP desirability field is

$$\boldsymbol{z}^g = (z_i^g)_{i\in\mathcal{X}}, \qquad z_i^g = \exp[-V_i^g/\lambda], \qquad z_g^g = 1.$$

Under uniform $\boldsymbol{\pi}$,

$$\psi_i^g = z_i^g.$$

For general reversible $\boldsymbol{\pi}$,

$$\psi_i^g = \sqrt{\frac{\pi_i}{\pi_g}}\, z_i^g.$$

The graph distance $d(i,j)$ is the minimum number of accessible transitions in a path from i to j. For an accessible oriented edge $e = (i,j)$, the edge gradient is

$$\nabla_e f = f_j - f_i.$$

When a continuous spatial gradient $\nabla f(\boldsymbol{x}_i)$ is written on an embedded graph, it denotes the local least-squares or interpolation gradient obtained from neighboring assemblies and their coordinates.

C Linearization of the recurrent population

Let r_i denote the raw firing rate in the recurrent rate equation

$$\tau_u \dot{r}_i = -r_i + g_s \sum_j S_{ij}\phi(r_j) - g_{\mathrm{inh}}\phi(r_i) + I_i(t), \tag{12}$$

where

$$I_i(t) = I_i^{\mathrm{bg}} + I_i^{\mathrm{EC}}(\boldsymbol{y}_t) + I_i^{\mathrm{goal}}(g,t).$$

The local inhibitory term reduces the local gain and the effective time constant in the first-order dynamics.

C.1 Expansion around the goal-free fixed point

During goal-free behavior,

$$I_i^{\mathrm{goal}}(g,t) = 0.$$

For fixed current position $\boldsymbol{y}_t$, the background state $\boldsymbol{r}^0$ satisfies

$$0 = -r_i^0 + g_s \sum_j S_{ij}\phi(r_j^0) - g_{\mathrm{inh}}\phi(r_i^0) + I_i^0, \tag{13}$$

where

$$I_i^0 = I_i^{\mathrm{bg}} + I_i^{\mathrm{EC}}(\boldsymbol{y}_t).$$

The recall equations below are stated for fixed $\boldsymbol{y}_t$, or for a quasi-static planning pause in which the change in $\boldsymbol{r}^0(\boldsymbol{y}_t)$ is negligible over the relaxation time. If $\boldsymbol{y}_t$ varies during the relaxation, the perturbation equation includes the additional forcing term $-\tau_u \dot{r}_i^0(\boldsymbol{y}_t)$. Define the goal-induced perturbation and raw input perturbation by

$$u_i = r_i - r_i^0, \qquad \delta I_i^{\mathrm{raw}} = I_i(t) - I_i^0.$$

Let

$$\rho_i = \phi'(r_i^0).$$

The Taylor estimates assume $\phi \in C^2$ on the relevant bounded activity interval. For a piecewise-smooth nonlinearity, the same estimates apply on intervals that do not cross a kink or hard saturation boundary. Taylor expansion gives

$$\phi(r_i^0 + u_i) = \phi(r_i^0) + \rho_i u_i + \frac{1}{2}\phi''(r_i^0)u_i^2 + \mathcal{O}(|u_i|^3).$$

Substituting into Eq. (12) and subtracting Eq. (13) yields

$$\tau_u \dot{u}_i = -u_i + g_s \sum_j S_{ij}\rho_j u_j - g_{\mathrm{inh}}\rho_i u_i + \delta I_i^{\mathrm{raw}} + \mathcal{R}_i(\boldsymbol{u}), \tag{14}$$

where

$$\begin{aligned}\mathcal{R}_i(\boldsymbol{u}) = {} & g_s \sum_j S_{ij}\Big[\phi(r_j^0 + u_j) - \phi(r_j^0) - \rho_j u_j\Big] \\ & - g_{\mathrm{inh}}\Big[\phi(r_i^0 + u_i) - \phi(r_i^0) - \rho_i u_i\Big].\end{aligned} \tag{15}$$

On any bounded activity interval there is a finite constant C_ϕ, depending on g_s, g_{inh}, $\boldsymbol{S}$, and the supremum of $|\phi''|$ on that interval, such that

$$|\mathcal{R}_i(\boldsymbol{u})| \le C_\phi \left(|u_i|^2 + \sum_j |S_{ij}|\,|u_j|^2 \right). \tag{16}$$

C.2 Homogeneous-gain scalar-discount reduction

The exact scalar-discount theory is obtained when the background gain is homogeneous:

$$\rho_i = \rho > 0 \qquad \text{for all } i.$$

Then Eq. (14) becomes

$$\tau_u \dot{\boldsymbol{u}} = -(1 + g_{\mathrm{inh}}\rho)\boldsymbol{u} + g_s \rho\, \boldsymbol{S}\boldsymbol{u} + \delta \boldsymbol{I}^{\mathrm{raw}} + \boldsymbol{\mathcal{R}}(\boldsymbol{u}).$$

Define

$$c_\ell = 1 + g_{\mathrm{inh}}\rho, \qquad \tau = \frac{\tau_u}{c_\ell}, \qquad \alpha = \frac{g_s \rho}{c_\ell}, \tag{17}$$

and

$$\delta \widetilde{\boldsymbol{I}} = \frac{\delta \boldsymbol{I}^{\mathrm{raw}}}{c_\ell}, \qquad \widetilde{\boldsymbol{\mathcal{R}}} = \frac{\boldsymbol{\mathcal{R}}}{c_\ell}. \tag{18}$$

The homogeneous-gain dynamics are

$$\tau \dot{\boldsymbol{u}} = -(\boldsymbol{I} - \alpha \boldsymbol{S})\boldsymbol{u} + \delta \widetilde{\boldsymbol{I}} + \widetilde{\boldsymbol{\mathcal{R}}}(\boldsymbol{u}). \tag{19}$$

Dropping the nonlinear remainder gives

$$\tau \dot{\boldsymbol{u}} = -(\boldsymbol{I} - \alpha \boldsymbol{S})\boldsymbol{u} + \delta \widetilde{\boldsymbol{I}}. \tag{20}$$

The condition

$$0 < \alpha < 1$$

is the gain constraint

$$0 < g_s \rho < 1 + g_{\mathrm{inh}}\rho. \tag{21}$$

Since $\boldsymbol{S}$ is symmetric and similar to a Markov kernel, its spectrum lies in $[-1, 1]$. Along an eigenvector of $\boldsymbol{S}$ with eigenvalue $\lambda_n(\boldsymbol{S})$, the homogeneous linearized dynamics have decay rate

$$\frac{1 - \alpha \lambda_n(\boldsymbol{S})}{\tau}.$$

For $\alpha \in (0, 1)$,

$$1 - \alpha \lambda_n(\boldsymbol{S}) \geq 1 - \alpha > 0.$$

Thus the linearized fixed point is stable and $\boldsymbol{I} - \alpha \boldsymbol{S}$ is invertible. Similarly,

$$\|\boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g}\|_2 \leq 1,$$

so $\boldsymbol{I} - \alpha \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g}$ is invertible for every $\alpha \in (0, 1)$.

In the scaled variables, the nonlinear remainder satisfies

$$|\widetilde{\mathcal{R}}_i(\boldsymbol{u})| \leq \widetilde{C}_\phi \left(|u_i|^2 + \sum_j |S_{ij}|\, |u_j|^2 \right) \tag{22}$$

for a finite constant $\widetilde{C}_\phi$.

C.3 Diagonal-gain operator

A localized entorhinal drive can make the background r_i^0 spatially nonuniform. For a saturating nonlinearity, this gives nonuniform gains $\rho_i = \phi'(r_i^0)$. Define

$$\boldsymbol{D}_\rho = \mathrm{diag}(\rho_1, \ldots, \rho_N), \qquad \boldsymbol{D}_\ell = \boldsymbol{I} + g_{\mathrm{inh}} \boldsymbol{D}_\rho.$$

The first-order part of Eq. (14) is

$$\tau_u \dot{\boldsymbol{u}} = -\big(\boldsymbol{D}_\ell - g_s \boldsymbol{S}\boldsymbol{D}_\rho\big)\boldsymbol{u} + \delta \boldsymbol{I}^{\mathrm{raw}}. \tag{23}$$

At steady state, or after using local diagonal time constants, the effective recurrent operator is

$$\boldsymbol{K} = g_s \boldsymbol{D}_\ell^{-1} \boldsymbol{S} \boldsymbol{D}_\rho. \tag{24}$$

The scalar-discount theorem uses the homogeneous case

$$\boldsymbol{K} = \alpha \boldsymbol{S}.$$

For nonuniform gain, the scalar-discount equivalence is treated as a perturbation of this case, or as a state-dependent gain operator.

Let

$$\boldsymbol{E}_K = \boldsymbol{K} - \alpha \boldsymbol{S}.$$

If

$$\|\boldsymbol{E}_K\|_2 < 1 - \alpha,$$

then $\boldsymbol{I} - \boldsymbol{K}$ is invertible and

$$\|(\boldsymbol{I} - \boldsymbol{K})^{-1} - (\boldsymbol{I} - \alpha \boldsymbol{S})^{-1}\|_2 \leq \frac{\|\boldsymbol{E}_K\|_2}{(1-\alpha)(1-\alpha-\|\boldsymbol{E}_K\|_2)}. \tag{25}$$

When $\boldsymbol{S}$ is symmetric and $0 < \rho_{\min} \leq \rho_i \leq \rho_{\max} < \infty$, $\boldsymbol{K}$ is diagonally symmetrizable. Indeed,

$$K_{ij} = g_s \frac{\rho_j}{1 + g_{\text{inh}} \rho_i} S_{ij}.$$

With

$$\omega_i = \rho_i (1 + g_{\text{inh}} \rho_i),$$

one has

$$\omega_i K_{ij} = g_s \rho_i \rho_j S_{ij} = \omega_j K_{ji}.$$

Thus $\boldsymbol{K}$ is self-adjoint in the weighted inner product induced by $\boldsymbol{\omega}$. It is not generally a stochastic reversible transition kernel, because its row sums and effective discount can vary across states.

D LMDP desirability and exact linear network equivalence

D.1 Bellman equation with KL control cost

At each non-goal state $i \in \mathcal{I}_g$, the agent chooses a controlled next-state distribution

$$\mu(\cdot \mid i)$$

supported on the accessible successors of i. The per-step cost is

$$q_0 + \lambda D_{\text{KL}}(\mu(\cdot \mid i) \,\|\, P_0(\cdot \mid i)),$$

where $q_0 > 0$ and $\lambda > 0$. The terminal value is

$$V(g) = 0.$$

For $i \in \mathcal{I}_g$, the Bellman equation is

$$V(i) = q_0 + \min_{\mu(\cdot|i)} \left\{ \lambda \sum_j \mu(j \mid i) \log \frac{\mu(j \mid i)}{P_0(j \mid i)} + \sum_j \mu(j \mid i) V(j) \right\}. \tag{26}$$

The Lagrange multiplier calculation for the constraint $\sum_j \mu(j \mid i) = 1$ gives

$$\mu^*(j \mid i) = \frac{P_0(j \mid i) \exp[-V(j)/\lambda]}{\sum_k P_0(k \mid i) \exp[-V(k)/\lambda]}. \tag{27}$$

The minimized bracket in Eq. (26) is

$$-\lambda \log \sum_j P_0(j \mid i) \exp[-V(j)/\lambda].$$

Thus

$$V(i) = q_0 - \lambda \log \sum_j P_0(j \mid i) \exp[-V(j)/\lambda]. \tag{28}$$

D.2 Linear desirability equation

Define

$$z(i) = \exp[-V(i)/\lambda], \qquad \gamma = \exp(-q_0/\lambda).$$

Exponentiating Eq. (28) gives

$$z(i) = \gamma \sum_j P_0(j \mid i) z(j), \qquad i \in \mathcal{I}_g, \tag{29}$$

with boundary condition

$$z(g) = 1.$$

Splitting the sum into non-goal and goal terms yields

$$(\boldsymbol{I} - \gamma \boldsymbol{P}_{0,\mathcal{I}_g\mathcal{I}_g}) \boldsymbol{z}_{\mathcal{I}_g} = \gamma \boldsymbol{p}_g, \qquad (\boldsymbol{p}_g)_i = P_0(g \mid i). \tag{30}$$

The optimal policy in desirability coordinates is

$$\mu^*(j \mid i) = \frac{P_0(j \mid i) z(j)}{\sum_k P_0(k \mid i) z(k)} = \frac{\gamma P_0(j \mid i) z(j)}{z(i)}. \tag{31}$$

On the finite goal-connected component, $z(i) > 0$ for all i, and $\mu^*(\cdot \mid i)$ has the same support as $P_0(\cdot \mid i)$. Outside the goal-reachable component, $z = 0$ and the finite-value policy formula is not used.

Since $\gamma \in (0, 1)$,

$$(\boldsymbol{I} - \gamma \boldsymbol{P}_{0,\mathcal{I}_g\mathcal{I}_g})^{-1} = \sum_{t \geq 0} \gamma^t \boldsymbol{P}^t_{0,\mathcal{I}_g\mathcal{I}_g}.$$

Therefore

$$\boldsymbol{z}_{\mathcal{I}_g} = \sum_{L \geq 1} \gamma^L \boldsymbol{P}^{L-1}_{0,\mathcal{I}_g\mathcal{I}_g} \boldsymbol{p}_g. \tag{32}$$

If τ_g is the first hitting time of g under the passive chain, then

$$z(i) = \mathbb{E}_{\boldsymbol{P}_0, i}\left[\gamma^{\tau_g} \mathbb{1}_{\{\tau_g < \infty\}}\right]. \tag{33}$$

On the goal-connected finite component, $\mathbb{P}_i(\tau_g < \infty) = 1$.

D.3 Exact linear network–LMDP theorem

Theorem 2 (Linear network–LMDP equivalence)**.** *Assume that $\boldsymbol{P}_0$ is finite, row-stochastic, non-negative, irreducible on the goal-connected component, and reversible with respect to positive $\boldsymbol{\pi}$. Let*

$$\boldsymbol{S} = \boldsymbol{\Pi}^{1/2} \boldsymbol{P}_0 \boldsymbol{\Pi}^{-1/2}.$$

Assume homogeneous linear gain, so the background-subtracted goal-induced network dynamics are governed by $\boldsymbol{I} - \alpha \boldsymbol{S}$ with

$$0 < \alpha < 1.$$

During goal recall, impose a true Dirichlet clamp

$$u_g = \bar{u}_g$$

and let the non-goal assemblies solve

$$(\boldsymbol{I} - \alpha \boldsymbol{S}_{\mathcal{I}_g\mathcal{I}_g}) \boldsymbol{u}_{\mathcal{I}_g} = \alpha \boldsymbol{S}_{\mathcal{I}_g g} \bar{u}_g. \tag{34}$$

Set

$$\alpha = \gamma = \exp(-q_0/\lambda).$$

Then the normalized network field satisfies

$$\psi^g_i = \frac{u_i}{\bar{u}_g} = \sqrt{\frac{\pi_i}{\pi_g}}\, z^g_i \tag{35}$$

for every state in the goal-connected component, where z^g is the LMDP desirability field with boundary condition $z^g_g = 1$.

Proof. Define the symmetric-coordinate desirability

$$\tilde{z}_i = \sqrt{\pi_i}\, z_i.$$

Multiplying Eq. (30) by $\mathbf{\Pi}^{1/2}$ gives

$$(\boldsymbol{I} - \gamma \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g})\tilde{\boldsymbol{z}}_{\mathcal{I}_g} = \gamma \tilde{\boldsymbol{p}}_g,$$

where

$$(\tilde{\boldsymbol{p}}_g)_i = \sqrt{\pi_i} P_0(g \mid i).$$

Using detailed balance,

$$\sqrt{\pi_i} P_0(g \mid i) = \sqrt{\pi_g}\sqrt{\frac{\pi_i}{\pi_g}} P_0(g \mid i) = \sqrt{\pi_g} S_{ig}.$$

Hence

$$(\boldsymbol{I} - \gamma \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g})\tilde{\boldsymbol{z}}_{\mathcal{I}_g} = \gamma \sqrt{\pi_g} \boldsymbol{S}_{\mathcal{I}_g g}. \tag{36}$$

Set

$$u_i = \frac{\bar{u}_g}{\sqrt{\pi_g}} \tilde{z}_i = \bar{u}_g \sqrt{\frac{\pi_i}{\pi_g}} z_i.$$

Then $u_g = \bar{u}_g z_g = \bar{u}_g$, and Eq. (36) becomes Eq. (34). Since $\boldsymbol{I} - \alpha \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g}$ is invertible, the solution is unique. □

A source-normalized full-resolvent representation is equivalent to the Dirichlet clamp. Let

$$\boldsymbol{R} = (\boldsymbol{I} - \alpha \boldsymbol{S})^{-1}.$$

Solving

$$(\boldsymbol{I} - \alpha \boldsymbol{S})\tilde{\boldsymbol{u}}^{(g)} = c_g \boldsymbol{e}_g$$

gives

$$\tilde{\boldsymbol{u}}^{(g)} = c_g \boldsymbol{R} \boldsymbol{e}_g.$$

Choosing

$$c_g = \frac{\bar{u}_g}{R_{gg}}$$

enforces $\tilde{\boldsymbol{u}}^{(g)}_g = \bar{u}_g$. For $i \neq g$, the equation has no source term, so

$$(\boldsymbol{I} - \alpha \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g})\tilde{\boldsymbol{u}}^{(g)}_{\mathcal{I}_g} = \alpha \boldsymbol{S}_{\mathcal{I}_g g} \bar{u}_g,$$

which is the same Dirichlet field.

D.4 Directed transitions: diagonal similarity

Theorem 1 uses reversibility to obtain a symmetric recurrent matrix. The identity between the clamped steady state and the desirability field does not need symmetry; it needs the recurrent matrix to be similar to the passive kernel through a positive diagonal matrix. This subsection proves Corollary 1 of the main text.

Let $\boldsymbol{P}_0$ be any row-stochastic kernel on the N states, reversible or not, and let $\gamma \in (0, 1)$. The desirability equation

$$z^g(i) = \gamma \sum_j P_0(j \mid i)\, z^g(j) \quad (i \neq g), \qquad z^g(g) = 1,$$

has a unique solution because the spectral radius of $\gamma \boldsymbol{P}_{0, \mathcal{I}_g \mathcal{I}_g}$ is at most $\gamma < 1$. Let $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_N)$ with $d_i > 0$, and define

$$\boldsymbol{S} := \boldsymbol{D} \boldsymbol{P}_0 \boldsymbol{D}^{-1}, \qquad S_{ij} = \frac{d_i}{d_j} P_0(j \mid i).$$

Set $\alpha = \gamma$ and define the candidate field $u_i := \bar{u}_g\, (d_i / d_g)\, z^g(i)$, so that $u_g = \bar{u}_g$. For $i \neq g$,

$$\alpha \sum_j S_{ij} u_j = \gamma \sum_j \frac{d_i}{d_j} P_0(j \mid i)\, \bar{u}_g \frac{d_j}{d_g} z^g(j) = \bar{u}_g \frac{d_i}{d_g}\, \gamma \sum_j P_0(j \mid i) z^g(j) = \bar{u}_g \frac{d_i}{d_g}\, z^g(i) = u_i.$$

Hence $\boldsymbol{u}$ solves the clamped steady-state equation $(\boldsymbol{I} - \alpha \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g})\boldsymbol{u}_{\mathcal{I}_g} = \alpha \boldsymbol{S}_{\mathcal{I}_g g} \bar{u}_g$ of Eq. (3). The solution is unique because $\boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g}$ is similar to $\boldsymbol{P}_{0,\mathcal{I}_g \mathcal{I}_g}$ and therefore shares its spectrum, so the spectral radius of $\alpha \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g}$ is below one. This proves $u_i/\bar{u}_g = (d_i/d_g) z^g(i)$.

For the policy, $P_0(j \mid i) = (d_j/d_i) S_{ij}$ and $z^g(j) = (d_g/\bar{u}_g)\, u_j/d_j$, so

$$\mu^*(j \mid i) \;\propto\; P_0(j \mid i)\, z^g(j) = \frac{d_g}{d_i\, \bar{u}_g}\, S_{ij}\, u_j \;\propto\; S_{ij}\, u_j,$$

because the prefactor does not depend on j. The optimal next-step distribution is the normalized vector of branch currents $S_{ij} u_j$ arriving at unit i, exactly as in the reversible case. The choice $\boldsymbol{D} = \boldsymbol{\Pi}^{1/2}$ for a reversible kernel gives the symmetric matrix of Eq. (5) and recovers Theorem 1; the choice $\boldsymbol{D} = \boldsymbol{I}$ gives $S_{ij} = P_0(j \mid i)$ and $u_i/\bar{u}_g = z^g(i)$ with no occupancy factor. App. M.5 verifies the corollary numerically for three positive diagonals, including $\boldsymbol{D} = \boldsymbol{I}$. Because the readout depends on the field only through $\nabla \log$, a non-uniform $\boldsymbol{D}$ adds the goal-independent bias $\nabla \log d$ to the readout direction, the analogue of the $\frac{1}{2} \nabla \log \pi$ bias of App. F.4.

Synaptic orientation. In the paper's convention S_{ij} is the synapse from unit j onto unit i. The corollary stores the legal move $i \to j$ in the synapse $j \to i$, so the clamped activity propagates from the goal backward along legal routes. A causal plasticity kernel, which potentiates a synapse when its presynaptic unit fires before the postsynaptic unit, writes $j \to i$ whenever a sequence plays j and then i. Reverse replay of the executed trajectory $i \to j$ plays exactly this order and therefore writes $\boldsymbol{S}$; forward replay of the same trajectory writes the transpose $\boldsymbol{S}^\top$, the predictive operator whose resolvent column from the current state is the successor representation of Sec. 2.3. With a symmetric kernel and bidirectional sampling the two coincide, which is the reversible case of the main text.

E Reversibility, learned transition kernels, and normalization

E.1 Symmetric network coordinate

Assume detailed balance:

$$\pi_i P_0(j \mid i) = \pi_j P_0(i \mid j).$$

With

$$\boldsymbol{S} = \boldsymbol{\Pi}^{1/2} \boldsymbol{P}_0 \boldsymbol{\Pi}^{-1/2},$$

one has

$$S_{ij} = \sqrt{\frac{\pi_i}{\pi_j}} P_0(j \mid i).$$

Then

$$S_{ji} = \sqrt{\frac{\pi_j}{\pi_i}} P_0(i \mid j) = \sqrt{\frac{\pi_i}{\pi_j}} P_0(j \mid i) = S_{ij}.$$

Thus $\boldsymbol{S}$ is symmetric. This is the recurrent coordinate used in Theorem 2.

E.2 Transition-gated symmetric accumulation

Let $a_i(t) \geq 0$ be the activity of assembly i. The transition-gated accumulation rule is

$$W_{ij}^{\mathrm{tr}} = \sum_{\tau} k(\tau) \mathbb{E}[a_i(t) a_j(t+\tau) \Xi_{ij}(t, \tau)]\,, \tag{37}$$

where

$$k(\tau) = k(-\tau), \qquad k(\tau) \geq 0.$$

The gate $\Xi_{ij}(t, \tau) \geq 0$ marks an accessible transition between assemblies i and j inside the plasticity window. Physical movement and internally generated sequences enter through the same transition gate. The nonnegativity of a_i, k, and Ξ gives

$$W_{ij}^{\mathrm{tr}} \geq 0.$$

Define the lagged gated correlation

$$C_{ij}(\tau) = \mathbb{E}[a_i(t) a_j(t+\tau) \Xi_{ij}(t, \tau)]\,.$$

The ideal reversible sampling condition is

$$C_{ij}(\tau) = C_{ji}(-\tau) \tag{38}$$

for every accessible unordered edge $\{i, j\}$ and lag τ. This condition states joint invariance of the activity pair and transition gate under reversal of the sampled edge. Under Eq. (38),

$$\begin{aligned} W^{\mathrm{tr}}_{ji} &= \sum_{\tau} k(\tau) C_{ji}(\tau) \\ &= \sum_{\tau} k(\tau) C_{ij}(-\tau) \\ &= \sum_{\tau'} k(-\tau') C_{ij}(\tau') \\ &= \sum_{\tau'} k(\tau') C_{ij}(\tau') = W^{\mathrm{tr}}_{ij}. \end{aligned} \tag{39}$$

Thus $\boldsymbol{W}^{\mathrm{tr}}$ is symmetric in expectation under stationary bidirectional sampling with a symmetric nonnegative temporal kernel and a jointly time-reversal-invariant transition gate. Directed or policy-biased sequence statistics enter as deviations from this symmetrized effective transition memory and are handled by the perturbative nonreversible terms in App. F.2.

If a transition is blocked, the gate is absent for that pair, and

$$W^{\mathrm{tr}}_{ij} = 0.$$

The support of $\boldsymbol{W}^{\mathrm{tr}}$ is therefore the accessible transition graph.

E.3 Lazy normalization

Let

$$d_i = \sum_{j \neq i} W^{\mathrm{tr}}_{ij}.$$

Choose a fixed global ϵ satisfying

$$0 < \epsilon \leq \frac{1}{\max_i d_i}. \tag{40}$$

The lazy normalization is

$$P_0(j \mid i) = \epsilon W^{\mathrm{tr}}_{ij} \quad (j \neq i), \qquad P_0(i \mid i) = 1 - \epsilon d_i. \tag{41}$$

Every row sums to one. If $\boldsymbol{W}^{\mathrm{tr}}$ is symmetric, then

$$P_0(j \mid i) = P_0(i \mid j)$$

for all i, j, so $\boldsymbol{P}_0$ is symmetric and reversible with uniform stationary measure

$$\pi_i = \frac{1}{N}.$$

In this regime,

$$\boldsymbol{S} = \boldsymbol{P}_0, \qquad \sqrt{\boldsymbol{\pi}} \propto \mathbf{1}.$$

For local environmental edits, ϵ is kept fixed across the pre-edit and post-edit graphs and chosen to satisfy the laziness condition for both graphs. If ϵ were globally reset after an edit, all off-diagonal weights could be rescaled, producing a nonlocal operator change.

E.4 Standard random-walk normalization

Another common normalization is

$$P_0(j \mid i) = \frac{W^{\mathrm{tr}}_{ij}}{d_i} \qquad (d_i > 0). \tag{42}$$

For isolated states with $d_i = 0$, one either removes the state from the navigable component or sets $P_0(i \mid i) = 1$ as a closed singleton component. On a connected component with $d_i > 0$, symmetric $\boldsymbol{W}^{\mathrm{tr}}$ gives detailed balance with

$$\pi_i = \frac{d_i}{\sum_\ell d_\ell}.$$

The symmetric recurrent coordinate is

$$S_{ij} = \sqrt{\frac{\pi_i}{\pi_j}} P_0(j \mid i) = \frac{W_{ij}^{\mathrm{tr}}}{\sqrt{d_i d_j}}. \tag{43}$$

F Finite sampling, nonuniform $\boldsymbol{\pi}$, and nonreversible perturbations

F.1 Reversible projection of empirical transition counts

Finite sampling can create asymmetric empirical transition counts. Let $\widehat{\boldsymbol{W}}$ denote the empirical nonnegative count matrix. A reversible projection is obtained by

$$\widehat{\boldsymbol{W}}^{\mathrm{sym}} = \frac{1}{2}(\widehat{\boldsymbol{W}} + \widehat{\boldsymbol{W}}^\top). \tag{44}$$

Applying the lazy normalization of Eq. (41) to $\widehat{\boldsymbol{W}}^{\mathrm{sym}}$ gives an exactly symmetric passive kernel and hence uniform $\boldsymbol{\pi}$. Applying the random-walk normalization to $\widehat{\boldsymbol{W}}^{\mathrm{sym}}$ gives an exactly reversible passive kernel with $\pi_i \propto \widehat{d}_i$. The exact theorem applies to these reversible projected kernels.

F.2 Perturbation by irreversible sampling error

Let $\bar{\boldsymbol{P}}$ be an ideal reversible kernel with stationary measure $\bar{\boldsymbol{\pi}}$, and let

$$\bar{\boldsymbol{S}} = \bar{\boldsymbol{\Pi}}^{1/2} \bar{\boldsymbol{P}} \bar{\boldsymbol{\Pi}}^{-1/2}$$

be its symmetric coordinate. Suppose a raw empirical kernel has the form

$$\boldsymbol{P}^{\mathrm{emp}} = \bar{\boldsymbol{P}} + \Delta \boldsymbol{P}, \qquad \Delta \boldsymbol{P}\, \mathbf{1} = 0.$$

In the $\bar{\boldsymbol{\pi}}$-symmetric coordinate define

$$\boldsymbol{E} = \bar{\boldsymbol{\Pi}}^{1/2} \Delta \boldsymbol{P}\, \bar{\boldsymbol{\Pi}}^{-1/2}. \tag{45}$$

Then the raw empirical recurrent operator in this coordinate is

$$\boldsymbol{S}^{\mathrm{emp}} = \bar{\boldsymbol{S}} + \boldsymbol{E}.$$

Decompose

$$\boldsymbol{E} = \boldsymbol{E}_{\mathrm{sym}} + \boldsymbol{E}_{\mathrm{asym}}, \qquad \boldsymbol{E}_{\mathrm{sym}} = \frac{1}{2}(\boldsymbol{E} + \boldsymbol{E}^\top), \qquad \boldsymbol{E}_{\mathrm{asym}} = \frac{1}{2}(\boldsymbol{E} - \boldsymbol{E}^\top).$$

The antisymmetric part $\boldsymbol{E}_{\mathrm{asym}}$ measures irreversible circulation in the empirical transition statistics.

For the full resolvent, if

$$\alpha \|\boldsymbol{E}\|_2 < 1 - \alpha,$$

then

$$\|(\boldsymbol{I} - \alpha(\bar{\boldsymbol{S}} + \boldsymbol{E}))^{-1} - (\boldsymbol{I} - \alpha \bar{\boldsymbol{S}})^{-1}\|_2 \le \frac{\alpha \|\boldsymbol{E}\|_2}{(1-\alpha)(1 - \alpha - \alpha \|\boldsymbol{E}\|_2)}. \tag{46}$$

The same bound holds for Dirichlet restrictions with $\boldsymbol{E}$ replaced by $\boldsymbol{E}_{\mathcal{I}_g \mathcal{I}_g}$. Thus finite sampling asymmetry and irreversible circulation give controlled corrections to the recurrent field when their operator norm is small relative to the stability margin $1 - \alpha$. Exact network–LMDP equivalence with the symmetric recurrent coordinate is retained for the reversible projection $\bar{\boldsymbol{P}}$.

For a clamped goal, write

$$\boldsymbol{A}_0 = \boldsymbol{I} - \alpha \bar{\boldsymbol{S}}_{\mathcal{I}_g \mathcal{I}_g}, \qquad \boldsymbol{A}_E = \boldsymbol{I} - \alpha(\bar{\boldsymbol{S}} + \boldsymbol{E})_{\mathcal{I}_g \mathcal{I}_g}.$$

If $\boldsymbol{b}_0$ is the reversible boundary source and $\boldsymbol{b}_E = \boldsymbol{b}_0 + \delta\boldsymbol{b}$ is the empirical boundary source, then

$$\boldsymbol{u}_E = \boldsymbol{A}_E^{-1}\boldsymbol{b}_E, \qquad \boldsymbol{u}_0 = \boldsymbol{A}_0^{-1}\boldsymbol{b}_0,$$

and

$$\boldsymbol{u}_E - \boldsymbol{u}_0 = (\boldsymbol{A}_E^{-1} - \boldsymbol{A}_0^{-1})\boldsymbol{b}_0 + \boldsymbol{A}_E^{-1}\delta\boldsymbol{b}. \tag{47}$$

Equation (46) controls the first term, and

$$\|\boldsymbol{A}_E^{-1}\|_2 \le \frac{1}{1-\alpha-\alpha\|\boldsymbol{E}\|_2}$$

controls the second.

F.3 Perturbation of a reversible stationary measure

When the learned kernel is kept reversible and the stationary measure changes by sampling or normalization, the change in $\boldsymbol{\pi}$ is controlled by the spectral gap. Let

$$\boldsymbol{P}_\varepsilon = \bar{\boldsymbol{P}} + \Delta\boldsymbol{P}$$

be row-stochastic and reversible up to the retained order, with stationary measure

$$\boldsymbol{\pi} = \bar{\boldsymbol{\pi}} + \delta\boldsymbol{\pi}.$$

Use the row-vector norm

$$\|\boldsymbol{r}\|_{\bar{\boldsymbol{\pi}}^{-1}} = \left(\sum_i \frac{r_i^2}{\bar{\pi}_i}\right)^{1/2}$$

and its compatible operator norm. The stationarity equations give, to first order,

$$\delta\boldsymbol{\pi}(\boldsymbol{I} - \bar{\boldsymbol{P}}) = \bar{\boldsymbol{\pi}}\Delta\boldsymbol{P}, \qquad \delta\boldsymbol{\pi}\,\mathbf{1} = 0.$$

On the zero-sum subspace, the reversible chain has inverse norm bounded by the reciprocal spectral gap

$$\mathrm{gap} = 1 - \lambda_2(\bar{\boldsymbol{P}}) > 0.$$

Thus

$$\delta\boldsymbol{\pi} = \bar{\boldsymbol{\pi}}\Delta\boldsymbol{P}\,(\boldsymbol{I} - \bar{\boldsymbol{P}})^{\#} + \mathcal{O}(|\Delta\boldsymbol{P}|^2), \tag{48}$$

and, if $\boldsymbol{P}_\varepsilon$ remains irreducible and

$$\|\Delta\boldsymbol{P}\|_{\bar{\boldsymbol{\pi}}^{-1}\to\bar{\boldsymbol{\pi}}^{-1}} < \mathrm{gap},$$

$$\|\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}\|_{\bar{\boldsymbol{\pi}}^{-1}} \le \frac{\|\bar{\boldsymbol{\pi}}\Delta\boldsymbol{P}\|_{\bar{\boldsymbol{\pi}}^{-1}}}{\mathrm{gap} - \|\Delta\boldsymbol{P}\|_{\bar{\boldsymbol{\pi}}^{-1}\to\bar{\boldsymbol{\pi}}^{-1}}}. \tag{49}$$

F.4 Readout bias from nonuniform stationary measure

For a reversible kernel with nonuniform $\boldsymbol{\pi}$, Theorem 2 gives

$$\psi_i^g = \sqrt{\frac{\pi_i}{\pi_g}}\, z_i^g.$$

Taking logarithms,

$$\log\psi_i^g = \log z_i^g + \frac{1}{2}\log\pi_i - \frac{1}{2}\log\pi_g. \tag{50}$$

For an accessible edge $e = (i, j)$,

$$\nabla_e \log\psi^g = \nabla_e \log z^g + \frac{1}{2}\nabla_e \log\pi, \tag{51}$$

where

$$\nabla_e \log\pi = \log\pi_j - \log\pi_i.$$

In a local embedded approximation, this edge relation corresponds to

$$\nabla \log\psi^g = \nabla \log z^g + \frac{1}{2}\nabla \log\pi. \tag{52}$$

Let

$$\boldsymbol{v}_0 = \boldsymbol{\Sigma}_G \nabla \log z^g, \qquad \boldsymbol{b}_\pi = \frac{1}{2}\boldsymbol{\Sigma}_G \nabla \log \pi.$$

When $\boldsymbol{v}_0 \neq 0$ and $\|\boldsymbol{b}_\pi\| < \|\boldsymbol{v}_0\|$,

$$\sin \angle(\boldsymbol{v}_0 + \boldsymbol{b}_\pi, \boldsymbol{v}_0) \leq \frac{\|\boldsymbol{b}_\pi\|}{\|\boldsymbol{v}_0 + \boldsymbol{b}_\pi\|} \leq \frac{\|\boldsymbol{b}_\pi\|}{\|\boldsymbol{v}_0\| - \|\boldsymbol{b}_\pi\|}. \tag{53}$$

An exact corrected readout in the network coordinate uses

$$\frac{\psi_i^g}{\sqrt{\pi_i}} = \frac{z_i^g}{\sqrt{\pi_g}},$$

whose log-gradient is exactly $\nabla \log z^g$.

G Local place-kernel readout

G.1 Continuous local moment expansion

Let

$$f(\boldsymbol{x}) = \psi^g(\boldsymbol{x}) > 0$$

or, under nonuniform $\boldsymbol{\pi}$, the corrected field

$$f(\boldsymbol{x}) = \frac{\psi^g(\boldsymbol{x})}{\sqrt{\pi(\boldsymbol{x})}}.$$

Fix the animal's current position $\boldsymbol{y}$, and write

$$\boldsymbol{\xi} = \boldsymbol{x} - \boldsymbol{y}.$$

The local population-vector readout is

$$\Delta \boldsymbol{x}_{\mathrm{pop}}(\boldsymbol{y}; f) = \frac{\int \boldsymbol{\xi}\, \mathcal{G}_{\boldsymbol{y}}(\boldsymbol{y} + \boldsymbol{\xi}) f(\boldsymbol{y} + \boldsymbol{\xi})\, d\boldsymbol{\xi}}{\int \mathcal{G}_{\boldsymbol{y}}(\boldsymbol{y} + \boldsymbol{\xi}) f(\boldsymbol{y} + \boldsymbol{\xi})\, d\boldsymbol{\xi}}. \tag{54}$$

Let

$$M_0(\boldsymbol{y}) = \int \mathcal{G}_{\boldsymbol{y}}(\boldsymbol{y} + \boldsymbol{\xi})\, d\boldsymbol{\xi},$$

and define expectation under the normalized place kernel:

$$\mathbb{E}_{\mathcal{G}}[h(\boldsymbol{\xi})] = \frac{1}{M_0(\boldsymbol{y})} \int h(\boldsymbol{\xi}) \mathcal{G}_{\boldsymbol{y}}(\boldsymbol{y} + \boldsymbol{\xi})\, d\boldsymbol{\xi}.$$

The kernel mean and covariance are

$$\boldsymbol{\mu}_G(\boldsymbol{y}) = \mathbb{E}_{\mathcal{G}}[\boldsymbol{\xi}],$$
$$\boldsymbol{\Sigma}_G(\boldsymbol{y}) = \mathbb{E}_{\mathcal{G}}[(\boldsymbol{\xi} - \boldsymbol{\mu}_G)(\boldsymbol{\xi} - \boldsymbol{\mu}_G)^\top].$$

Expanding

$$f(\boldsymbol{y} + \boldsymbol{\xi}) = f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^\top \boldsymbol{\xi} + \frac{1}{2}\boldsymbol{\xi}^\top \nabla^2 f(\boldsymbol{y}) \boldsymbol{\xi} + \mathcal{O}(\|\boldsymbol{\xi}\|^3)$$

gives, to first order in the local field variation,

$$\Delta \boldsymbol{x}_{\mathrm{pop}}(\boldsymbol{y}; f) = \boldsymbol{\mu}_G(\boldsymbol{y}) + \boldsymbol{\Sigma}_G(\boldsymbol{y}) \nabla \log f(\boldsymbol{y}) + \text{higher-order kernel-width terms.} \tag{55}$$

The term $\boldsymbol{\mu}_G(\boldsymbol{y})$ is the resting displacement produced by the local kernel alone:

$$\Delta \boldsymbol{x}_{\mathrm{pop}}(\boldsymbol{y}; 1) = \boldsymbol{\mu}_G(\boldsymbol{y}).$$

The goal-dependent centered readout is therefore

$$\Delta \boldsymbol{x}_{\mathrm{goal}}(\boldsymbol{y}; f) = \Delta \boldsymbol{x}_{\mathrm{pop}}(\boldsymbol{y}; f) - \Delta \boldsymbol{x}_{\mathrm{pop}}(\boldsymbol{y}; 1) = \boldsymbol{\Sigma}_G(\boldsymbol{y}) \nabla \log f(\boldsymbol{y}) + \text{higher-order terms.} \tag{56}$$

When the local kernel is centered,

$$\boldsymbol{\mu}_G(\boldsymbol{y}) = 0,$$

the raw and centered readouts agree to leading order.

G.2 Centered-kernel correction

For a centered kernel, define

$$\boldsymbol{\Sigma}_G(\boldsymbol{y}) = \frac{1}{M_0(\boldsymbol{y})} \int \boldsymbol{\xi}\boldsymbol{\xi}^\top \mathcal{G}_{\boldsymbol{y}}(\boldsymbol{y}+\boldsymbol{\xi})\, d\boldsymbol{\xi}.$$

Then

$$\Delta \boldsymbol{x}_{\mathrm{pop}}(\boldsymbol{y}; f) = \boldsymbol{\Sigma}_G(\boldsymbol{y}) \nabla \log f(\boldsymbol{y}) + \text{higher-order terms}. \tag{57}$$

Using repeated-index summation over spatial indices, define the third moment tensor

$$T_{abc}(\boldsymbol{y}) = \frac{1}{M_0(\boldsymbol{y})} \int \xi_a \xi_b \xi_c \mathcal{G}_{\boldsymbol{y}}(\boldsymbol{y}+\boldsymbol{\xi})\, d\boldsymbol{\xi}.$$

For a centered kernel,

$$\begin{aligned}[\Delta \boldsymbol{x}_{\mathrm{pop}}]_a = [\boldsymbol{\Sigma}_G \nabla \log f]_a + \frac{1}{2} T_{abc} \frac{\partial_b \partial_c f}{f} \\ - \frac{1}{2} [\boldsymbol{\Sigma}_G \nabla \log f]_a (\boldsymbol{\Sigma}_G)_{bc} \frac{\partial_b \partial_c f}{f} + \mathcal{O}(\sigma_G^4/\ell_f^3),\end{aligned} \tag{58}$$

where σ_G is the local kernel-width scale and ℓ_f is the local length scale on which f varies. If the kernel has local reflection symmetry, then $T_{abc} = 0$, and the first shape correction is fourth order in the kernel width.

G.3 Reading the raw rate

The readout of Eq. (7) is written in terms of the deviation $\psi^g \propto u = r - r^0$. The following lemma, Proposition 1 of the main text, shows that a readout cell can multiply its place kernel by the raw rate r without any subtraction of the resting state, and that the population-vector direction is exactly unchanged.

Lemma (centered kernel). Let $\mathcal{G} := \mathcal{G}_{\boldsymbol{y}_t}$ be a nonnegative kernel with mass $M_0 := \int \mathcal{G}(\boldsymbol{x})\, d\boldsymbol{x}$ and vanishing first moment about $\boldsymbol{y}_t$, $\int (\boldsymbol{x} - \boldsymbol{y}_t)\mathcal{G}(\boldsymbol{x})\, d\boldsymbol{x} = \mathbf{0}$. For any field $f \geq 0$ with $\int \mathcal{G} f > 0$ and any constant $c \geq 0$,

$$\Delta \boldsymbol{x}_{\mathrm{pop}}[c+f] := \frac{\int (\boldsymbol{x} - \boldsymbol{y}_t)\, \mathcal{G}(\boldsymbol{x})\, \big(c + f(\boldsymbol{x})\big)\, d\boldsymbol{x}}{\int \mathcal{G}(\boldsymbol{x})\, \big(c + f(\boldsymbol{x})\big)\, d\boldsymbol{x}} = \frac{\int \mathcal{G} f}{c M_0 + \int \mathcal{G} f}\, \Delta \boldsymbol{x}_{\mathrm{pop}}[f].$$

Proof. The numerator splits as $c \int (\boldsymbol{x} - \boldsymbol{y}_t)\mathcal{G} + \int (\boldsymbol{x} - \boldsymbol{y}_t)\mathcal{G} f$; the first term vanishes by centering. The denominator is $cM_0 + \int \mathcal{G} f$. Dividing numerator and denominator by $\int \mathcal{G} f$ gives the stated factor times $\Delta \boldsymbol{x}_{\mathrm{pop}}[f]$. □

The factor lies in $(0, 1]$ and depends on the goal only through the local mass $\int \mathcal{G} f$, so the direction of the population vector computed from the raw rate $r = c + u$ equals the direction computed from u, and only the magnitude is reduced by the resting contribution. Setting $f = \psi^g$ and $c = r^0/\bar{u}_g$ gives Proposition 1. Because the result holds exactly for every centered kernel, it holds for the anisotropic kernels of App. G.4 and for the discrete population sums used in the experiments, where centering is enforced by subtracting the goal-free population vector (App. M.4).

Position-dependent resting profile. If the resting rate varies with position, $r^0(\boldsymbol{x}) = c + \varrho(\boldsymbol{x})$, the numerator acquires the additional term $\int (\boldsymbol{x} - \boldsymbol{y}_t)\mathcal{G}\varrho$, the population vector of the resting profile. This term does not depend on the goal. At rest the entorhinal bump that sets ϱ is centered on $\boldsymbol{y}_t$, so for a kernel with reflection symmetry about $\boldsymbol{y}_t$ the term vanishes; in general it is a fixed offset measured once in the absence of recall, and the centered readout of App. M.4 subtracts it. In every case the goal-dependent part of the population vector is $\int (\boldsymbol{x} - \boldsymbol{y}_t)\mathcal{G} u$, the same quantity that enters Eq. (7).

G.4 Accessible directions

The covariance $\mathbf{\Sigma}_G(\boldsymbol{y})$ records the directions sampled by the local place kernel. In open space it is close to isotropic:

$$\mathbf{\Sigma}_G(\boldsymbol{y}) \approx \sigma_G^2 \boldsymbol{I}.$$

Near a wall, variance in the blocked direction is suppressed. Inside a narrow corridor, the covariance is elongated along the corridor axis. The goal-dependent vector

$$\mathbf{\Sigma}_G(\boldsymbol{y}) \nabla \log f(\boldsymbol{y})$$

therefore projects the local log-gradient onto directions represented in the local place-kernel support. A nonzero kernel mean $\boldsymbol{\mu}_G$ contributes a resting drift term, and the centered readout of Eq. (56) isolates the goal-dependent component.

H Relation between the readout and the LMDP optimal policy

At state i, the LMDP optimal next-step distribution is

$$\mu^*(j \mid i) = \frac{P_0(j \mid i) z(j)}{\sum_k P_0(k \mid i) z(k)}.$$

Let

$$\boldsymbol{\xi}_j = \boldsymbol{x}_j - \boldsymbol{x}_i.$$

The optimal-policy mean displacement is

$$\boldsymbol{m}^*(i, g) = \sum_j \boldsymbol{\xi}_j \mu^*(j \mid i). \tag{59}$$

In the uniform-$\boldsymbol{\pi}$ case, $z = \psi^g$, and

$$\boldsymbol{m}^*(i, g) = \frac{\sum_j \boldsymbol{\xi}_j P_0(j \mid i) \psi^g(j)}{\sum_j P_0(j \mid i) \psi^g(j)}. \tag{60}$$

Define the passive local mean and covariance

$$\bar{\boldsymbol{\xi}}_i = \sum_j P_0(j \mid i) \boldsymbol{\xi}_j,$$

$$\mathbf{\Sigma}_{\boldsymbol{P}_0}(i) = \sum_j P_0(j \mid i) (\boldsymbol{\xi}_j - \bar{\boldsymbol{\xi}}_i)(\boldsymbol{\xi}_j - \bar{\boldsymbol{\xi}}_i)^\top.$$

A local expansion gives

$$\boldsymbol{m}^*(i, g) = \bar{\boldsymbol{\xi}}_i + \mathbf{\Sigma}_{\boldsymbol{P}_0}(i) \nabla \log \psi^g(\boldsymbol{x}_i) + \text{higher-order local terms}. \tag{61}$$

Thus the centered optimal displacement is

$$\boldsymbol{m}^*(i, g) - \bar{\boldsymbol{\xi}}_i = \mathbf{\Sigma}_{\boldsymbol{P}_0}(i) \nabla \log \psi^g(\boldsymbol{x}_i) + \text{higher-order local terms}. \tag{62}$$

If the passive local motion has zero mean, $\bar{\boldsymbol{\xi}}_i = 0$, then Eq. (62) reduces to the uncentered expression. Comparing Eq. (62) with Eq. (56), the place-kernel readout matches the local LMDP policy mean to leading order when $\boldsymbol{\mu}_G(\boldsymbol{y}) \approx \bar{\boldsymbol{\xi}}_i$ and $\mathbf{\Sigma}_G(\boldsymbol{y}) \approx \mathbf{\Sigma}_{\boldsymbol{P}_0}(i)$. Otherwise it is a geometrically preconditioned log-desirability readout, with preconditioner $\mathbf{\Sigma}_G(\boldsymbol{y})$ rather than $\mathbf{\Sigma}_{\boldsymbol{P}_0}(i)$.

For general reversible $\boldsymbol{\pi}$,

$$z(j) = \sqrt{\frac{\pi_g}{\pi_j}} \psi^g(j).$$

The exact LMDP policy can be written as

$$\mu^*(j \mid i) \propto P_0(j \mid i) \frac{\psi^g(j)}{\sqrt{\pi_j}}. \tag{63}$$

Equivalently,

$$\mu^*(j \mid i) \propto S_{ij} \psi^g(j),$$

because the factor $1/\sqrt{\pi_i}$ is constant for fixed i. Hence the exact policy readout under nonuniform $\boldsymbol{\pi}$ uses the corrected field

$$\boldsymbol{\psi}^g / \sqrt{\boldsymbol{\pi}}.$$

I Nonlinear perturbation under strong goal recall

The exact equivalence in Theorem 2 holds for the homogeneous linearized dynamics. Strong goal recall can drive the clamped goal assembly and nearby assemblies into the nonlinear range of ϕ. The log-gradient readout is invariant under spatially constant multiplicative factors, so the relevant nonlinear effect is the residual spatial shape change after absorbing the leading amplitude renormalization. The bounds below control this distortion in the perturbative regime measured by A_{eff}, including large raw clamps whose saturating output keeps A_{eff} within the small-gain range.

I.1 Nonlinear steady-state equation

Work in the homogeneous-gain scaled variables of App. C.2. Let

$$A = \bar{u}_g$$

be the clamped goal perturbation. The effective recurrent output emitted by the clamped goal, expressed in linearized activity units, is

$$A_{\mathrm{eff}}(A) = \frac{\phi(r_g^0 + A) - \phi(r_g^0)}{\phi'(r_g^0)}. \tag{64}$$

This definition assumes

$$\phi'(r_g^0) > 0.$$

For small A,

$$A_{\mathrm{eff}}(A) = A + \mathcal{O}(A^2).$$

For bounded saturating ϕ and fixed $\phi'(r_g^0) > 0$, $A_{\mathrm{eff}}(A)$ is bounded as $A \to \infty$.

On $\mathcal{I}_g$, the nonlinear steady-state equation is

$$(\boldsymbol{I} - \alpha \boldsymbol{S}_{\mathcal{I}_g\mathcal{I}_g})\boldsymbol{u}_{\mathcal{I}_g} = \alpha \boldsymbol{S}_{\mathcal{I}_g g} A_{\mathrm{eff}}(A) + \boldsymbol{\mathcal{R}}(\boldsymbol{u}_{\mathcal{I}_g}; A), \tag{65}$$

where $\boldsymbol{\mathcal{R}}$ contains the quadratic and higher-order nonlinear terms away from the clamped goal.

Define

$$\boldsymbol{G}_g(\alpha) = (\boldsymbol{I} - \alpha \boldsymbol{S}_{\mathcal{I}_g\mathcal{I}_g})^{-1},$$

and define the unit-boundary linear field

$$\boldsymbol{H}^g_{\mathcal{I}_g} = \boldsymbol{G}_g(\alpha)\, \alpha \boldsymbol{S}_{\mathcal{I}_g g}. \tag{66}$$

By boundary convention,

$$H^g_g = 1, \qquad \boldsymbol{H}^g|_{\mathcal{I}_g} = \boldsymbol{H}^g_{\mathcal{I}_g}.$$

The full nonlinear field can be written as

$$\boldsymbol{u}_{\mathcal{I}_g} = A_{\mathrm{eff}} \boldsymbol{H}^g_{\mathcal{I}_g} + \boldsymbol{\eta}_{\mathcal{I}_g}, \qquad \boldsymbol{\eta}_{\mathcal{I}_g} = \boldsymbol{G}_g(\alpha)\boldsymbol{\mathcal{R}}(\boldsymbol{u}_{\mathcal{I}_g}; A). \tag{67}$$

I.2 Linear exponential envelope

For the reversible passive chain,

$$H^g_i = \sqrt{\frac{\pi_i}{\pi_g}}\, \mathbb{E}_{\boldsymbol{P}_0, i}\left[\alpha^{\tau_g} \mathbb{1}_{\{\tau_g < \infty\}}\right].$$

Since $\tau_g \geq d(i, g)$,

$$|H^g_i| \leq C_h e^{-\kappa d(i,g)}, \qquad C_h = \sup_i \sqrt{\frac{\pi_i}{\pi_g}}. \tag{68}$$

In the uniform-$\boldsymbol{\pi}$ case one may take $C_h = 1$. This is a path-length envelope. The sharp asymptotic decay of a screened Green function can depend on graph geometry and screening mass.

The Dirichlet Green kernel has the corresponding envelope

$$\begin{aligned} 0 \leq [\boldsymbol{G}_g(\alpha)]_{ij} &= \sqrt{\frac{\pi_i}{\pi_j}} \sum_{t \geq 0} \alpha^t [\boldsymbol{P}^t_{0, \mathcal{I}_g\mathcal{I}_g}]_{ij} \\ &\leq C_\pi \frac{\alpha^{d(i,j)}}{1-\alpha} = C_\pi \frac{e^{-\kappa d(i,j)}}{1-\alpha}, \end{aligned} \tag{69}$$

where

$$C_\pi = \sup_{i,j} \sqrt{\frac{\pi_i}{\pi_j}}.$$

I.3 Small-gain control of the nonlinear solution

Define the weighted norm

$$\|\boldsymbol{v}\|_H = \sup_{i\in\mathcal{I}_g} \frac{|v_i|}{H_i^g}.$$

Because the component is finite and $H_i^g > 0$, this norm is finite. Let

$$\boldsymbol{Q} = (Q_i)_{i\in\mathcal{I}_g}, \qquad Q_i = (H_i^g)^2 + \sum_j |S_{ij}|(H_j^g)^2,$$

and define

$$M_Q = \sup_{i\in\mathcal{I}_g} \frac{[\boldsymbol{G}_g(\alpha)\boldsymbol{Q}]_i}{H_i^g}.$$

On a finite component, $M_Q < \infty$. For graph families, a uniform version of this estimate requires a uniform bound on neighboring ratios H_j^g/H_i^g on the support of $\boldsymbol{S}$, or equivalently a uniform Harnack-type constant for the screened field. From Eq. (22),

$$\|\boldsymbol{G}_g(\alpha)\boldsymbol{\mathcal{R}}(\boldsymbol{v}; A)\|_H \le C_Q\|\boldsymbol{v}\|_H^2, \qquad C_Q = \widetilde{C}_\phi M_Q,$$

as long as $\boldsymbol{v}$ remains inside the activity interval on which the Taylor remainder bound holds.

If

$$4C_Q|A_{\text{eff}}| < 1$$

and the corresponding local Lipschitz constant of $\boldsymbol{G}_g\boldsymbol{\mathcal{R}}$ on the ball $\|\boldsymbol{v}\|_H \le 2|A_{\text{eff}}|$ is smaller than one, the nonlinear fixed-point map

$$\mathcal{T}(\boldsymbol{v}) = A_{\text{eff}}\boldsymbol{H}^g + \boldsymbol{G}_g(\alpha)\boldsymbol{\mathcal{R}}(\boldsymbol{v}; A)$$

is a contraction on that ball. The nonlinear solution exists uniquely in the ball and satisfies

$$\|\boldsymbol{u}\|_H \le 2|A_{\text{eff}}|, \qquad \|\boldsymbol{\eta}\|_H \le 4C_Q|A_{\text{eff}}|^2. \tag{70}$$

Consequently,

$$|\eta_i| \le C_\eta|A_{\text{eff}}|^2 H_i^g \le C_\eta C_h|A_{\text{eff}}|^2 e^{-\kappa d(i,g)}. \tag{71}$$

The absolute correction has the same universal path-length envelope as the linear field. The ratio $\eta_i/(A_{\text{eff}}H_i^g)$ is controlled by $\mathcal{O}(|A_{\text{eff}}|)$ under the small-gain condition and is generally not forced to decay by Eq. (71) alone.

I.4 Far-field amplitude projection

The following geometric condition captures the far-field structure needed to separate a scalar amplitude shift from a residual shape correction.

Far-field ratio condition. There exist constants $\beta > 0$, $C_{\text{far}} < \infty$, and linear functionals $\mathcal{A}_g$ such that, for every source $\boldsymbol{v}$ with finite weighted source norm

$$\|\boldsymbol{v}\|_{2,g} = \sup_j \frac{|v_j|}{(H_j^g)^2} < \infty,$$

one has

$$\boldsymbol{G}_g(\alpha)\boldsymbol{v} = \mathcal{A}_g[\boldsymbol{v}]\boldsymbol{H}^g + \boldsymbol{\varepsilon}_g[\boldsymbol{v}], \tag{72}$$

with

$$|(\boldsymbol{\varepsilon}_g[\boldsymbol{v}])_i| \le C_{\text{far}}\|\boldsymbol{v}\|_{2,g} H_i^g e^{-\beta d(i,g)}. \tag{73}$$

This condition holds for standard single-channel far-field geometries such as homogeneous screened one-dimensional chains and regular corridors, and for graph regions where localized source responses have a unique outgoing positive harmonic profile. On arbitrary finite graphs it is an explicit geometric assumption. The residual log-gradient and shell-radius estimates below are conditional on this far-field ratio condition.

Applying Eq. (72) to

$$\boldsymbol{v} = \boldsymbol{\mathcal{R}}(\boldsymbol{u}; A)$$

gives

$$\boldsymbol{\eta} = \mathcal{A}_g[\boldsymbol{\mathcal{R}}]\boldsymbol{H}^g + \boldsymbol{\zeta}, \qquad |\zeta_i| \le C_\zeta |A_{\text{eff}}|^2 H_i^g e^{-\beta d(i,g)}. \tag{74}$$

Define the renormalized amplitude

$$A_{\text{ren}} = A_{\text{eff}} + \mathcal{A}_g[\boldsymbol{\mathcal{R}}]. \tag{75}$$

Then

$$u_i = A_{\text{ren}} H_i^g + \zeta_i. \tag{76}$$

For sufficiently small $|A_{\text{eff}}|$,

$$A_{\text{ren}} = A_{\text{eff}} \left(1 + \mathcal{O}(|A_{\text{eff}}|)\right).$$

On any region where

$$A_{\text{ren}} > 0, \qquad \left| \frac{\zeta_i}{A_{\text{ren}} H_i^g} \right| < 1,$$

the log-gradient is well defined. Write

$$u_i = A_{\text{ren}} H_i^g (1 + \varrho_i), \qquad \varrho_i = \frac{\zeta_i}{A_{\text{ren}} H_i^g}.$$

Then

$$|\varrho_i| \le C_\varrho |A_{\text{eff}}| e^{-\beta d(i,g)}. \tag{77}$$

For an accessible edge $e = (i, j)$,

$$\nabla_e \log u = \nabla_e \log H^g + \nabla_e \log(1 + \varrho). \tag{78}$$

If $\|\varrho\|_\infty \le \varepsilon < 1$, then

$$|\nabla_e \log(1 + \varrho)| \le \frac{|\varrho_j - \varrho_i|}{1 - \varepsilon}. \tag{79}$$

Using Eq. (77), edges whose endpoints are at distance at least r from the goal satisfy

$$|\nabla_e \log u - \nabla_e \log H^g| \le C_\nabla |A_{\text{eff}}| e^{-\beta r}. \tag{80}$$

For a target residual relative shape tolerance ε, the affected shell radius is

$$r_\varepsilon = \left[\frac{1}{\beta} \log \left(\frac{C_\varrho |A_{\text{eff}}(A)|}{\varepsilon} \right) \right]_+, \qquad [x]_+ = \max\{x, 0\}. \tag{81}$$

In the linear recall range $A_{\text{eff}}(A) \approx A$, multiplying recall strength by c shifts the radius by

$$\Delta r_\varepsilon = \frac{\log c}{\beta}.$$

The residual decay length is $\ell_\beta = 1/\beta$. An e-fold amplitude change shifts the shell by one ℓ_β; a tenfold change shifts it by $(\log 10)\ell_\beta$. For bounded saturating ϕ, $A_{\text{eff}}(A)$ is bounded, and the shell radius in Eq. (81) has a finite upper bound.

J Weak-cost limit and Green functions

The recurrent Laplacian $\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{S}$ (App. B) coincides with the discrete Laplacian of (Zuo et al., 2026) up to a positive scalar. Under the lazy normalization of Sec. 3.1 with rate ϵ and symmetric transition counts $\boldsymbol{W}^{\text{tr}}$, one has $\boldsymbol{S} = \boldsymbol{P}_0$ and uniform $\boldsymbol{\pi}$, so

$$\boldsymbol{L} = \epsilon \left(\boldsymbol{D}_{\text{deg}} - \boldsymbol{W}^{\text{tr}} \right), \qquad \boldsymbol{D}_{\text{deg}} = \text{diag}(d_1, \ldots, d_N), \qquad d_i = \sum_{j \ne i} W_{ij}^{\text{tr}}. \tag{82}$$

The right-hand side is the discrete Laplacian $\boldsymbol{L}_{\text{pos}} = \boldsymbol{D}_{\text{deg}} - \boldsymbol{A}_{\text{adj}}$ of (Zuo et al., 2026) with adjacency matrix $\boldsymbol{A}_{\text{adj}} = \boldsymbol{W}^{\text{tr}}$; eigenvectors of $\boldsymbol{L}$ and $\boldsymbol{S}$ therefore coincide with the Laplacian eigenfunctions of (Zuo et al., 2026).

J.1 Screened Dirichlet operator

The goal-clamped linear equation is

$$(\boldsymbol{I} - \alpha \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g}) \boldsymbol{u}_{\mathcal{I}_g} = \alpha \boldsymbol{S}_{\mathcal{I}_g g} \bar{u}_g.$$

With

$$\boldsymbol{L}_g^{\mathrm{Dir}} = \boldsymbol{I}_{\mathcal{I}_g} - \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g}, \qquad m = \alpha^{-1} - 1,$$

one has

$$\begin{aligned} \boldsymbol{I}_{\mathcal{I}_g} - \alpha \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g} &= \alpha \left[\alpha^{-1} \boldsymbol{I}_{\mathcal{I}_g} - \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g} \right] \\ &= \alpha \left(\boldsymbol{L}_g^{\mathrm{Dir}} + m \boldsymbol{I}_{\mathcal{I}_g} \right). \end{aligned} \tag{83}$$

Therefore

$$(\boldsymbol{L}_g^{\mathrm{Dir}} + m \boldsymbol{I}_{\mathcal{I}_g}) \boldsymbol{u}_{\mathcal{I}_g} = \boldsymbol{S}_{\mathcal{I}_g g} \bar{u}_g. \tag{84}$$

The screened Dirichlet Green operator is

$$\boldsymbol{G}_g(m) = (\boldsymbol{L}_g^{\mathrm{Dir}} + m \boldsymbol{I}_{\mathcal{I}_g})^{-1}. \tag{85}$$

Thus

$$\boldsymbol{u}_{\mathcal{I}_g} = \boldsymbol{G}_g(m) \boldsymbol{S}_{\mathcal{I}_g g} \bar{u}_g. \tag{86}$$

Since $\boldsymbol{L}_g^{\mathrm{Dir}} \succ 0$, let

$$\boldsymbol{L}_g^{\mathrm{Dir}} \boldsymbol{\varphi}_n = \lambda_n \boldsymbol{\varphi}_n, \qquad 0 < \lambda_1 \le \lambda_2 \le \cdots.$$

Then

$$\boldsymbol{G}_g(m) = \sum_n \frac{\boldsymbol{\varphi}_n \boldsymbol{\varphi}_n^\top}{\lambda_n + m}. \tag{87}$$

As $m \to 0$,

$$\boldsymbol{G}_g(m) \to (\boldsymbol{L}_g^{\mathrm{Dir}})^{-1}. \tag{88}$$

J.2 Boundary-driven weak-cost limit

The boundary-driven clamped-goal field satisfies

$$\frac{\boldsymbol{u}_{\mathcal{I}_g}}{\bar{u}_g} = \boldsymbol{G}_g(m) \boldsymbol{S}_{\mathcal{I}_g g}.$$

Taking $m \to 0$ gives

$$\lim_{m \to 0} \frac{\boldsymbol{u}_{\mathcal{I}_g}}{\bar{u}_g} = (\boldsymbol{L}_g^{\mathrm{Dir}})^{-1} \boldsymbol{S}_{\mathcal{I}_g g}. \tag{89}$$

This vector can be evaluated using the null vector of $\boldsymbol{L}$. Since

$$\boldsymbol{L} \sqrt{\boldsymbol{\pi}} = 0,$$

the non-goal rows give

$$(\boldsymbol{I}_{\mathcal{I}_g} - \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g}) \sqrt{\boldsymbol{\pi}}_{\mathcal{I}_g} = \boldsymbol{S}_{\mathcal{I}_g g} \sqrt{\pi_g}.$$

Hence

$$(\boldsymbol{L}_g^{\mathrm{Dir}})^{-1} \boldsymbol{S}_{\mathcal{I}_g g} = \frac{\sqrt{\boldsymbol{\pi}}_{\mathcal{I}_g}}{\sqrt{\pi_g}}. \tag{90}$$

Therefore

$$\lim_{m \to 0} \psi_i^g = \sqrt{\frac{\pi_i}{\pi_g}}. \tag{91}$$

In desirability coordinates,

$$z_i^g = \sqrt{\frac{\pi_g}{\pi_i}} \psi_i^g,$$

so

$$\lim_{m \to 0} z_i^g = 1 \tag{92}$$

on the goal-connected component. This is the hitting-probability field.

Let

$$\epsilon_q = \frac{q_0}{\lambda}.$$

Since

$$\alpha = e^{-\epsilon_q},$$

Eq. (33) gives

$$z_i^g = \mathbb{E}_i[e^{-\epsilon_q \tau_g}].$$

On a finite component,

$$z_i^g = 1 - \epsilon_q \mathbb{E}_i[\tau_g] + \frac{\epsilon_q^2}{2}\mathbb{E}_i[\tau_g^2] + \mathcal{O}(\epsilon_q^3 \mathbb{E}_i[\tau_g^3]). \tag{93}$$

Thus

$$\log z_i^g = -\epsilon_q \mathbb{E}_i[\tau_g] + \frac{\epsilon_q^2}{2}\operatorname{Var}_i(\tau_g) + \mathcal{O}(\epsilon_q^3), \tag{94}$$

and

$$V_i^g = -\lambda \log z_i^g = q_0 \mathbb{E}_i[\tau_g] - \frac{q_0^2}{2\lambda}\operatorname{Var}_i(\tau_g) + \mathcal{O}(q_0^3/\lambda^2). \tag{95}$$

The log-gradient signal in desirability coordinates satisfies, for an edge $e = (i, j)$,

$$\nabla_e \log z^g = -\frac{q_0}{\lambda}\nabla_e \mathbb{E}[\tau_g] + \mathcal{O}\big((q_0/\lambda)^2\big). \tag{96}$$

Thus the spatial variation of the weak-cost desirability is carried by the hitting-time correction.

J.3 Dirichlet source-response Green function

For a localized source $\boldsymbol{s}$ on $\mathcal{I}_g$, consider

$$(\boldsymbol{L}_g^{\mathrm{Dir}} + m\boldsymbol{I})\boldsymbol{y}_m = \boldsymbol{s}. \tag{97}$$

Then

$$\boldsymbol{y}_m = \boldsymbol{G}_g(m)\boldsymbol{s},$$

and

$$\boldsymbol{y}_m \to (\boldsymbol{L}_g^{\mathrm{Dir}})^{-1}\boldsymbol{s} \qquad (m \to 0). \tag{98}$$

If $\boldsymbol{s} = \boldsymbol{e}_a$ for $a \in \mathcal{I}_g$, this limit is the a-th column of the Dirichlet Green function.

The boundary-driven field and an interior localized-source response have distinct weak-cost limits:

$$\boldsymbol{G}_g(m)\boldsymbol{S}_{\mathcal{I}_g g} \to \frac{\sqrt{\boldsymbol{\pi}}_{\mathcal{I}_g}}{\sqrt{\pi_g}},$$

whereas

$$\boldsymbol{G}_g(m)\boldsymbol{e}_a \to (\boldsymbol{L}_g^{\mathrm{Dir}})^{-1}\boldsymbol{e}_a.$$

J.4 Full-graph mean-zero Green function

On a connected component, the full Laplacian

$$\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{S}$$

has null vector

$$\boldsymbol{\varphi}_0 = \sqrt{\boldsymbol{\pi}}.$$

Let

$$\boldsymbol{P}_\perp = \boldsymbol{I} - \boldsymbol{\varphi}_0\boldsymbol{\varphi}_0^\top$$

be the Euclidean orthogonal projection onto the subspace orthogonal to $\sqrt{\boldsymbol{\pi}}$. The screened full-graph Green operator is

$$\boldsymbol{G}(m) = (\boldsymbol{L} + m\boldsymbol{I})^{-1}.$$

With orthonormal eigenpairs

$$\boldsymbol{L}\boldsymbol{\varphi}_n = \lambda_n \boldsymbol{\varphi}_n, \qquad n \geq 1, \qquad \lambda_n > 0,$$

the spectral expansion is

$$\boldsymbol{G}(m) = \frac{1}{m}\boldsymbol{\varphi}_0\boldsymbol{\varphi}_0^\top + \sum_{n\geq 1}\frac{\boldsymbol{\varphi}_n\boldsymbol{\varphi}_n^\top}{\lambda_n + m}. \tag{99}$$

For a source $\boldsymbol{s}$ satisfying

$$\langle \boldsymbol{s}, \boldsymbol{\varphi}_0\rangle = 0,$$

the divergent mode is absent and

$$\boldsymbol{G}(m)\boldsymbol{s} \to \boldsymbol{L}^\dagger \boldsymbol{s} \qquad (m \to 0), \tag{100}$$

where

$$\boldsymbol{L}^\dagger = \sum_{n\geq 1}\frac{\boldsymbol{\varphi}_n\boldsymbol{\varphi}_n^\top}{\lambda_n}$$

is the Moore–Penrose inverse on the $\sqrt{\boldsymbol{\pi}}$-orthogonal subspace.

For a point source at g in the symmetric network coordinate, the mean-zero source is

$$\boldsymbol{s}^{(g)} = \boldsymbol{P}_\perp \boldsymbol{e}_g = \boldsymbol{e}_g - \sqrt{\pi_g}\sqrt{\boldsymbol{\pi}}. \tag{101}$$

The corresponding full-graph mean-zero Green column is

$$\boldsymbol{g}_0^{(g)} = \boldsymbol{L}^\dagger\left(\boldsymbol{e}_g - \sqrt{\pi_g}\sqrt{\boldsymbol{\pi}}\right). \tag{102}$$

For any vector $\boldsymbol{w}$ in the symmetric network coordinate, the correct centering operation is

$$\mathcal{C}_{\boldsymbol{\pi}}(\boldsymbol{w}) = \boldsymbol{P}_\perp \boldsymbol{w} = \boldsymbol{w} - \sqrt{\boldsymbol{\pi}}\,\langle\sqrt{\boldsymbol{\pi}}, \boldsymbol{w}\rangle. \tag{103}$$

A scale-free shape normalization is

$$\mathcal{N}_{\boldsymbol{\pi}}(\boldsymbol{w}) = \frac{\mathcal{C}_{\boldsymbol{\pi}}(\boldsymbol{w})}{\|\mathcal{C}_{\boldsymbol{\pi}}(\boldsymbol{w})\|_2}, \qquad \mathcal{C}_{\boldsymbol{\pi}}(\boldsymbol{w}) \neq 0. \tag{104}$$

Arithmetic mean-centering coincides with Eq. (103) exactly in the uniform-$\boldsymbol{\pi}$ case.

For an uncentered point response,

$$\boldsymbol{G}(m)\boldsymbol{e}_g = \frac{\sqrt{\pi_g}}{m}\sqrt{\boldsymbol{\pi}} + \boldsymbol{L}^\dagger\left(\boldsymbol{e}_g - \sqrt{\pi_g}\sqrt{\boldsymbol{\pi}}\right) + \mathcal{O}(m). \tag{105}$$

Subtracting the $\sqrt{\boldsymbol{\pi}}$ component isolates the finite mean-zero Green function. If the full graph has several connected components, $\boldsymbol{P}_\perp$ is replaced by the projection onto the orthogonal complement of the full nullspace of $\boldsymbol{L}$.

K Local environmental edits and low-rank structure

K.1 Rank of a local connectivity edit

Let

$$\boldsymbol{S}' = \boldsymbol{S} + \Delta\boldsymbol{S}$$

be the recurrent matrix after a local environmental change. Closing or opening k undirected edges changes $\mathcal{O}(k)$ off-diagonal entries. Under the lazy normalization with fixed ϵ, the same edit changes the self-loop entries of the affected rows so that row sums remain one. Under standard random-walk normalization, the same rank scaling holds on bounded-degree graphs, because a changed edge rescales only the outgoing weights incident to the affected vertices. Thus

$$\operatorname{rank}(\Delta\boldsymbol{S}) \leq r_{\mathrm{edit}}, \qquad r_{\mathrm{edit}} = \mathcal{O}(k), \tag{106}$$

where k counts changed local matrix entries up to a bounded-degree constant.

Write a rank factorization

$$\Delta\boldsymbol{S} = \boldsymbol{U}_\Delta \boldsymbol{C}_\Delta \boldsymbol{V}_\Delta^\top, \tag{107}$$

with $\boldsymbol{U}_\Delta, \boldsymbol{V}_\Delta \in \mathbb{R}^{N\times r_{\mathrm{edit}}}$ and nonsingular $\boldsymbol{C}_\Delta \in \mathbb{R}^{r_{\mathrm{edit}}\times r_{\mathrm{edit}}}$, after discarding zero singular directions. A single directed entry change $\Delta S_{ab} = \delta$ is the rank-one matrix $\delta\boldsymbol{e}_a\boldsymbol{e}_b^\top$. A symmetric edge edit with lazy row-stochastic renormalization is a sum of a constant number of such rank-one terms.

K.2 Fixed-goal Dirichlet update

For a fixed goal g, define

$$\boldsymbol{A}_g = \boldsymbol{I} - \alpha \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g}, \qquad \boldsymbol{b}^{(g)} = \alpha \boldsymbol{S}_{\mathcal{I}_g g} \bar{u}_g.$$

The original field is

$$\boldsymbol{u}^{(g)} = \boldsymbol{A}_g^{-1} \boldsymbol{b}^{(g)}.$$

After the edit,

$$\boldsymbol{A}'_g = \boldsymbol{I} - \alpha \boldsymbol{S}'_{\mathcal{I}_g \mathcal{I}_g} = \boldsymbol{A}_g - \alpha \Delta \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g},$$

and

$$\boldsymbol{b}'^{(g)} = \boldsymbol{b}^{(g)} + \delta \boldsymbol{b}^{(g)}, \qquad \delta \boldsymbol{b}^{(g)} = \alpha \Delta \boldsymbol{S}_{\mathcal{I}_g g} \bar{u}_g.$$

Factor

$$-\alpha \Delta \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g} = \boldsymbol{U}_g \boldsymbol{C}_g \boldsymbol{V}_g^\top. \tag{108}$$

The factorization is chosen with nonsingular $\boldsymbol{C}_g$ on the retained rank. Then

$$\boldsymbol{A}'_g = \boldsymbol{A}_g + \boldsymbol{U}_g \boldsymbol{C}_g \boldsymbol{V}_g^\top.$$

Woodbury gives

$$(\boldsymbol{A}'_g)^{-1} = \boldsymbol{A}_g^{-1} - \boldsymbol{A}_g^{-1} \boldsymbol{U}_g \left(\boldsymbol{C}_g^{-1} + \boldsymbol{V}_g^\top \boldsymbol{A}_g^{-1} \boldsymbol{U}_g \right)^{-1} \boldsymbol{V}_g^\top \boldsymbol{A}_g^{-1}. \tag{109}$$

Therefore

$$\begin{aligned} \boldsymbol{u}'^{(g)} - \boldsymbol{u}^{(g)} &= \boldsymbol{A}_g^{-1} \delta \boldsymbol{b}^{(g)} \\ &\quad - \boldsymbol{A}_g^{-1} \boldsymbol{U}_g \left(\boldsymbol{C}_g^{-1} + \boldsymbol{V}_g^\top \boldsymbol{A}_g^{-1} \boldsymbol{U}_g \right)^{-1} \boldsymbol{V}_g^\top \boldsymbol{A}_g^{-1} \left(\boldsymbol{b}^{(g)} + \delta \boldsymbol{b}^{(g)} \right). \end{aligned} \tag{110}$$

The fixed-goal field change lies in the span of

$$\boldsymbol{A}_g^{-1} \boldsymbol{U}_g$$

together with

$$\boldsymbol{A}_g^{-1} \delta \boldsymbol{b}^{(g)}.$$

Under fixed-ϵ lazy normalization, if no edited matrix entry lies in the goal column $\boldsymbol{S}_{\mathcal{I}_g g}$, then

$$\delta \boldsymbol{b}^{(g)} = 0,$$

and the field change lies in an $\mathcal{O}(k)$-dimensional subspace. Under random-walk normalization, $\delta \boldsymbol{b}^{(g)}$ can also change when an edit changes the degree of a vertex adjacent to g, or changes d_g. When the boundary column changes, $\delta \boldsymbol{b}^{(g)}$ is supported on the locally edited boundary entries under lazy normalization and on the affected goal-neighborhood entries under bounded-degree random-walk normalization, contributing another $\mathcal{O}(k)$-dimensional subspace.

K.3 Shared full-resolvent update

Let

$$\boldsymbol{R} = (\boldsymbol{I} - \alpha \boldsymbol{S})^{-1}, \qquad \boldsymbol{R}' = (\boldsymbol{I} - \alpha \boldsymbol{S}')^{-1}.$$

Factor

$$-\alpha \Delta \boldsymbol{S} = \boldsymbol{U} \boldsymbol{C} \boldsymbol{V}^\top. \tag{111}$$

The factorization is chosen with nonsingular $\boldsymbol{C}$ on the retained rank. Then

$$\boldsymbol{I} - \alpha \boldsymbol{S}' = \boldsymbol{I} - \alpha \boldsymbol{S} + \boldsymbol{U} \boldsymbol{C} \boldsymbol{V}^\top.$$

Woodbury gives

$$\boldsymbol{R}' = \boldsymbol{R} - \boldsymbol{R}\boldsymbol{U} \left(\boldsymbol{C}^{-1} + \boldsymbol{V}^\top \boldsymbol{R}\boldsymbol{U} \right)^{-1} \boldsymbol{V}^\top \boldsymbol{R}. \tag{112}$$

Hence

$$\boldsymbol{R}' - \boldsymbol{R} = -\boldsymbol{R}\boldsymbol{U} \left(\boldsymbol{C}^{-1} + \boldsymbol{V}^\top \boldsymbol{R}\boldsymbol{U} \right)^{-1} \boldsymbol{V}^\top \boldsymbol{R}, \tag{113}$$

and

$$\operatorname{rank}(\boldsymbol{R}' - \boldsymbol{R}) \le r_{\text{edit}} = \mathcal{O}(k). \tag{114}$$

For a fixed source matrix $\boldsymbol{B}$ independent of the edit,

$$\boldsymbol{F} = \boldsymbol{R}\boldsymbol{B}, \qquad \boldsymbol{F}' = \boldsymbol{R}'\boldsymbol{B},$$

so

$$\boldsymbol{F}' - \boldsymbol{F} = (\boldsymbol{R}' - \boldsymbol{R})\boldsymbol{B} \tag{115}$$

has rank at most r_{edit}. If the source matrix changes by a local low-rank term $\Delta\boldsymbol{B}$, then

$$\boldsymbol{F}' = \boldsymbol{R}'(\boldsymbol{B} + \Delta\boldsymbol{B}),$$

and

$$\boldsymbol{F}' - \boldsymbol{F} = (\boldsymbol{R}' - \boldsymbol{R})\boldsymbol{B} + \boldsymbol{R}'\Delta\boldsymbol{B}, \tag{116}$$

with

$$\operatorname{rank}(\boldsymbol{F}' - \boldsymbol{F}) \le r_{\text{edit}} + \operatorname{rank}(\Delta\boldsymbol{B}).$$

K.4 Unit-clamped all-goal fields

A unit-clamped goal field can be represented through the full resolvent. For a source $c_g\boldsymbol{e}_g$,

$$\boldsymbol{u}^{(g)} = c_g\boldsymbol{R}\boldsymbol{e}_g.$$

Choosing

$$c_g = \frac{\bar{u}}{R_{gg}}$$

enforces

$$u_g^{(g)} = \bar{u}.$$

For all goals with common clamp amplitude $\bar{u}$, define

$$\boldsymbol{\Psi} = \boldsymbol{R}\boldsymbol{D}_{\text{clamp}}, \qquad \boldsymbol{D}_{\text{clamp}} = \bar{u}\ \operatorname{diag}(R_{11}^{-1}, \dots, R_{NN}^{-1}). \tag{117}$$

If $\boldsymbol{S} \ge 0$, then

$$\boldsymbol{R} = \boldsymbol{I} + \alpha\boldsymbol{S} + \alpha^2\boldsymbol{S}^2 + \cdots$$

has nonnegative entries and

$$R_{gg} \ge 1,$$

so $\boldsymbol{D}_{\text{clamp}}$ is well defined.

After the edit,

$$\boldsymbol{\Psi}' = \boldsymbol{R}'\boldsymbol{D}'_{\text{clamp}}, \qquad \boldsymbol{D}'_{\text{clamp}} = \bar{u}\ \operatorname{diag}(R_{11}'^{-1}, \dots, R_{NN}'^{-1}).$$

Thus

$$\boldsymbol{\Psi}' - \boldsymbol{\Psi} = (\boldsymbol{R}' - \boldsymbol{R})\boldsymbol{D}_{\text{clamp}} + \boldsymbol{R}'(\boldsymbol{D}'_{\text{clamp}} - \boldsymbol{D}_{\text{clamp}}). \tag{118}$$

The first term has rank at most $r_{\text{edit}} = \mathcal{O}(k)$. The second term is the source-renormalization term required to keep every goal clamped to the same amplitude. Its algebraic rank can be full because the diagonal entries R_{gg} can change for many goals.

The full-rank source-renormalization term is spatially localized in goal index under an exponential resolvent envelope. Let $\mathcal{V}_{\text{edit}}$ be the set of vertices incident to the edited entries. Suppose

$$|R_{ij}| \le C_R e^{-\kappa_R d(i,j)}, \qquad |R'_{ij}| \le C'_R e^{-\kappa_R d(i,j)}$$

and the Woodbury middle factor

$$\left(\boldsymbol{C}^{-1} + \boldsymbol{V}^\top\boldsymbol{R}\boldsymbol{U}\right)^{-1}$$

is uniformly bounded, with $\boldsymbol{U}$ and $\boldsymbol{V}$ supported on $\mathcal{V}_{\text{edit}}$ and its edited neighborhood. Then Eq. (113) gives

$$|\Delta R_{gg}| = |R'_{gg} - R_{gg}| \le C_{\Delta R} e^{-2\kappa_R d(g, \mathcal{V}_{\text{edit}})}. \tag{119}$$

Since $R_{gg} \ge 1$ and $R'_{gg} \ge 1$,

$$|D'_{\text{clamp},gg} - D_{\text{clamp},gg}| = \bar{u}\left|\frac{1}{R'_{gg}} - \frac{1}{R_{gg}}\right| \le C_D e^{-2\kappa_R d(g, \mathcal{V}_{\text{edit}})}. \tag{120}$$

For a truncation radius R_{edit}, retain only goals satisfying

$$d(g, \mathcal{V}_{\mathrm{edit}}) \le R_{\mathrm{edit}}.$$

The retained source-renormalization term has rank at most the number of goals in that edited-neighborhood ball. The discarded diagonal entries obey

$$|D'_{\mathrm{clamp},gg} - D_{\mathrm{clamp},gg}| \le C_D e^{-2\kappa_R R_{\mathrm{edit}}}.$$

On graphs with subexponential or sufficiently slow volume growth, the discarded Frobenius energy decays exponentially in R_{edit}. As α approaches one, the decay rate κ_R may decrease and resolvent constants may grow, so this locality estimate weakens in the weak-cost limit.

For each individual goal, multiplying the field by a scalar leaves the log-gradient unchanged:

$$\nabla \log(c_g f^{(g)}) = \nabla \log f^{(g)} \qquad (c_g > 0).$$

Thus diagonal source renormalization affects raw all-goal field matrices and cancels exactly in per-goal log-gradient readouts.

L Numerical experiment details

This section gives the numerical protocol used for all simulations in Sec. 4. The purpose of the simulations is to check four claims made in the main text and in the theoretical appendix. First, the linear goal-clamped network must reproduce the LMDP desirability field with the stationary-measure correction in Theorem 1. Second, the same readout must remain stable when a strong goal clamp pushes local neurons into the nonlinear range of the activation function. Third, transition-gated BTSP must write enough legal local connectivity during goal-free exploration to plan through a maze with obstacles, and a local environmental edit must produce the low-rank remapping predicted by App. K.4. Fourth, the same recurrent resolvent must connect to the screened Green function in App. J.4, and the effect of asymmetric BTSP sampling must be controlled by the perturbation scale derived in App. F.2. Additional experiments that test the scope of these assumptions, measure the realized relaxation dynamics, and compare the circuit with learned goal-conditioned models are reported in App. M.

L.1 Common graph, field, and readout conventions

Each open grid cell was represented by one assembly. Coordinates were integer cell centers, denoted by $\boldsymbol{x}_i \in \mathbb{R}^2$. Unless a specific experiment below states a different normalization, the accessible transition memory was a symmetric weighted grid graph with eight-neighbor connectivity. A horizontal or vertical step and a diagonal step were both allowed when the two end cells were open. Diagonal corner cutting was removed: a diagonal edge was kept only when the two associated axial edges were also legal. The raw edge weight was

$$W_{ij} = \exp\left(-\frac{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|_2^2}{2\sigma_e^2}\right), \qquad \sigma_e = 1. \tag{121}$$

Thus axial edges had weight $e^{-1/2}$, diagonal edges had weight e^{-1}, and blocked transitions had weight zero. The matrix was symmetrized after edge construction.

For the equivalence, BTSP-learning, low-rank, and Green-limit simulations, the symmetric transition memory $\boldsymbol{W}$ was converted into a passive kernel by the standard random-walk normalization of App. E.4:

$$P_0(j \mid i) = \frac{W_{ij}}{d_i}, \qquad d_i = \sum_j W_{ij}, \qquad \pi_i = \frac{d_i}{\sum_\ell d_\ell}. \tag{122}$$

The recurrent matrix used in the linear network was the symmetric coordinate

$$\boldsymbol{S} = \boldsymbol{\Pi}^{1/2} \boldsymbol{P}_0 \boldsymbol{\Pi}^{-1/2}. \tag{123}$$

This choice makes the stationary-measure factor in Theorem 1 visible near doors, walls, and boundaries. The asymmetric BTSP experiment used a fixed-ϵ lazy normalization instead; that case is described separately in App. L.5.

For a recalled goal g, the LMDP desirability field was computed from

$$(\boldsymbol{I}_{\mathcal{I}_g} - \alpha \boldsymbol{P}_{0,\mathcal{I}_g\mathcal{I}_g})\boldsymbol{z}^g_{\mathcal{I}_g} = \alpha \boldsymbol{P}_{0,\mathcal{I}_g g}, \qquad z^g_g = 1. \tag{124}$$

The linear recurrent field was computed from

$$(\boldsymbol{I}_{\mathcal{I}_g} - \alpha \boldsymbol{S}_{\mathcal{I}_g\mathcal{I}_g})\boldsymbol{u}_{\mathcal{I}_g} = \alpha \boldsymbol{S}_{\mathcal{I}_g g}\bar{u}_g, \qquad u_g = \bar{u}_g. \tag{125}$$

The normalized network field was $\psi^g_i = u_i/\bar{u}_g$. When $\boldsymbol{\pi}$ was nonuniform, the LMDP-coordinate field was

$$\widehat{z}_i = \sqrt{\frac{\pi_g}{\pi_i}}\,\psi^g_i, \qquad \widehat{z}_g = 1. \tag{126}$$

This is the numerical version of Eq. (6). Sparse linear systems were solved directly with SciPy sparse solvers.

The place-tuned readout was the discrete version of Eq. (7). At a current state i, a graph-geodesic Gaussian place kernel was formed around i,

$$\mathcal{G}_i(j) = \exp\left[-\frac{d_{\mathrm{graph}}(i,j)^2}{2\sigma_G^2}\right], \qquad d_{\mathrm{graph}}(i,j) \le 3\sigma_G, \tag{127}$$

where graph distance used Euclidean edge lengths on the legal graph. The population-vector readout was

$$\Delta\boldsymbol{x}_{\mathrm{pop}}(i;\boldsymbol{f}) = \frac{\sum_j \mathcal{G}_i(j) f_j (\boldsymbol{x}_j - \boldsymbol{x}_i)}{\sum_j \mathcal{G}_i(j) f_j}, \tag{128}$$

with $\boldsymbol{f} = \boldsymbol{z}^g$, $\widehat{\boldsymbol{z}}$, or the relevant normalized source-response field. Readout-angle errors were computed as the angle between two such vectors at the same state, omitting states where either vector had norm below numerical precision.

For rollout figures, the continuous readout vector was projected onto the legal neighboring moves. If $\boldsymbol{v}_i$ is the readout at state i, each neighbor j received the score

$$\mathrm{score}(j \mid i) = \frac{(\boldsymbol{x}_j - \boldsymbol{x}_i)^\top \boldsymbol{v}_i}{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|_2 \, \|\boldsymbol{v}_i\|_2} + w_f \frac{f_j - f_i}{|f_i| + 10^{-12}} - w_\ell N_j, \tag{129}$$

where N_j is the number of previous visits to j. The next state was the legal neighbor with the largest score. The pure readout rollouts used $w_f = w_\ell = 0$. Some diagnostic rollouts used a small field-gain or loop penalty, and the exact values are listed below. A route was counted as successful when it reached the goal state, or when the bug-trap protocol used a specified goal-zone radius. Route length was the sum of Euclidean lengths of successive grid moves. Route ratio was route length divided by the shortest legal graph path length under the same success criterion.

Relative field errors were reported as

$$\varepsilon_{\mathrm{field}}(\boldsymbol{a},\boldsymbol{b}) = \frac{\|\boldsymbol{a} - \boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2 + 10^{-15}}. \tag{130}$$

Field correlations were Pearson correlations across all open cells. All plots of fields used $\log(\boldsymbol{f} + 10^{-10})$ or $\log(\boldsymbol{f} + 10^{-12})$ only for display; metrics were computed on the untransformed fields unless stated otherwise.

L.2 Equivalence and strong-recall robustness

The first experiment checks the exact theorem before adding learning or maze changes. The maze was a two-room diagnostic environment with a zero-thickness vertical wall. The default size was 68×48 grid cells, with the wall between columns 33 and 34, a near door over rows 10–15, and a far door over rows 34–40. The start was $(9, 12)$, the goal was $(58, 37)$, and the legal graph had 3036 states and 11709 undirected edges. This maze was chosen because the bottlenecks make $\boldsymbol{\pi}$ mildly nonuniform, so the factor $\sqrt{\pi_i/\pi_g}$ in Theorem 1 can be tested directly.

For the linear equivalence check, Eq. (124) and Eq. (125) were solved at

$$\alpha \in \{0.90, 0.94, 0.97, 0.985, 0.992, 0.996\}, \qquad \bar{u}_g = 1. \tag{131}$$

The network field was converted to the LMDP coordinate by Eq. (126). At the main setting $\alpha = 0.985$, the corrected field matched the LMDP desirability with relative error 9.55×10^{-16} and correlation 1.000. Across the whole α grid, the corrected relative error stayed between 6.29×10^{-16} and 9.59×10^{-16}. The uncorrected field $\boldsymbol{\psi}^g$ differed from $\boldsymbol{z}^g$ by 3.17% at $\alpha = 0.985$, matching the size of the stationary-measure correction. This panel tests the algebraic identity in Theorem 1; the uncorrected points show why the $\sqrt{\boldsymbol{\pi}}$ factor is needed in a nonuniform environment.

The nonlinear recall test used the same maze and $\alpha = 0.985$. The activity nonlinearity was

$$\phi(u) = \tanh u. \tag{132}$$

For each clamp amplitude

$$A = \bar{u}_g \in \{0.05, 0.10, 0.20, 0.40, 0.80, 1.60, 3.20, 5.00\}, \tag{133}$$

the non-goal fixed point was iterated from the linear solution using

$$\boldsymbol{u}_{\mathcal{I}_g} = \alpha \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g} \tanh(\boldsymbol{u}_{\mathcal{I}_g}) + \alpha \boldsymbol{S}_{\mathcal{I}_g g} \tanh A, \qquad u_g = A. \tag{134}$$

Damped fixed-point iteration used damping factor 0.72 and relative tolerance 2×10^{-11} for the one-dimensional amplitude sweep. The nonlinear field was first mapped to the LMDP coordinate by Eq. (126) and then aligned to the linear field by a single scalar least-squares factor. This scalar alignment removes the uniform amplitude change that cannot affect Eq. (7).

The visible nonlinear shell was measured through local gain loss. For the tanh unit,

$$1 - \frac{\phi'(u)}{\phi'(0)} = \tanh^2 u. \tag{135}$$

Cells were binned by graph distance from the goal. In each distance bin, the 90th percentile of Eq. (135) was computed. The shell radius was the largest distance bin whose 90th percentile exceeded 3×10^{-3}. This radius measures where the local gain has visibly left the small-signal regime. At $A = 5$, the radius was eight graph-distance cells. Outside that shell, the population-vector readout remained close to the linear prediction: the mean angle error was 0.080°, and the 99th percentile was 0.60°. At a probe location near the door and far from the goal, the angle error was 0.015°. The shell radius grew approximately as $\log A$, with $R^2 = 0.94$. This directly tests the shell-radius estimate in App. I.4: strong recall changes the local shape near the clamp, while the action-relevant log-gradient remains stable outside that local region.

A two-parameter stress sweep repeated the nonlinear solve for every α in Eq. (131) and every A in Eq. (133), using tolerance 5×10^{-10}. The displayed main row $\alpha = 0.985$ stayed below a 1° outside-shell 99th percentile across the whole recall range; its maximum was 0.60°. This stress map separates two quantities that can be conflated visually: the field amplitude can be strongly compressed near the goal, yet the far-field readout that drives behavior changes by less than one degree in the main regime.

L.3 Goal-free BTSP learning and local low-rank remapping

The second experiment asks whether the recurrent operator needed for planning can be written by local goal-free transition events. It also asks whether a local environmental edit produces the low-rank field remapping predicted by App. K.4. The experiment has two parts: a challenging bug-trap navigation task and a smaller two-door remapping diagnostic.

Goal-free exploration was simulated as short novelty-biased bouts on the legal graph. At the beginning of each bout, the animal was placed in a low-visit region, defined by the lowest quartile of current visit counts. During a bout, the next state was sampled among legal neighbors using the score

$$q(j \mid i) = -1.20\, n_j - 0.80\, n_{\{i,j\}} + 0.55\, h_{ij} + 0.05\, \xi_j. \tag{136}$$

Here n_j is the visit count of state j, $n_{\{i,j\}}$ is the visit count of the undirected edge $\{i, j\}$, h_{ij} is the cosine alignment between the proposed step and the current heading, and ξ_j is a standard normal random variable. A softmax with temperature 0.55 converted these scores into choice probabilities. This exploration rule has no goal input and no shortest-path information. It samples under-visited local transitions while keeping a mild heading persistence.

Each sampled transition $i \to j$ added one count to both W_{ij} and W_{ji}. Thus the learned memory is the symmetric BTSP transition memory in Sec. 3.1. A self-only numerical floor of 10^{-7} was added before normalization so that isolated rows during early sparse learning did not make the sparse linear system singular. This floor did not add untraversed transition geometry. Because every event was sampled from the legal graph, the learned edge precision was exactly 1.000 in all BTSP-learning runs; the main variable was edge recall.

Bug-trap navigation. The bug-trap maze was 112×76 cells. The start was $(102, 45)$, the goal was $(72, 43)$, and the legal graph had 7285 states and 27669 undirected edges. The start was only 30.1 Euclidean cells from the goal, but the shortest legal route to the goal zone had length 110.43. The goal-zone radius was 1.5 cells. This geometry makes the task useful scientifically: a local cue pointing straight toward the goal hits a wall, so successful navigation requires a field that has already integrated the global obstacle layout.

The bug-trap field used $\alpha = 0.993$. The readout width was $\sigma_G = 3.2$, and pure readout rollouts used $w_f = w_\ell = 0$, maximum length 3600 steps, and goal-zone radius 1.5. Five exploration seeds were used: $4, 15, 27, 41, 57$. Performance was measured at $1.0, 1.5, 2.0, 2.5$, and 3.0 goal-free transitions per assembly. At the displayed milestone, 3.0 transitions per assembly corresponded to 21855 sampled transition events.

At this milestone, edge recall was 68.9% on average across the five seeds, precision was 1.000, and the learned field correlated with the ideal-maze field at 0.988 ± 0.003 (mean $\pm$ SEM). All five seeds reached the goal zone using the pure population-vector readout. The median route ratio was 1.16. For the displayed seed, the pure readout route took 110 legal steps with route ratio 1.16; the first readout vector was 93.8° away from the direct Euclidean vector to the goal. A Euclidean greedy baseline, which chose the legal neighbor most aligned with the direct goal vector and used only a small revisit penalty, stalled at the wall after 16 legal steps. This result tests the core claim that local goal-free BTSP can write a map whose relaxation solves a global obstacle-avoidance query.

Two-door remapping diagnostic. The low-rank remapping diagnostic used a smaller 60×42 two-door maze with the wall between columns 29 and 30, a near door over rows 8–13, and a far door over rows 29–35. The start was $(8, 10)$, the goal was $(51, 32)$, and the graph had 2320 states and 8905 undirected edges. The discount was $\alpha = 0.988$, the readout width was $\sigma_G = 2.9$, and rollouts used maximum length 2200, zero goal radius, and loop penalty $w_\ell = 0.02$. The exploration checkpoints were $0.15, 0.30, 0.50, 0.80, 1.10, 1.50, 2.00$, and 2.80 transitions per assembly, again across seeds $4, 15, 27, 41, 57$.

The displayed remapping point was 2.0 transitions per assembly. At this sparse milestone, the learned graph had mean edge recall 49.4%, precision 1.000, and learned-versus-ideal field correlation 0.973 ± 0.006. All five seeds reached the goal, with median route ratio 1.27. For the displayed seed, the open-door route ratio was 1.20.

The local edit closed the near door by removing the 16 coordinate-level crossing edges that passed through that door. No other learned weights were changed, and no new exploration was run. After the edit, the next relaxation rerouted the displayed goal through the far door; the displayed closed-door route ratio was 1.09. For the Woodbury check, the effective support of $\Delta \boldsymbol{S}$ above numerical threshold 10^{-10} involved 26 rows. A direct solve on the closed-door operator and the Woodbury update of App. K.2 agreed with relative error 1.82×10^{-15}.

To test whether the same local edit has low rank across many goals, 34 goal states were selected by taking spatial quantiles of the score $\boldsymbol{q}^\top \boldsymbol{x}_i$, with $\boldsymbol{q} = (1, 0.7)^\top$. For each sampled goal, the open-door and closed-door linear network fields were solved, corrected to the LMDP coordinate by Eq. (126), and differenced. These 34 field differences were stacked as columns of a matrix. The first two singular modes explained 99.46% of its total squared energy. This is the numerical version of the low-rank statement in Sec. 3.2: a local edit produces a small set of propagation patterns that explains the field change for many recalled goals.

L.4 Weak-cost Green-function limit

The Green-limit experiment used the same 60×42 two-door diagnostic maze as the remapping diagnostic. It checks the operator identity in App. J.4 and clarifies which Green object is being visual-

ized. The clamped boundary field ψ^g has a flat hitting-probability limit in desirability coordinates as $\alpha \to 1$. To isolate the Laplacian Green object used for spectral navigation, the numerical experiment used the full-graph point-source response

$$\boldsymbol{r}_\alpha^{(g)} = (\boldsymbol{I} - \alpha \boldsymbol{S})^{-1} \boldsymbol{e}_g. \tag{137}$$

With $\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{S}$ and

$$m_\alpha = \frac{1-\alpha}{\alpha}, \tag{138}$$

the exact algebraic relation is

$$\alpha \boldsymbol{r}_\alpha^{(g)} = (\boldsymbol{L} + m_\alpha \boldsymbol{I})^{-1} \boldsymbol{e}_g. \tag{139}$$

This identity was tested at

$$\alpha \in \{0.90, 0.94, 0.97, 0.985, 0.992, 0.996, 0.998, 0.999, 0.9995\}. \tag{140}$$

The relative error in Eq. (139) stayed at floating-point precision, with maximum value 2.67×10^{-13} at $\alpha = 0.9995$.

The weak-cost comparison used the mean-zero Laplacian Green column in the symmetric network coordinate. Let

$$\boldsymbol{\varphi}_0 = \sqrt{\boldsymbol{\pi}}, \qquad \boldsymbol{P}_\perp = \boldsymbol{I} - \boldsymbol{\varphi}_0 \boldsymbol{\varphi}_0^\top. \tag{141}$$

The reference was

$$\boldsymbol{G}_0^{(g)} = \boldsymbol{L}^\dagger \left(\boldsymbol{e}_g - \sqrt{\pi_g} \sqrt{\boldsymbol{\pi}} \right), \tag{142}$$

computed by a sparse least-squares solve with absolute and relative tolerances 10^{-10}. For each m_α, the screened Green field was projected by $\boldsymbol{P}_\perp$ and normalized to unit Euclidean norm before comparison with Eq. (142). The resulting shape error decreased from 1.168 at $m_\alpha = 0.111$ to 0.162 at $m_\alpha = 5.0 \times 10^{-4}$, showing convergence to the Laplacian Green shape as the cost decreases.

The spectral reconstruction in Fig. 5 used the 72 smallest eigenpairs of $\boldsymbol{L}$, computed by sparse symmetric eigensolver. The screened Green expansion was

$$\boldsymbol{G}_{m_\alpha}^{(g)} = \sum_n \frac{(\boldsymbol{\varphi}_n)_g \, \boldsymbol{\varphi}_n}{\lambda_n + m_\alpha}. \tag{143}$$

Panel A of Fig. 5 shows the non-null gains normalized by the first non-null gain and the shape error after adding successive low-frequency modes. This panel checks that the Green field is dominated by low-frequency Laplacian modes in the weak-cost regime, matching the spectral interpretation in Sec. 2.4.

For the displayed field, $\alpha = 0.996$, so $m_\alpha = 0.00402$. The plotted field was $\boldsymbol{G}_{m_\alpha}^{(g)}$ normalized by its goal value; white contours show the independently computed recurrent source response $\boldsymbol{r}_\alpha^{(g)}$, also normalized by its goal value. The two agree with relative error 2.2×10^{-15}. The rollout in the displayed panel used $\sigma_G = 2.9$, maximum length 1800, zero goal radius, and loop penalty $w_\ell = 0.02$.

L.5 Asymmetric BTSP perturbation and bilateral replay

The exact theorem assumes a reversible passive kernel. Empirical BTSP sampling can contain directional residue from biased movement, finite sampling, or forward-dominant sequence statistics. The asymmetric BTSP experiment asks a more precise question than generic robustness: whether all deviations from the reversible reference are ordered by the single dimensionless perturbation scale

$$\chi = \frac{\alpha \|\boldsymbol{E}_{\mathrm{asym}}\|_2}{1-\alpha} \tag{144}$$

from App. F.2. This question is quantitative and falsifiable. If η, α, maze position, or sampling seed controlled the error separately, field error and readout error would scatter when plotted against χ. A collapse onto one curve supports the perturbation analysis.

The experiment used the same 112×76 bug-trap maze as the main BTSP navigation task. The start, goal, graph size, and goal-zone radius were the same as in App. L.3. A directed effective count matrix $\boldsymbol{C}$ was built from the legal directed edges of the maze. The symmetric legal support came from the same edge weights in Eq. (121). A weak circulation bias was then added to make C_{ij} and C_{ji} slightly different while keeping the support legal.

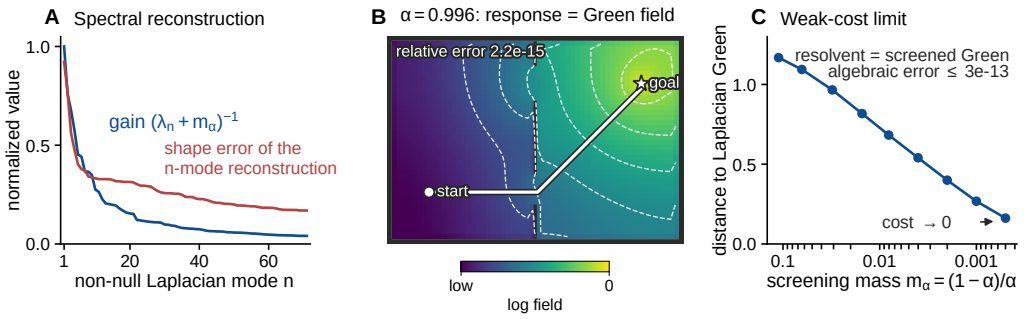


Figure 5: **The recurrent source response is a screened Green function and approaches the Laplacian Green shape in the weak-cost limit.** (**A**) Spectral reconstruction of the screened Green response from the low-frequency eigenmodes of $\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{S}$: normalized gains $(\lambda_n + m_\alpha)^{-1}$ (blue) and the shape error of the reconstruction after adding successive modes (red). (**B**) At $\alpha = 0.996$, the source response of the recurrent network (white contours) matches the screened Green field (color) to relative error 2.2×10^{-15}; the white route is the population-vector rollout. (**C**) As $m_\alpha = (1-\alpha)/\alpha$ decreases, the mean-zero screened Green shape approaches the Laplacian pseudoinverse Green column. The algebraic identity between the recurrent resolvent and the screened Green operator holds to 3×10^{-13} or better throughout the sweep.

For each directed edge $i \to j$, define the unit step

$$\boldsymbol{s}_{ij} = \frac{\boldsymbol{x}_j - \boldsymbol{x}_i}{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|_2}, \tag{145}$$

the edge midpoint $\boldsymbol{m}_{ij} = (\boldsymbol{x}_i + \boldsymbol{x}_j)/2$, and a circulation center $\boldsymbol{c}$. The tangent direction around that center was

$$\boldsymbol{t}_{ij} = \frac{[-(m_{ij,y} - c_y),\ m_{ij,x} - c_x]^\top}{\|[-(m_{ij,y} - c_y),\ m_{ij,x} - c_x]^\top\|_2}. \tag{146}$$

The directional score was

$$s_{ij}^{\text{flow}} = \text{clip}(\boldsymbol{s}_{ij}^\top \boldsymbol{t}_{ij}, -1, 1), \tag{147}$$

and the directed count rate was

$$C_{ij} = W_{ij} \exp(\rho s_{ij}^{\text{flow}}). \tag{148}$$

Across sampling seeds, the center was drawn as $\boldsymbol{c} = (70, 38) + \boldsymbol{\xi}$, with standard deviations $(1.4, 1.0)$ in the two coordinates, and the strength was $\rho = 0.050 \exp(\zeta)$, with $\zeta \sim \mathcal{N}(0, 0.07^2)$. The sampling seeds were $4, 15, 27, 41, 57$. This construction gives a controlled directional residue without adding illegal edges or changing the maze.

The asymmetry knob was

$$\boldsymbol{W}^\eta = (1+\eta)\boldsymbol{C} + (1-\eta)\boldsymbol{C}^\top = \boldsymbol{C} + \boldsymbol{C}^\top + \eta(\boldsymbol{C} - \boldsymbol{C}^\top), \tag{149}$$

with

$$\eta \in \{0, 0.05, 0.10, 0.20, 0.40, 0.60, 0.80\}. \tag{150}$$

The off-diagonal symmetric support $\boldsymbol{C} + \boldsymbol{C}^\top$ was held fixed, and η scaled the antisymmetric residue $\boldsymbol{C} - \boldsymbol{C}^\top$. For each sampling seed, a single global lazy-normalization constant was chosen as

$$\epsilon = \frac{0.88}{\max_\eta \max_i \sum_{j \neq i} W_{ij}^\eta}. \tag{151}$$

The raw asymmetric passive kernel was then

$$P_\eta(j \mid i) = \epsilon W_{ij}^\eta \quad (j \neq i), \qquad P_\eta(i \mid i) = 1 - \epsilon \sum_{j \neq i} W_{ij}^\eta. \tag{152}$$

This fixed-ϵ rule keeps the maximum off-diagonal row mass at 0.88, so every self-loop is nonnegative. The diagonal lazy term changes only to keep each row stochastic. It contributes no antisymmetric part.

The reversible reference for the same seed was $\boldsymbol{P}_0$, obtained by setting $\eta = 0$. For each η, the antisymmetric perturbation in the uniform matrix coordinate was

$$\boldsymbol{E}^{\eta}_{\mathrm{asym}} = \frac{1}{2}\left[(\boldsymbol{P}_\eta - \boldsymbol{P}_0) - (\boldsymbol{P}_\eta - \boldsymbol{P}_0)^\top\right] = \eta\epsilon(\boldsymbol{C} - \boldsymbol{C}^\top). \tag{153}$$

The spectral norm $\|\boldsymbol{E}^{\eta}_{\mathrm{asym}}\|_2$ was computed with a largest-singular-value sparse solver. The sweep used

$$\alpha \in \{0.970, 0.985, 0.993, 0.997\}. \tag{154}$$

Together with the five seeds and seven η values, this gave 140 raw asymmetric conditions.

For each raw condition, the Dirichlet field $\boldsymbol{z}^g_\eta$ was solved from Eq. (124) with $\boldsymbol{P}_\eta$ in place of $\boldsymbol{P}_0$. The reference field $\boldsymbol{z}^g_0$ used the same seed and α, with $\eta = 0$. The field error was $\|\boldsymbol{z}^g_\eta - \boldsymbol{z}^g_0\|_2/\|\boldsymbol{z}^g_0\|_2$. The readout-angle error was the median angle across 148 probe states. Probe states were selected on a stride-seven grid, excluding states within five graph steps of the goal, and the set was augmented by the start and three corridor or chamber states $(52, 25)$, $(84, 43)$, and $(96, 24)$. This probe set tests readout changes away from the clamped goal, where behavior is chosen.

Rollouts in this experiment used the pure population-vector rule with $\sigma_G = 3.2$, $w_f = w_\ell = 0$, maximum length 1200, and goal-zone radius 1.5. Route ratio was normalized by the same shortest path to the goal zone, 110.43, used in the bug-trap BTSP task.

The bilateral replay control projected each raw asymmetric transition memory back to its reversible part:

$$\frac{1}{2}\left(\boldsymbol{W}^\eta + (\boldsymbol{W}^\eta)^\top\right) = \boldsymbol{C} + \boldsymbol{C}^\top. \tag{155}$$

After the same fixed-ϵ lazy normalization, this control is exactly $\boldsymbol{P}_0$ for the corresponding seed. It represents a circuit in which forward and reverse replay, or other bilateral sampling, cancels the directional residue before the recurrent map is used, consistent with bidirectional hippocampal replay motifs (Pfeiffer and Foster, 2013; Widloski and Foster, 2022). The control was plotted at the same raw χ values, so it asks whether the same asymmetric experience can be rescued by a reversible projection.

Figure 6 is organized as a mechanistic chain. The perturbation knob η scales only the directional residue in Eq. (149), while the quantitative axis used later is χ, the perturbation size after multiplication by the recurrent resolvent gain $\alpha/(1-\alpha)$. This matters because the same raw asymmetry is more dangerous when α is closer to one.

Panels A–C show one representative seed at $\alpha = 0.985$. The displayed conditions are $\eta = 0$, $\eta = 0.10$, and $\eta = 0.60$, with $\chi = 0.00$, 0.19, and 1.11. Panel A is the reference log field $\log \boldsymbol{z}^g_0$ with the reference rollout. Panels B and C show the signed difference $\Delta \log \boldsymbol{z}^g_\eta = \log \boldsymbol{z}^g_\eta - \log \boldsymbol{z}^g_0$, with red indicating that the asymmetric field is higher than the reference and blue indicating that it is lower, together with the rollout on the asymmetric field (white) and the reference rollout (grey). The perturbation visibly twists the field, and the twist grows continuously with χ. At $\eta = 0$ and $\eta = 0.10$, the rollout reaches the goal with route ratio 1.02. At $\eta = 0.60$, the rollout still reaches the goal, with route ratio 1.13, after taking a visibly shifted route. Complete raw failures occur in the larger sweep shown in panel F.

Panels D and E test the scaling law. Color encodes α, each point is one seed–η condition, and both panels use a logarithmic χ axis. The field error and the median readout-angle error both collapse onto single increasing trends when plotted against χ. In the green band $\chi \le 0.3$, the raw asymmetric branch has median field error 0.0040, 75th-percentile field error 0.0070, and median readout-angle error $0.13°$. All raw rollouts in this band succeed, with median route ratio 1.03. For $1 \le \chi < 2$, median field error rises to 0.053, median readout-angle error to $1.88°$, and median route ratio to 1.14; two raw rollouts fail at $\alpha = 0.997$, $\eta = 0.20$, with $\chi \approx 1.7$. For $2 \le \chi < 8$, median field error is 0.088, median readout-angle error is $3.88°$, and median route ratio is 1.24. These numbers show that the theory gives more than a qualitative robustness claim: it identifies the scale on which deviations should be compared across different α and η values.

Panel F gives the biological interpretation of the same scale. The green band marks the low-χ region used as the biological safe zone, motivated by small measured directional imbalance in BTSP induction and BTSP-like recurrent plasticity (Bittner et al., 2017; Li et al., 2024). In this region, raw asymmetric BTSP produces fields and rollouts close to the reversible reference. Outside the band,

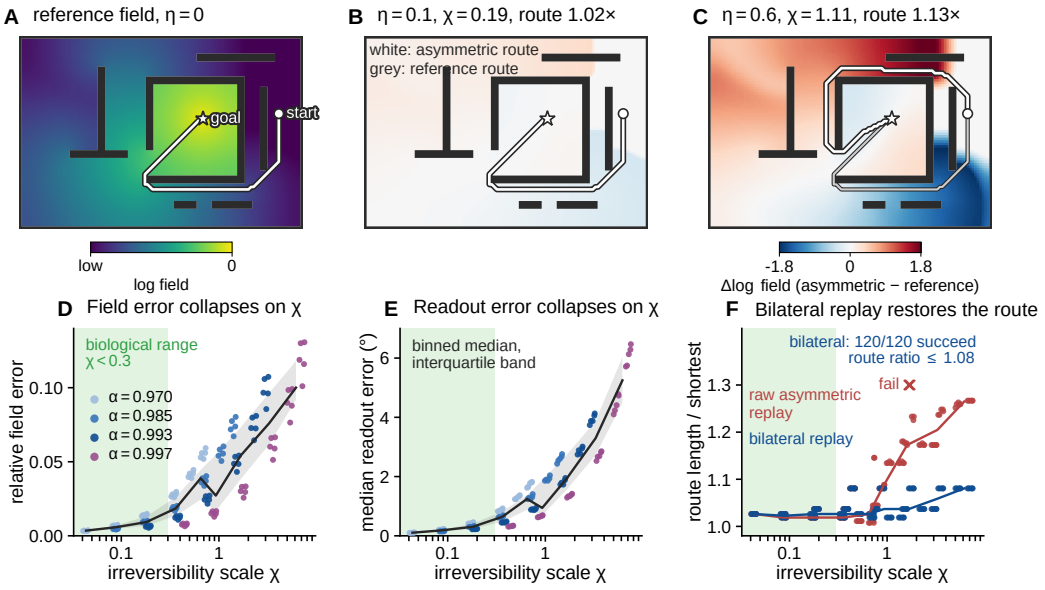


Figure 6: **Asymmetric BTSP deviations are controlled by a single irreversibility scale.** The perturbation knob η scales the directional residue of the effective BTSP memory; the relevant error scale is $\chi = \alpha\|\boldsymbol{E}_{\mathrm{asym}}\|_2/(1-\alpha)$, which combines the raw asymmetry with the recurrent amplification. (**A**) Reference field $\log \boldsymbol{z}_0^g$ in the bug-trap maze at $\alpha = 0.985$ with the reference route. (**B**, **C**) Signed deviations $\Delta \log \boldsymbol{z}_\eta^g = \log \boldsymbol{z}_\eta^g - \log \boldsymbol{z}_0^g$ for $\eta = 0.1$ ($\chi = 0.19$) and $\eta = 0.6$ ($\chi = 1.11$); red means the asymmetric field is higher than the reference. The white route is the population-vector rollout on the asymmetric field, the grey route the reference; low χ leaves the route unchanged (ratio 1.02), larger χ produces a visible detour (ratio 1.13). (**D**) Relative field error collapses when plotted against χ across seeds, discounts (color), and η values; the black curve is the binned median and the grey band the interquartile range. The green band marks $\chi < 0.3$. (**E**) The median readout-angle error follows the same χ-ordered trend. (**F**) Raw asymmetric replay lengthens routes at large χ and fails twice (crosses); bilateral replay projects the same experience back to the reversible support and keeps every route within $1.08\times$ the shortest.

raw replay degrades and can fail. The bilateral projection in Eq. (155) keeps every rollout successful across the full sweep, with route ratios between 1.02 and 1.08. Thus the same plot carries two conclusions. Moderate asymmetry leaves the theorem's prediction effectively intact, and bidirectional replay provides a concrete circuit-level recovery mechanism when raw directional residue grows large.

M Additional experiments: scope, robustness, and comparisons

This section reports a suite of additional experiments (Exp. 01–35) that test the assumptions of Table 1, measure the realized relaxation dynamics, and compare the circuit with learned goal-conditioned models at matched experience; they supply the numbers quoted in Sec. 4.3 and Fig. 4. All experiments run from the released code with fixed seeds and single-threaded numerical libraries (App. N); every condition, including failures, is retained in the result files. Unless stated otherwise, "field error" is the relative L_2 distance between the recurrent field and the direct LMDP solution in the same coordinate, "policy TV" is the total-variation distance between the branch-current policy of Theorem 1 and the optimal policy μ^*, and "stochastic arrival" counts rollouts that sample the policy and reach the exact goal within a 4,096-step horizon. Table 2 lists the headline result of each experiment; the subsections give protocols and details.

Exp.	Question	Headline result
01	realized relaxation dynamics	slow-rate relative error $\leq 1.8 \times 10^{-10}$; monotone energy in every ODE condition
02	exact policy and readout	field error $\leq 3.8 \times 10^{-16}$; branch-current policy TV $\leq 2.8 \times 10^{-16}$
03	directed support	field error $\leq 6.8 \times 10^{-16}$; policy TV $\leq 2.8 \times 10^{-16}$ (18 conditions)
04	overlapping population code	full-rank normalized-field error $\leq 1.4 \times 10^{-14}$
05	local opening and closure events	opening learned to zero field error after 40 balanced events; blocked-transition gate reroutes $8/8$
06	state-dependent costs	field/policy errors $\leq 3.8 \times 10^{-16}/5.2 \times 10^{-16}$
07	edit algebra and storage	$\mathrm{rank}(\Delta \boldsymbol{R}) = \mathrm{rank}(\Delta \boldsymbol{S})$ in every edit; Woodbury error $\leq 2.6 \times 10^{-14}$
08	asymmetry coordinate audit	closed-form and measured antisymmetric coordinates agree to 4.4×10^{-15}
09	transition-support errors	one spurious wall edge gives mean safe arrival 0.86; a count-matched missing edge keeps 1.00
10	structured perturbations	first-order prediction Spearman $\rho = 0.9998$ over 360 conditions
11	strong-recall shells	greedy arrival $103{,}190/103{,}190$; shell $\leq 5.2\%$ of states
12	horizon–latency law	horizon versus 5% latency Spearman $\rho = 0.990$
13	amortized caches versus exact query	exact recurrent query $768/768$ held-out rollouts before and after an edit; frozen spectral caches $864/3{,}072$
14	irregular place-center graphs	$702/702$ collision-free routes; field/policy errors $\leq 1.3 \times 10^{-15}/4.0 \times 10^{-16}$
15	matched-experience representations	sparse operator and SR/SF fields differ by $\leq 2.7 \times 10^{-15}$; all models $4{,}096/4{,}096$ at the highest budget
16	place-cell dropout	$1{,}080/1{,}080$ routes through 80% dropout, uniform and clustered
17	gain heterogeneity and activity noise	gain success $2{,}185{,}200/2{,}185{,}200$; noise success $2{,}884{,}382/2{,}913{,}600$ state starts
18	anisotropic population kernels	moment matching improves $24/24$ condition medians; routes $26/26$
19	directed acquisition	direction-selective rule: zero illegal entries, minimum safe arrival 1.000
20	continuum refinement	monotone field convergence $4/4$; routes $12/12$
21	anytime action and warm starts	median current-action speedup $82\times$ ($\alpha \geq 0.985$); pre-edit initial error $\leq 1.3 \times 10^{-4}$
22	learned architecture family	tabular KL-Q policy error $\leq 1.4 \times 10^{-15}$; width-1,024 goal-coordinate networks $38{,}760/49{,}152$ held-out arrivals
23	experience-driven learners	recurrent operator $40{,}960/40{,}960$ arrivals across all budgets; sampled learners 39–40% on the bug trap at 10^6 transitions
24	fixed-support dropout	field error $\leq 1.4 \times 10^{-14}$; no new survivor edges
25	gain boundary factorial	$13{,}930{,}648/13{,}930{,}650$ state starts arrive through 30% heterogeneity
26	natural directed learning	highest-exposure safe routes $160/160$
27	passive acquisition of an opening	$30/30$ routes, zero field and policy discrepancy at 320,000 transitions
28	amplitude–noise range laws	range-coordinate median Spearman $\rho = 0.933$
29	kernel validity surface	moment correction improves $414/512$ condition–goal cells
30	ϵ scaling and large maps	identity through $38{,}556$ states (1.2×10^{-14}); routes $3/3$
31	irregular continuum convergence	field routes $2{,}880/2{,}880$; recovered screening error $\leq 1.8 \times 10^{-11}$
32	goal-conditioned actor–critic	10^6-transition held-out arrival $25{,}450/30{,}720$; zero illegal execution mass
33	state-matched kernel policy	policy TV and KL exactly 0; vector routes $8/8$
34	sparse acquisition after dropout	native routes $475/480$ through 30% and $689/720$ through 50% dropout
35	time-varying stochastic dynamics	fixed-CV arrival $600/600$ at the noise ceiling; fixed-Fano range law $\rho = 0.966$–0.979

Table 2: Headline results of the additional experiments of App. M; experiment numbers follow the released code.

M.1 Realized relaxation dynamics, latency, and anytime readiness

Exp. 01 integrates the continuous-time dynamics of Eq. (2) instead of solving its fixed point. On the $1{,}472$-state diagnostic map the measured slowest relaxation rate agrees with the spectral prediction δ_g/τ, where δ_g is the spectral gap of $\boldsymbol{I} - \alpha \boldsymbol{S}_{\mathcal{I}_g \mathcal{I}_g}$, to relative error 1.8×10^{-10}, and the energy $\frac{1}{2}\boldsymbol{u}^\top(\boldsymbol{I} - \alpha \boldsymbol{S})\boldsymbol{u} - \boldsymbol{b}_g^\top \boldsymbol{u}$ decreases monotonically in every recorded condition, as required by the gradient-flow form of the symmetric dynamics. Exp. 12 sweeps $\alpha \in \{0.9, 0.94, 0.97, 0.98, 0.985, 0.99, 0.993\}$ over two maps and three goal strata (28 conditions). The cold-start latency to 5% field convergence spans 17–$304\,\tau$ and follows the discounted horizon $\alpha/(1-\alpha)$ with Spearman $\rho = 0.990$; the identity $T_{\mathrm{slow}}\delta_g/\tau = 1$ holds to 1.1×10^{-16}. Warm starts shorten the transient: reusing the field of a nearby goal gives a median speedup of $1.17\times$ ($1.36\times$ for $\alpha \geq 0.97$), and after a local edit the pre-edit field begins inside the 1% criterion in all seven edit conditions (maximum initial error 0.008).

Exp. 21 extends the sweep to seven α values on both maps (42 conditions) and separates whole-field convergence from the readiness of the action at the current state. The current-state argmax is defined in all 42 conditions; in the 24 conditions with $\alpha \geq 0.985$ the current action is correct before the whole field has converged to 5% in $24/24$ cells, with median speedup $82.4\times$ and minimum $1.99\times$ (Fig. 4D shows the bug trap at $\alpha = 0.993$: $11\,\tau$ versus $331\,\tau$). Measured slow rates agree with the spectral prediction to 1.9×10^{-12}. Under structured operator perturbations every system remains stable, the actual field response follows the inverse spectral gap ($\rho = 0.974$) and the first-order prediction ($\rho = 0.997$), and branch-current policy TV stays at most 0.002 in condition mean and 0.015 statewise. The pre-edit field starts inside the 5% and 1% criteria in all 42 conditions (maximum initial error 1.3×10^{-4}).

M.2 Strong recall: distance-resolved shells and behavior

Exp. 11 extends the recall-amplitude sweep of Sec. 4.1 to $A \in [0.1, 100]$ on both maps with $\phi = \tanh$, evaluating every non-goal state. Greedy rollouts of the nonlinear policy reach the exact goal in $103{,}190/103{,}190$ state starts, the same count as the linear reference, and the complete route-ratio distribution is identical between the two policies in every condition. Tie-aware first-action agreement is $103{,}142/103{,}190$ (minimum condition-level agreement 99.90%). The shell of distorted field occupies at most 5.21% of non-goal states (at $A = 100$: 158 states on the diagnostic map and 225 on the bug trap, maximum geodesic radii 8.7 and 9.7 cells). Outside the shell the maximum mean policy TV is 1.6×10^{-4} and the largest statewise TV is 0.003; the largest single-state TV, 0.76, occurs inside the shell adjacent to the goal. Stochastic rollouts started inside the shell are captured by the goal in $33{,}792/33{,}792$ cases, and the distance-stratified stochastic assay reaches $327{,}680/327{,}680$. Fig. 2F plots the shell fraction and arrival against A.

M.3 Gain heterogeneity, activity noise, and operator perturbations

Exp. 10 perturbs the operator in 18 structured directions (360 conditions, 288 with nonzero amplitude). All perturbed systems remain stable; the first-order resolvent predictor of the field response has Spearman $\rho = 0.9998$ with the realized error, so the relevant diagnostic is the perturbation structure relative to the goal-clamped stability margin. Across 40 $\tanh$ conditions the residual-resolvent identity holds to 3.9×10^{-13}; under the strongest saturation the raw field error reaches 0.69 while the branchwise policy TV has condition mean 0.006 and worst-state value 0.22, and we retain this regime as the boundary of the linear description.

Exp. 17 and Exp. 25 draw heterogeneous unit gains ρ_i from three spatial families at six depths and three α values on both maps, with 30 paired realizations per cell (2,700 conditions). Through 30% gain modulation, $13{,}930{,}648$ of $13{,}930{,}650$ state starts reach the goal (minimum condition success 99.973%), with median field error 0.020 and policy TV 0.006; the first-order prediction of the field response has pooled $\rho = 0.962$. In Exp. 17 the gain assay reaches $2{,}185{,}200/2{,}185{,}200$ starts and the activity-noise assay $2{,}884{,}382/2{,}913{,}600$. Exp. 28 and Exp. 35 vary the goal amplitude over three decades together with readout noise, dynamic noise, synaptic jitter, and fresh-innovation noise in the recurrent dynamics (50 fixed streams). Under 1% multiplicative synaptic jitter the maximum condition-mean policy TV is 0.004 and the minimum all-state success 91.9%; the logarithmic range law that links amplitude, noise, and the distance over which the readout is reliable has median Spearman $\rho = 0.933$ (Exp. 28) and $\rho = 0.966$–0.979 across map–α cells (Exp. 35). With fixed coefficient of variation the policy is amplitude invariant ($600/600$ arrivals at the noise ceiling); with fixed Fano factor, raising the amplitude from 0.3 to 30 reduces the median policy TV from 0.126 to 0.049 and raises median all-state arrival from 45% to 68%, the expected signature of a signal that grows faster than its noise.

M.4 Exact policy readout and anisotropic population kernels

Exp. 02 evaluates the branch-current policy of Theorem 1 on the $3{,}036$-state diagnostic map and the $7{,}285$-state bug trap: the recurrent field matches the direct LMDP solution to 3.8×10^{-16}, the local synaptic readout $S_{ij}\psi_j$ matches μ^* to 2.8×10^{-16} in TV, and the Bellman residual is at most 1.7×10^{-16}. The source-at-current-state query and the goal clamp of Fig. 1D are verified on the same operator. For the population-vector readout of Eq. (7), the raw readout on the bug trap has median angular error 9.3° relative to μ^*; near walls the flat-field kernel drift has median norm 1.5 grid units

over $1{,}932$ states, and subtracting this goal-independent baseline (the centering of App. G.3) reduces the error to median 3.9° and p90 26.8° near walls. Raw and centered population readouts reach the goal with route ratios 1.033 and 1.034; Dijkstra gives 1.000, and the Euclidean-direction baseline terminates in a repeated cycle.

Exp. 18 applies determinant-preserving anisotropic Gaussian kernels with condition numbers $\{1, 2, 5, 10, 50\}$ and orientations $\{0^\circ, 45^\circ, 90^\circ\}$ at every non-goal state of both maps ($134{,}147$ state–condition rows). The goal-independent moment correction $\boldsymbol{\Sigma}_{\mathrm{passive}}\boldsymbol{\Sigma}^{+}_{\mathrm{kernel}}\Delta\boldsymbol{x}_{\mathrm{centered}}$, which uses only the kernel and passive-transition covariances, improves the condition-level median angle in all $24/24$ anisotropic cells, from a raw range of 0.4°–55.4° to 0.7°–15.3°; raw readout reaches the goal in $23/26$ conditions, the moment-matched readout in $26/26$ with median route ratio 1.034. Exp. 29 maps the validity surface over 64 width–condition cells and four goals per map and finds the correction helpful in $414/512$ regular cells. Exp. 33 uses the population kernel whose second moment matches the passive successor moments at every state: its policy agrees with μ^* with TV and KL exactly zero and reaches the goal in $8/8$ vector routes. Locality, centering, accessible support, and moment matching are therefore the conditions under which Eq. (7) reproduces the optimal policy.

M.5 Directed transitions

Exp. 03 builds a 25×17 map (425 states) with a wall carrying two one-way gates, one passable only left to right and one only right to left. For three positive diagonals $\boldsymbol{D}$ (identity, an exponential axis gradient, and a spatial wave), two goals, and $\alpha \in \{0.85, 0.93, 0.97\}$ (18 conditions), the recurrent realization $\boldsymbol{S} = \boldsymbol{D}\boldsymbol{P}_0\boldsymbol{D}^{-1}$ of Corollary 1 reproduces the directed LMDP field to relative error 6.8×10^{-16} and the branch-current policy to 2.8×10^{-16} in TV, with Bellman residual 2.2×10^{-16}; integrating the ODE from zero activity reaches the same field (final error 8.4×10^{-5} at the recorded time) with no monotonicity violation. The symmetric projection of the same map assigns 0.52–0.71 of its probability mass to an illegal reverse crossing of a gate (Fig. 4A).

Exp. 19 and Exp. 26 ask whether the directed operator is acquired from experience. A direction-selective rule that writes the synapse $j \to i$ for each executed transition $i \to j$ (App. D.4) produces zero illegal reverse entries in every condition; from one to 32 events per entry the median field error falls from 0.031 to 0.006 and the median policy TV from 0.026 to 0.005, with minimum safe-arrival probability 1.000. Natural random walks on two directed environments reach full legal-edge coverage at the highest exposure, with median field error 0.025 and policy TV 0.041; routes are safe in $160/160$ conditions and arrive in $158/160$. A control that mixes in reverse-oriented writing creates up to 62 illegal entries and 167 unsafe routes, so the orientation of the plasticity rule, not the exposure, decides whether one-way structure is respected.

M.6 Overlapping population codes

Exp. 04 replaces the one-unit-per-location coordinate by two nonnegative full-rank overlapping encoders (minimum rank 140) and evaluates three goals each. In the recoded population the normalized desirability field is reproduced to 1.4×10^{-14}, the source response to 1.2×10^{-14}, and the policy to 7.3×10^{-14} in TV, and the edit-rank inequality of App. K holds in every condition: the computation is invariant under full-rank recoding, as stated in Sec. 2.5. Undercomplete codes are a different matter. Over a factorial of seven feature counts, four widths, three α values, three goal strata, and ten paired layouts (2,520 conditions), only 850 decoded fields remain positive everywhere, and among those the oracle-assisted route reaches the goal in $444/850$; exact transformed operators are dense with at least 47% negative weights. The invariance therefore holds for full-rank codes; compression below full rank requires additional structure.

M.7 Local opening and closure events and transition-support errors

Exp. 05 streams local events through a $1{,}472$-state map with a ten-edge doorway interface. A positive transition-event rule that adds 25% of the native weight per observed crossing learns a newly opened doorway to median field error 4.1×10^{-4} after 16 events and to exactly zero after 40 balanced events (four per doorway edge). Exp. 27 confirms the same acquisition from goal-free random walks in the true open map: at $320{,}000$ transitions every one of 30 trajectories covers the door edges, the recurrent field agrees with the learned-belief LMDP field within 5.2×10^{-15}, and field error and policy TV relative to the true open map are exactly zero (Fig. 3F). Closure requires evidence of blocking: without

observed transitions the field is unchanged (error 0.0017 before and after 16 elapsed intervals), and the positive-only rule succeeds in $0/8$ streaming episodes. A blocked-transition gate that masks the annotated interface reroutes $8/8$ episodes, and a signed depotentiation extension, motivated by the bidirectionality of BTSP (Milstein et al., 2021), also reaches $8/8$ but needs a median of 40 blocked attempts and 540 decision steps with median route ratio 11.6; the gate is the fast mechanism, and signed LTD a slow proof of principle.

Exp. 07 audits the low-rank algebra. The shared propagation correction obeys $\mathrm{rank}(\Delta \boldsymbol{R}) = \mathrm{rank}(\Delta \boldsymbol{S})$ in every edit and the Woodbury identity of App. K.3 to 2.6×10^{-14}; rescaling each clamp column changes the policy by at most 7.0×10^{-17}. The raw unit-clamped all-goal field change can be numerically full rank because every goal column is renormalized, while its 99%-energy rank is 4–7; "low rank" therefore refers to the exact propagation correction, and the 99.5% of Fig. 3E is the empirical effective compression. Exp. 09 compares false-positive and false-negative edges around obstacles on one fixed map: a single spurious wall edge carries mean maximum illegal action mass 0.33 and reduces the mean safe-arrival probability to 0.86 (range 0.79–0.91 over 20 anchors), whereas a count-matched missing legal edge keeps it at 1.00. Precision of the transition gate matters more than recall in this setting, which is the regime in which BTSP operates (precision 1.00 in Fig. 3A).

M.8 State-dependent costs

Exp. 06 adds a state-dependent cost $q(i)$ to the LMDP objective, realized as a state-dependent gain in the recurrent operator, for one 2,320-state map, three goals, and five hazard strengths. The recurrent field matches the normative solution to 3.8×10^{-16}, the policy to 5.2×10^{-16}, and the Bellman equation to 3.3×10^{-16}; within each cost strength the operator is identical across goals, and only the clamped unit changes. At the strongest hazard the expected exposure falls to 1.6%, 0.9%, and 1.6% of its constant-cost value for the near, mid, and far goals while the expected step counts change by factors 1.31, 1.13, and 0.99. Terrain, effort, and risk costs therefore enter the same linear family, as anticipated in Sec. 5.

M.9 Irregular place-center graphs and the continuum limit

Exp. 14 replaces the grid by irregular graphs of place centers in continuous open and barrier arenas, with edges between centers whose connecting segment does not cross a barrier. Across the arena specifications, $702/702$ population-vector routes reach the goal without touching a barrier, and field and policy errors relative to the LMDP on the same graph are at most 1.3×10^{-15} and 4.0×10^{-16} (Fig. 4B shows a 1,024-center barrier arena). Exp. 20 refines regular grids toward a fixed screened-Poisson continuum problem: the field error decreases monotonically in $4/4$ curves with empirical convergence order 1.64–1.69, the coarsest-to-finest error falls by at least $9.7\times$, and all $12/12$ routes arrive with maximum route ratio 1.08. Exp. 31 repeats the refinement on irregular layouts (20 seeds, six goals, three sizes, two layouts, two arenas, two kernel widths): field routes arrive in $2{,}880/2{,}880$ conditions and moment-readout routes in $2{,}860/2{,}880$, the recovered screening constant differs from κ by at most 1.8×10^{-11}, and field and readout converge with median orders 1.45 and 1.63.

M.10 Partial coverage by place cells

Exp. 16 removes place cells from both maps, uniformly or in spatial clusters, at nested fractions up to 80%, and reconstructs the survivor graph from the geometry of the remaining centers. All $1{,}080/1{,}080$ routes reach the exact goal (Fig. 4C). Exp. 24 keeps the original support and only deletes units, redirecting the deleted mass to self-loops or renormalizing survivor rows: the field identity holds to 1.4×10^{-14} with no new survivor edges, and connected routes are retained through median prefixes of 30–40% loss, after which the fixed support disconnects the protected assemblies; geometry-aware reconstruction is what carries Exp. 16 through 80%. Exp. 34 writes the sparse operator from scratch by goal-free BTSP after dropout: native learned-edge coverage has median 1.000 and minimum 0.992, the operator error has median 0.019, and native routes arrive in $360/360$ conditions through 20% dropout, $475/480$ through 30%, and $689/720$ through 50% (884/1,080 over the full 0–80% grid). A degree-stratified relabeling control yields $0/1{,}080$ arrivals, so the result reflects the learned geometry and not the degree distribution.

M.11 Comparisons with cached and learned goal-conditioned models

Four families were compared with the recurrent operator at matched experience. (i) *Oracles.* The direct LMDP solution is the ground truth of Theorem 1 and is matched to machine precision (Exp. 02). A Dijkstra oracle gives route ratio 1.000 on the bug trap; the population readout gives 1.033; the Euclidean-direction baseline terminates in a repeated cycle. Dijkstra is a geometric shortest-path oracle, not the finite-temperature KL objective. (ii) *Successor representations.* Exp. 15 builds successor features from the same transition streams: the sparse operator and the full-state SR/SF fields differ by at most 2.7×10^{-15}, both reach $4{,}096/4{,}096$ stochastic arrivals at the highest budget with median policy TV 0.0019, and one shared closure event raises both from $831/2{,}048$ to $2{,}048/2{,}048$ on the edited map. The difference is storage: the operator keeps $\mathcal{O}(N)$ synapses and answers any goal by one relaxation, while a dense all-goal cache grows as N^2 ($252\times$ more float64 storage at $3{,}036$ states, $3{,}116\times$ at the largest map of Exp. 30). (iii) *Frozen caches and supervised readouts.* Exp. 13 fixes a goal split (32 training, 8 validation, 12 held-out goals) on the $3{,}036$-state map and registers an edit closing a near gate (32 weight entries). The exact recurrent query reaches the goal in $768/768$ paired held-out rollouts before and after the edit; validation-selected RBF readouts trained on 8/16/32 goal fields reach $1{,}357/2{,}304$ overall and improve smoothly with goal supervision; the four frozen spectral caches of the original map reach $864/3{,}072$ (28%). (iv) *Learned goal-conditioned models at absolute budgets.* Exp. 22 and Exp. 23 train goal-specific sampled soft-Q, held-out goal-coordinate TD value learners, tabular KL-Q, and shared goal-coordinate value networks (widths 256 and 1,024) on twenty streams per map with budgets from 10^3 to 10^6 transitions, evaluated by $4{,}096$-step stochastic rollouts against a same-seed exact-reference ceiling that reaches 100%. The recurrent operator reaches $40{,}960/40{,}960$ original-map arrivals across all budgets (median policy TV $0.109 \to 0.001$ from 10^3 to 10^6). At 10^6 transitions on the bug trap the tabular soft-Q learner reaches $2{,}019/5{,}120$ (39.4%) and goal-coordinate TD learners with 4/16/64 training goals $40.3/40.2/40.0\%$, while on the easier two-door map all learners exceed 98%. Tabular KL-Q, which fits the same linear family, matches the direct policy to 1.4×10^{-15} and reaches $16{,}384/16{,}384$; width-1,024 goal-coordinate networks reach $38{,}760/49{,}152$ held-out arrivals. Exp. 32 trains a sampled-count goal-conditioned soft actor–critic with a pooled reversible local operator: at 10^6 transitions on the bug trap it reaches 48.0%, 53.9%, and 96.2% of held-out-goal arrivals with 4, 16, and 64 training goals (99–100% on the two-door map), with zero illegal execution mass; after 10 observed closure events on the two-door map its arrival falls to 41–52%. The learned models approach the operator only with tens of training goals and lose the plan after a local edit, which is the per-goal supervision and the relearning that the zero-shot readout does not need.

M.12 Scale and the lazy rate

Exp. 30 verifies the fixed-ϵ construction of Eq. (9): across 40 ϵ cells on regular graphs the field error is at most 6.6×10^{-15} and the moving-policy TV at most 1.2×10^{-15}, so any admissible ϵ gives the same plan once α is matched. On three maps beyond the previous $7{,}285$-state endpoint, up to $38{,}556$ states, the recurrent and LMDP fields agree within 1.2×10^{-14} with policy TV 9.7×10^{-16}, all $3/3$ routes arrive, and the dense-to-sparse storage ratio reaches $3{,}116\times$. Exp. 07 fits the sparse storage over four regular-grid sizes as $N^{1.03}$ (96 bytes per state at $3{,}036$ states); bounded degree supplies the structural linear law (Fig. 4F). The remaining finite cost is latency: on the $7{,}285$-state cold bug trap the field error reaches 8×10^{-4} by the last sampled time while the policy is available on 94% of states with mean TV 0.027, and the current-action readiness of App. M.1 is what makes the plan usable before the field has converged everywhere.

N Reproduction

The code is released under the MIT license and has three parts: the core experiments of Sec. 4 and App. L (`experiments/`, with result tables in `results/`), the additional experiments of App. M (`extended/`), and the scripts that draw the figures of this paper (`paper_figures/`). Everything runs with Python 3.10 or later and

```
numpy>=1.24
scipy>=1.10
pandas>=2.0
matplotlib>=3.7
```

plus `pytest` for the unit tests. NumPy, SciPy, and pandas are distributed under the BSD 3-Clause license, Matplotlib under the PSF-based Matplotlib license, and pytest under the MIT license; no other third-party code or data is used. The scripts set `OPENBLAS_NUM_THREADS`, `OMP_NUM_THREADS`, and `MKL_NUM_THREADS` to one unless these variables are already set. Randomness enters through NumPy's `default_rng`; the seeds listed in this appendix fully determine the simulations.

From the repository root, the core experiments and their tables are reproduced by

```
python -m experiments.exp1_equivalence --outdir results --figdir figures --seed 2
python -m experiments.exp2_btsp_navigation_lowrank --outdir results --figdir figures --seed 4
python -m experiments.exp3_green_limit --outdir results --figdir figures --seed 5
python -m experiments.exp4_asymmetric_btsp --outdir results --figdir figures --seed 4
```

which write

```
results/exp1_equivalence_metrics.csv          results/exp2_lowrank_remap_metrics.csv
results/exp1_nonlinear_shell_metrics.csv      results/exp3_green_limit_metrics.csv
results/exp1_alpha_recall_sweep_metrics.csv   results/exp4_asymmetric_btsp_metrics.csv
results/exp2_btsp_learning_metrics.csv        results/exp4_asymmetric_btsp_summary.csv
results/exp2_challenge_navigation_metrics.csv
```

The additional experiments are run by

```
python -m extended.run_all
```

which executes the unit tests, runs Exp. 01–35 with fixed seeds and single-threaded numerical libraries, reruns the four core experiments in an isolated directory and compares their result tables with the stored ones, and records the environment, the command lines, and a SHA-256 hash of every result file. A single experiment is run by its module name, for example

```
python -m extended.experiments.exp21_anytime_horizon_robustness
```

Each experiment writes raw CSV files and a JSON summary to `extended/results/`; the numbers quoted in App. M are read from these files, and failed conditions are kept in the same files as successful ones.

Each figure is drawn by its own script in `paper_figures/` (Fig. 3, for example, by `fig3_btsp.py`), which recomputes the displayed fields, rollouts, and spectra from the experiment code and reads the sweep statistics from the result files above.

The stored tables were produced on macOS (arm64). Values reported at machine precision (10^{-16} to 10^{-13}) vary at that level between platforms, and the pseudoinverse shape errors of App. L.4 vary at the 10^{-8} level. The two-mode energy fraction of Fig. 3E depends on how ties are broken when the 34 goals are selected by spatial quantiles, because NumPy's default sort is not stable and its vectorized x86 implementation orders ties differently; across tie-breaking orders the first two modes carry 98.3–99.5% of the energy (99.46% with the reference order used in Fig. 3E). The README of the code release gives the setting that restores the reference order on x86 processors.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Sec. 1 state the specific contributions, and each is supported by the equivalence theorem and its directed corollary (Sec. 2.2), the readout analysis (Sec. 2.3), the BTSP learning model and low-rank update analysis (Sec. 3), and the experiments of Sec. 4. The claims are scoped to the assumptions listed in Table 1 and to the simulated environments.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Sec. 5 contains a dedicated "Limitations" paragraph (unspecified hippocampal subfield, idealized transition gate, the population-vector readout as an approximation of the exact branch-current policy), Table 1 lists every assumption with where it is relaxed or tested, and Apps. C, F, I, and M state the perturbative regimes and the failure modes retained in the result files.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best

judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Theorem 1 states its assumptions, including the linear-regime assumption, in the theorem; the proof is in App. D, the proof of Corollary 1 in App. D.4, and the proof of Proposition 1 in App. G.3. The supporting derivations for linearization, reversibility, readout, nonlinear perturbation, Green-function limits, and low-rank updates are in Apps. C–K.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: App. L specifies graph construction, maze geometries, solvers, tolerances, seeds, sweep values, metrics, and rollout rules; App. M gives the protocols of the additional experiments, and App. N the exact reproduction commands. The supplemental material contains the code, result files, and figure scripts.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The supplemental material includes the full experiment code with README files and requirements, a one-command runner for the additional experiments, result CSV and JSON files, and the figure scripts; App. N gives the exact commands.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: App. L gives maze construction, connectivity rules, kernel widths, discounts, recall amplitudes, seeds, solver tolerances, the exploration sampler, rollout scoring, and metrics; App. M gives the goal splits, budgets, horizons, and model configurations of the learned comparators.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Stochastic BTSP experiments report means and SEM across five exploration seeds (Sec. 4.2, App. L.3); the additional experiments report complete counts over all conditions, medians with interquartile ranges, and Wilson bounds where applicable (App. M). Deterministic algebraic checks report relative errors and tolerances instead of error bars.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: All experiments run on CPU on a single Apple-silicon laptop, without GPUs; the supplemental compute report lists the wall-clock time of each core experiment, and App. M lists the size of every sweep in the additional experiments.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: The work is theoretical and simulation-based on synthetic mazes, involves no human subjects, private data, or deployment, and conforms with the NeurIPS Code of Ethics.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: Sec. 5 discusses scientific implications and testable predictions. The work is foundational neuroscience theory with synthetic simulations and no direct deployment path, so a separate discussion of positive and negative societal impacts is not included.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper releases simulation code, synthetic maze procedures, result files, and figures; no pretrained models or scraped data with a risk of misuse are released.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: No external datasets, pretrained models, or third-party research code are used. The open-source packages used by the code (NumPy, SciPy, and pandas under the BSD 3-Clause license, Matplotlib under the PSF-based Matplotlib license, and pytest under the MIT license) are named with their licenses in App. N.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The released code (core experiments, additional experiments, and figure scripts) is documented by README files in the code release and by Apps. L–N, and is distributed under the MIT license.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research with human subjects; all results come from simulations.

    Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve human participants or identifiable private data, so IRB review is not applicable.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not an important, original, or non-standard component of the core scientific method, theory, simulations, or evaluation pipeline. Any use limited to writing, editing, or formatting would not affect the methodology, scientific rigor, or originality and therefore does not require declaration under this checklist item.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.