A LANDSCAPE THEORY OF ASSOCIATION: PATHS AND CHOICES IN REPRESENTATION SPACE

Anonymous authors
Paper under double-blind review

ABSTRACT

Association enables a cue to activate related representations, supporting memory retrieval and flexible reasoning. Understanding this process requires explaining both how an association unfolds and why one destination is selected over alternatives. However, descriptions based solely on representational similarity or energy barriers do not jointly explain these two aspects. We develop a landscape model in which modes of a representation density define associative states and calibrated gradient diffusion connects their basins. Our model separates three quantities: fixed-duration path likelihood, governed by the full Onsager–Machlup action; the weak-noise scale of direct escape, determined by exit barriers; and next-state choice, obtained by normalizing competing exit rates. By doing so, the framework links the representation landscape to associative paths and destination choices, revealing how local curvature shapes probable routes and how competition among accessible destinations governs selection. Across three controlled landscapes, free-time escape-rate slopes recover both barrier predictions within 5.45%; in an equal-barrier design, an action-independent moving-tube PDE recovers the complete ordering of six prespecified paths predicted by the full Onsager–Machlup action, whereas the quadratic-only truncation selects a different winner. On a five-well landscape, $160{,}000$ first-exit trajectories recover all four barrier-difference log-odds slopes within 4.5%; a fixed-center energy intervention raises the targeted exit probability by 21.8 percentage points on average and reverses the dominant destination. Together, these results quantitatively connect landscape geometry to associative paths, escape scales, and next-state choice.

1 INTRODUCTION

An association cue rarely has a single inevitable response. It can bring several memories or concepts to mind, and the tendency to move from one concept to another need not be symmetric. Free association norms depict this structure as a distribution of responses conditioned on a cue (Nelson et al., 2004; De Deyne et al., 2019). A process model of association must therefore go beyond identifying similar representations. It must define the states between which association occurs, characterize the paths that connect them, and explain how competing destinations give rise to a particular choice.

Existing methods provide different parts of this account. Semantic networks describe relations between concepts and support spreading activation or random-walk models of retrieval (Collins & Loftus, 1975). Distributed representations provide coordinates in which semantic relationships can be measured (Mikolov et al., 2013; Radford et al., 2021). Associative-memory models identify stored patterns with stable states of an energy function (Hopfield, 1982; Ramsauer et al., 2021). However, these works leave unresolved how to specify a continuous probability law over trajectories between representations: distance alone does not determine the intervening route, nor does attractor depth explain how multiple possible exits compete.

To address this gap, we model association as stochastic motion on a potential landscape derived from a distribution over representations. Given a smooth density p, we define the potential by $\Phi = -D_0 \log p + C$, with a fixed calibration scale $D_0 > 0$. The non-degenerate modes of p then correspond to stable equilibria of the induced gradient flow, while isotropic noise enables transitions between their basins of attraction. The density alone does not specify a unique stochastic process:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

gradient drift and isotropic diffusion are explicit modeling choices. These choices turn a static representation distribution into a dynamical model that yields precise and testable predictions about transition paths, escape behavior, and destination choices.

Interpreting these predictions requires distinguishing three related but mathematically different questions. The likelihood of a trajectory over a prescribed duration is a fixed-time, fixed-noise small-tube problem; escape from a metastable basin is a free-time, weak-noise problem; and selection of the next basin is a competing-exit problem that requires normalization across directly accessible destinations. Because these questions involve different limiting regimes and normalizations, their associated quantities are not interchangeable. In particular, a finite-time path action cannot serve as an unrestricted free-time transition cost, and an escape barrier alone does not define an association probability. Paths, escape rates, and destination choices must therefore be derived from the same landscape while being analyzed as distinct observables.

Our framework organizes the problem in the order in which its objects are constructed:

$$\text{density} \longrightarrow \text{potential and basins} \longrightarrow \begin{cases} \text{fixed-time path probabilities,} \\ \text{direct exit rates and destination choices.} \end{cases}$$

For paths, the Onsager–Machlup (OM) functional includes a divergence correction that measures local contraction of the drift. For escape, the Freidlin–Wentzell (FW) action gives a basin-restricted barrier. For choice, Eyring–Kramers (EK) theory provides local rate corrections, and normalizing the rates over directly accessible destinations yields a conditional distribution.

The contributions of this paper are threefold:

- We introduce a landscape theory of association that maps a representation density to a calibrated potential landscape and an explicitly specified stochastic dynamics. This construction provides a unified account of associative paths, metastable transitions, and next-state choices under transparent modeling assumptions and consistent noise scaling.
- We distinguish three related but mathematically different observables: finite-time path probabilities, free-time metastable escape, and normalized next-state choice. This separation yields two key consequences. First, equal barrier heights need not imply equally likely finite-time paths, because landscape curvature along the paths can alter their probability ordering. Second, leading-order next-state odds are governed by relative direct-exit barriers: the source-well depth cancels when comparing exits from the same basin, while basin adjacency determines which destinations are valid competitors.
- We conduct controlled numerical studies on known landscapes, including a barrier-rate benchmark, an equal-barrier path comparison verified with an independent PDE solver, and a competing-exit intervention. Together, these experiments isolate the relevant mechanisms and test the predicted path, rate, and choice behavior separately.

2 Related Work

Semantic association and associative memory. Free-association norms define a directed conditional distribution over responses (Nelson et al., 2004; De Deyne et al., 2019). Spreading activation, semantic foraging, and network random walks model retrieval over discrete relations (Collins & Loftus, 1975; Hills et al., 2012; Austerweil et al., 2012; Abbott et al., 2015), while distributional representations predict association from corpus geometry (Mikolov et al., 2013; Cattle & Ma, 2017; Rodriguez & Merlo, 2020). These methods provide strong endpoint and graph baselines, but they do not determine a continuous law over intermediate routes. Hopfield and dense associative memories instead identify memories with attractors of an energy function (Hopfield, 1982; Krotov & Hopfield, 2016; Ramsauer et al., 2021); cognitive landscape models use related ideas to quantify switching between fitted neural states (Ueltzhöffer et al., 2015; Ye et al., 2023). Our target is complementary: a representation-to-probability calculus in which local geometry ranks routes, basin topology identifies direct competitors, and relative exit rates normalize their choices.

Learning and using a representation landscape. Energy-based learning, score matching, and noise-contrastive estimation provide ways to estimate unnormalized densities or their gradients (Le-

Cun et al., 2007; Hyvärinen, 2005; Gutmann & Hyvärinen, 2010); score-based generative models then use $\nabla \log p$ to construct stochastic samplers (Song & Ermon, 2019; Song et al., 2021). Given a sufficiently regular density or score, we instead hold the calibrated landscape fixed and ask which attractors, routes, barriers, and exits its induced diffusion predicts. This use requires stable basin topology; it does not assume that a high-dimensional learned representation automatically supplies such topology. It also yields a non-circular test: learn the landscape without association responses and evaluate its predictions on held-out behavior (Jones et al., 2015).

Path probabilities and metastable transitions. The fixed-time layer specializes the classical Onsager–Machlup small-tube functional (Onsager & Machlup, 1953; Cugliandolo & Lecomte, 2017). The free-time weak-noise layer uses the Freidlin–Wentzell quasipotential and minimum-action theory (Freidlin & Wentzell, 1998; E et al., 2004), while competing Eyring–Kramers rates add curvature prefactors and normalization over direct exits (Bovier et al., 2004; Lelièvre et al., 2025). These theories are classical; our contribution is their unified decomposition for association, together with explicit boundaries between them. In particular, a quadratic discretization can omit the divergence correction and change a fixed-time path mode (Dutra et al., 2014), and the full Onsager–Machlup action is not a valid unrestricted free-time cost. Closest to our fixed-time computation, Raja et al. (2025) use scores and the Onsager–Machlup action for molecular transition paths. We connect that route object to two additional observables—metastable transition scale and normalized choice—without treating the three as one generic action.

3 Theory: From Stochastic Dynamics to Association Probabilities

Let $\mathcal{Z} = \mathbb{R}^d$ be a representation space with a specified Euclidean coordinate system. The latent state can be described by the Itô SDE

$$dX_t = -\nabla\Phi(X_t)\,dt + \sqrt{2D}\,dW_t, \qquad X_0 = x_0, \quad D > 0, \tag{1}$$

where $\Phi : \mathbb{R}^d \to \mathbb{R}$ is the cognitive potential, W_t is a standard d-dimensional Wiener Process, and D is a scalar diffusion coefficient controlling the noise strength.

Assumption 1 (Regularity and well-posedness)**.** *Let $\Phi \in C^3(\mathbb{R}^d)$. For every $D > 0$, the diffusion in Eq. (1) is nonexplosive. A sufficient condition is that, for each $D > 0$, there exist a function $\mathcal{V}_D \in C^2(\mathbb{R}^d; [1, \infty))$ with compact sublevel sets and a constant $c_D < \infty$ such that*

$$\mathcal{L}_D\mathcal{V}_D(x) \le c_D\mathcal{V}_D(x), \qquad x \in \mathbb{R}^d, \tag{2}$$

where $\mathcal{L}_D f = -\nabla\Phi \cdot \nabla f + D\Delta f$.

The C^3 regularity gives a locally Lipschitz drift and the derivatives required for Onsager–Machlup analysis; together with nonexplosion, verifiable by the stated Lyapunov criterion, it yields a unique global strong solution.

3.1 Finite-time Path Probabilities and the Onsager–Machlup Action

Within the stochastic dynamics in Eq. (1), we consider a finite-time question: among the candidate routes connecting two associative states a and b over a prescribed duration T, which is the most likely? Since an individual diffusion path has zero probability and path space admits no ordinary pointwise density, route likelihood must instead be compared through probabilities of shrinking tubes around reference paths. Write $\mu = -\nabla\Phi$ and define $\Gamma_{a,b}^T = \{\gamma \in H^1([0,T]; \mathbb{R}^d) : \gamma(0) = a, \gamma(T) = b\}$. For each $\gamma \in \Gamma_{a,b}^T$, let $B_\varepsilon(\gamma) = \{\sup_{t \le T} \|X_t - \gamma(t)\| < \varepsilon\}$ and denote by $\mathbb{P}_a$ the law of Eq. (1) initialized at $X_0 = a$.

Theorem 2 (Onsager–Machlup path comparison)**.** *Fix $T > 0$ and $D > 0$. Under Assumption 1, for every $\gamma, \eta \in \Gamma_{a,b}^T$, the uniform-tube probability ratio satisfies*

$$\lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}_a(B_\varepsilon(\gamma))}{\mathbb{P}_a(B_\varepsilon(\eta))} = \exp\left(-S_{\mathrm{OM},D}^T[\gamma] + S_{\mathrm{OM},D}^T[\eta]\right), \tag{3}$$

where

$$
\begin{aligned}
S^T_{\mathrm{OM},D}[\gamma] &= \int_0^T \left[\frac{1}{4D}\left\|\dot\gamma(t)-\mu(\gamma(t))\right\|^2 + \frac{1}{2}\nabla\cdot\mu(\gamma(t))\right] dt \\
&= \frac{1}{4D}\int_0^T \left\|\dot\gamma(t)+\nabla\Phi(\gamma(t))\right\|^2 dt - \frac{1}{2}\int_0^T \Delta\Phi(\gamma(t))\, dt,
\end{aligned}
\tag{4}
$$

Theorem 2 establishes a direct ordering of candidate routes: a smaller OM action corresponds to a larger asymptotic tube probability. Therefore, whenever the action reaches the minimum, any $\gamma^*_{\mathrm{OM}} \in \arg\min_{\gamma\in\Gamma^T_{a,b}} S^T_{\mathrm{OM},D}[\gamma]$ is a most-probable reference path from a to b in the small-tube sense. In this way, the probabilistic problem of comparing finite-time association routes is reduced to a variational problem over $\Gamma^T_{a,b}$. Eq. (4) further explains how the dynamics rank these routes. The quadratic term penalizes deviations from the deterministic drift, whereas the divergence term captures the local contraction or expansion of nearby stochastic trajectories. For the gradient drift $\mu = -\nabla\Phi$, the latter becomes $-\Delta\Phi/2$, hence holding the drift-mismatch term fixed, regions with positive $\Delta\Phi$ lower the action. Finite-time route preference therefore depends on the geometry encountered along the entire path, rather than only on endpoint distance or barrier height. Appendix A.1 provides the technical small-tube result and verifies the coefficients through Girsanov's formula, while the precise relation to the fixed-endpoint action of Raja et al. (2025) is derived in Appendix A.2.

3.2 Weak-Noise Transition Costs

The OM action in Section 3.1 ranks paths at fixed T and D but is not an unrestricted free-time transition cost: waiting near a non-degenerate minimum before transitioning accumulates a negative divergence contribution, so its free-time infimum is generally $-\infty$ (Appendix A.4). We therefore define the fixed-horizon endpoint cost $C^T_{\mathrm{OM},D}(a,b) := \inf_{\gamma\in\Gamma^T_{a,b}} S^T_{\mathrm{OM},D}[\gamma]$, which retains the finite-noise divergence correction and characterizes the most-probable association paths discussed above.

For the weak-noise statements below, we additionally assume that, for every finite T, the family $\{X^D\}_{D\downarrow 0}$ satisfies a good Freidlin–Wentzell large-deviation principle on $C([0,T];\mathbb{R}^d)$ with speed $(2D)^{-1}$ and rate functional I_T. This condition is separate from Assumption 1: nonexplosion ensures global well-posedness, whereas the large-deviation principle also requires exponential tightness of the small-noise family.

To quantify transition difficulty when the duration is not prescribed, we instead use the Freidlin–Wentzell (FW) action

$$
I_T[\gamma] := \frac{1}{2}\int_0^T \left\|\dot\gamma(t)+\nabla\Phi(\gamma(t))\right\|^2 dt. \tag{5}
$$

For each fixed path and duration,

$$
2D\, S^T_{\mathrm{OM},D}[\gamma] = I_T[\gamma] - D\int_0^T \Delta\Phi(\gamma(t))\, dt. \tag{6}
$$

For fixed γ and T, Eq. (6) makes I_T the leading weak-noise term; it does not justify interchanging $D \to 0$ with unrestricted minimization over T. Appendix A.3 treats the Euler–Maruyama joint-density objective.

Minimizing the FW action over both paths and durations defines the directed transition cost

$$
V(a,b) := \inf_{T>0}\inf_{\gamma\in\Gamma^T_{a,b}} I_T[\gamma]. \tag{7}
$$

With Φ fixed as D varies, V is noise-independent and enters the leading large-deviation exponent as $V/(2D)$.

Proposition 3 (Properties of the unrestricted transition cost)**.** *For any $a, b, c \in \mathbb{R}^d$:*

1. *$V(a,b) \geq 0$ and $V(a,a) = 0$;*
2. *$V(a,c) \leq V(a,b) + V(b,c)$;*
3. *If the deterministic relaxation initialized at a converges to b, then $V(a,b) = 0$.*

Consequently, V is directed and need not be symmetric.

The endpoint cost $V(a, b)$ ranges over paths that may cross intermediate basins. Next-state choice is governed by first exits through the competing boundary channels of the source basin. Let $\mathcal{B}_a$ be the deterministic basin of a local minimum a. For a direct neighbor b, let $\Sigma_{ab} \subset \partial\mathcal{B}_a$ be the boundary exit channel toward $\mathcal{B}_b$. Let $\Gamma^T_{a,\Sigma_{ab}}$ contain the H^1 paths from a to Σ_{ab}. The direct-exit cost is

$$V_{\mathrm{dir}}(a, b) := \inf_{T>0} \inf_{\substack{\gamma\in\Gamma^T_{a,\Sigma_{ab}} \\ \gamma([0,T))\subset\mathcal{B}_a}} I_T[\gamma]. \tag{8}$$

Proposition 4 (Direct-exit barrier)**.** *Let a and b be neighboring non-degenerate minima. Suppose an isolated index-one saddle s_{ab} minimizes Φ on both Σ_{ab} and Σ_{ba} and has descending branches to a and b. Then*

$$V_{\mathrm{dir}}(a, b) = 2\big(\Phi(s_{ab}) - \Phi(a)\big), \qquad V_{\mathrm{dir}}(b, a) = 2\big(\Phi(s_{ab}) - \Phi(b)\big). \tag{9}$$

The shared saddle fixes the crossing height, while the source-well depths make the direct-exit costs directional. Proofs appear in Appendix A.4.

3.3 From Transition Costs to Association Probabilities

Direct-exit costs set the weak-noise scales of competing transition rates. Let $\mathcal{M}$ be a finite collection of non-degenerate local minima of Φ, interpreted as metastable association states. For each $a \in \mathcal{M}$, let $\emptyset \neq \mathcal{C}(a) \subseteq \mathcal{M} \setminus \{a\}$ denote its direct-exit competitor set. Each $b \in \mathcal{C}(a)$ has an exit channel Σ_{ab} with cost $V_{\mathrm{dir}}(a, b)$ defined in Eq. (8).

Under the metastable conditions of Appendix A.5, the aggregate rate through the channels to b takes the form

$$k_D(a, b) = \kappa_{ab}(D) \exp\left(-\frac{V_{\mathrm{dir}}(a, b)}{2D}\right), \qquad b \in \mathcal{C}(a), \tag{10}$$

where $\kappa_{ab}(D) > 0$ is subexponential in $1/D$: $\lim_{D\to 0} 2D \log \kappa_{ab}(D) = 0$.

The barrier sets the exponential scale; the prefactor collects local geometric contributions from the source well and exit saddles (Bovier et al., 2004; Lelièvre et al., 2025).

Normalizing these direct rates gives the next-state association probability

$$P^{\mathrm{assoc}}_D(b \mid a) := \frac{k_D(a, b)}{\sum_{c\in\mathcal{C}(a)} k_D(a, c)}. \tag{11}$$

If $\mathcal{C}(a)$ contains every direct exit, this is the reduced jump process's next-state law; otherwise, it is conditional on exit through $\mathcal{C}(a)$.

Proposition 5 (Normalized weak-noise association law)**.** *Let $\mathcal{C}(a)$ be nonempty and finite, and suppose that its direct rates satisfy Eq. (10). Define $V^{\mathrm{dir}}_{\min}(a) := \min_{c\in\mathcal{C}(a)} V_{\mathrm{dir}}(a, c)$. Then*

$$P^{\mathrm{assoc}}_D(b \mid a) = \frac{\kappa_{ab}(D) \exp[-V_{\mathrm{dir}}(a, b)/(2D)]}{\sum_{c\in\mathcal{C}(a)} \kappa_{ac}(D) \exp[-V_{\mathrm{dir}}(a, c)/(2D)]}, \tag{12}$$

and

$$\lim_{D\to 0} \left[-2D \log P^{\mathrm{assoc}}_D(b \mid a)\right] = V_{\mathrm{dir}}(a, b) - V^{\mathrm{dir}}_{\min}(a). \tag{13}$$

Each barrier sets a direct rate's exponential scale; its excess over the lowest competing barrier sets the logarithmic scale of next-state choice. The corresponding prefactor-free choice rule is

$$\widehat{P}_D(b \mid a) = \frac{\exp[-V_{\mathrm{dir}}(a, b)/(2D)]}{\sum_{c\in\mathcal{C}(a)} \exp[-V_{\mathrm{dir}}(a, c)/(2D)]}. \tag{14}$$

With the potential held fixed, D controls how sharply this distribution concentrates on low-barrier exits. Curvature prefactors refine its finite-noise calibration, as tested in Section 4.3.

3.4 Learning the Potential and Identifying Association Attractors

We now derive Φ from a representation density and identify its non-degenerate modes as stable, basin-level association states. Let $p \in C^3(\mathbb{R}^d)$ be a strictly positive probability density with respect to Lebesgue measure in the Euclidean coordinates fixed in Section 3, and let $D_0 > 0$ be a calibration scale.

Proposition 6 (Density–potential–score correspondence)**.** *Define*

$$\Phi_p(z) = -D_0 \log p(z) + C, \tag{15}$$

and suppose that Φ_p satisfies Assumption 1. At noise level $D = D_0$, the diffusion in Eq. (1) is reversible with invariant density p, and its drift is $-\nabla\Phi_p(z) = D_0 \nabla \log p(z)$. Conversely, Eq. (15) is, up to the constant C, the unique gradient potential whose diffusion at scale D_0 is reversible with invariant density p.

If Φ_p is held fixed and $Z_p(D) := \int_{\mathbb{R}^d} p(z)^{D_0/D}\, dz < \infty$, then the reversible invariant density at runtime noise level D is $\pi_D(z) = Z_p(D)^{-1} p(z)^{D_0/D}$.

This construction separates landscape geometry from dynamical scale. The density p determines the dimensionless landscape $\Phi_p/D_0 = -\log p + \text{constant}$, where D_0 sets the drift and barrier scales. Varying D_0 rescales the drift speed and barrier heights without changing critical points, gradient-flow orbits, or basins; at fixed D_0, the ratio D_0/D controls stationary tempering.

A learned scalar energy model instantiates the construction directly:

$$\widehat{p}_\psi(z) \propto e^{-E_\psi(z)}, \qquad \Phi_\psi(z) = D_0 E_\psi(z), \qquad \mu_\psi(z) = -D_0 \nabla E_\psi(z) = D_0 s_\psi(z),$$

where $s_\psi = -\nabla E_\psi$ approximates $\nabla \log p$. Noise-contrastive estimation and score matching provide standard ways to learn E_ψ or its gradient (Gutmann & Hyvärinen, 2010; Hyvärinen, 2005). The scalar-energy parameterization ensures that the learned drift is conservative; a directly parameterized score field must satisfy the same condition for the potential and barrier theory to apply.

Proposition 7 (Representation modes are association attractors)**.** *Under the conditions of Proposition 6, a point m is a non-degenerate local maximum of p if and only if it is a non-degenerate local minimum of Φ_p. At such a point, $\nabla^2\Phi_p(m) = -(D_0/p(m))\nabla^2 p(m) \succ 0$; hence m is a locally exponentially stable equilibrium of $\dot{z} = -\nabla\Phi_p(z) = D_0 \nabla \log p(z)$.*

We therefore define the population-level association attractors as

$$\mathcal{M}_p := \left\{ m \in \mathbb{R}^d : \nabla p(m) = 0,\ \nabla^2 p(m) \prec 0 \right\}. \tag{16}$$

If s is a relevant saddle on a direct exit from the basin of m, then

$$\Phi_p(s) - \Phi_p(m) = D_0 \log \frac{p(m)}{p(s)}. \tag{17}$$

Equation (17) provides a direct link between representation geometry and transition difficulty: density modes define the states, gradient-flow basins organize them, and mode-to-saddle density drops determine their weak-noise barriers.

For the transition theory, we select a finite subset $\mathcal{A}_p \subset \mathcal{M}_p$ whose relevant basins are separated by positive barriers through non-degenerate index-one saddles. Their saddle-mediated adjacency defines the direct competitor set $\mathcal{C}(a)$. On this state space, the OM functional ranks fixed-duration paths, the FW action determines direct-exit costs, and normalized Eyring–Kramers rates determine next-state choice over $b \in \mathcal{C}(a)$. The learned density thereby supplies a unified phase portrait, barrier hierarchy, and probabilistic choice law for association.

4 Experiments

The theory in Section 3 yields three observable predictions: barriers set free-time escape scales, the full Onsager–Machlup action orders fixed-time path-tube probabilities, and direct-exit rates determine normalized next-state choice over saddle-defined neighbors. We test each mechanism on a controlled landscape designed to discriminate it. The Müller–Brown benchmark probes barrier-rate scaling across geometrically distinct exits. An equal-barrier two-channel landscape puts the

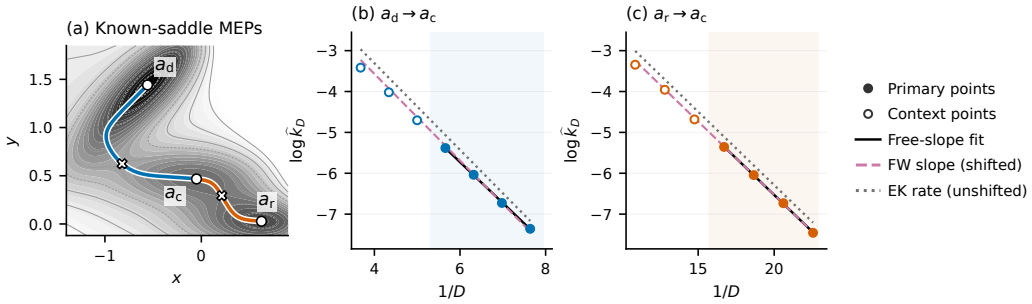

Figure 1: **Free-time escape geometry and barrier scaling.** (a) Scaled Müller–Brown potential with minima (circles), relevant index-one saddles (crosses), and reference minimum-energy paths for the deep-to-central and right-to-central exits. (b,c) Estimated natural-log escape rates $\log \widehat{k}_D$ against $1/D$. Filled points enter the prespecified primary fit; open points show higher-noise behavior. Black lines show unconstrained-slope fits; dashed lines fix the slope at the Freidlin–Wentzell prediction $-\Delta\Phi$ and fit only the vertical offset; dotted curves show the unshifted Eyring–Kramers predictions for absolute rates. Across barriers differing by nearly a factor of three, the fitted slopes recover $-\Delta\Phi$ to within 5.45% and 0.14%, respectively.

full action in conflict with its quadratic-only truncation, with an independent moving-tube PDE adjudicating their path-probability rankings. A five-well landscape tests barrier-difference log-odds, curvature-prefactor calibration, and a fixed-center intervention, using a nonadjacent state as a topological negative control.

4.1 Barrier Scaling of Free-Time Escape Rates

Barrier-resolved design and estimator. For gradient diffusion, Eqs. (9) and (10) give $\log k_D = \log \kappa(D) - \Delta\Phi/D$, where $\Delta\Phi = \Phi(s) - \Phi(a)$; thus the leading slope against $1/D$ is $-\Delta\Phi$. The scaled Müller–Brown potential supplies a demanding nonseparable benchmark through its curved saddle connections and anisotropic wells. Its deep-to-central and right-to-central exits have barriers 1.060347 and 0.359178, respectively. For each exit and each of seven noise levels, we simulate eight independent batches of 300 trajectories until they reach a target core, a competing core, or the fixed observation horizon. The resulting $33{,}600$ trajectories yield cause-specific rate estimates from target-event counts divided by total observed exposure. Our primary analysis fits $\log \widehat{k}_D = \alpha + \beta/D$ with an unconstrained slope over the four prespecified lowest-noise levels and constructs 95% bootstrap intervals by resampling whole batches. Complete event definitions and numerical audits appear in Appendix B.

Observed barrier scaling. For both exits, $\log \widehat{k}_D$ is nearly linear in $1/D$, with $R^2 = 0.99975$ and 0.99984. The unconstrained fits yield slopes of -1.00256 and -0.35866, recovering the corresponding barrier predictions -1.06035 and -0.35918 to within 5.45% and 0.14%, respectively. The right-exit bootstrap interval contains its theoretical slope; the deep-exit estimate is 5.45% shallower, with an interval that excludes the asymptotic value. Halving the time step and changing the initialization keep the resulting slope shifts within the prespecified equivalence margins. Taken together, the primary and robustness analyses recover barrier-controlled exponential scaling across exits that differ substantially in barrier height, local curvature, and transition geometry.

4.2 Equal Barriers, Unequal Fixed-Time Path Probabilities

Equal-barrier design. We next isolate fixed-time path probability by testing whether equal-barrier routes can nevertheless carry different small-tube probabilities. We construct an analytic two-well landscape whose upper and lower saddles have identical barrier heights but a $7.59\times$ contrast in stable-direction curvature. Equal barriers remove the free-time barrier preference, while the curvature contrast makes the divergence correction decisive: the full Onsager–Machlup action and its quadratic-only truncation select different channels. At $T = 8.5730$ and $D = 0.117397$, fixed-

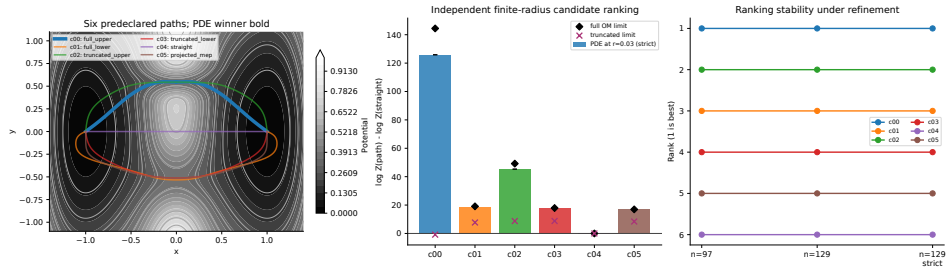

Figure 2: **An action-independent tube calculation selects the full-OM route.** Left: six prespecified fixed-endpoint paths on the equal-barrier landscape; full OM selects upper-channel $c00$, whereas its quadratic-only truncation selects lower-channel $c03$. Middle: backward-Kolmogorov log tube-probability ratios at $\varepsilon/\|b-a\| = .03$ relative to straight path $c04$; diamonds and crosses mark the full and truncated zero-radius limits. Right: the PDE order $c00 > c02 > c01 > c03 > c05 > c04$ persists under grid and solver-tolerance refinement, matching full OM exactly while the truncation predicts a different winner and ranking.

endpoint optimization selects the upper channel ($c00$) under the full action and the lower channel ($c03$) under the truncation. We augment these two winners with four optimized or projected alternatives, yielding a six-path candidate set fixed before any tube-probability calculation.

Action-independent tube probabilities. For each candidate γ, a backward-Kolmogorov solve on a moving absorbing tube computes $Z_\varepsilon(\gamma) = \mathbb{P}(B_\varepsilon(\gamma))$ from the drift and path coordinates, independently testing the conflicting action rankings. We evaluate five prespecified tube radii and verify the smallest-radius result by spatial-grid refinement and a 100-fold tighter integration tolerance. Appendix C reports the solver calibration, candidate construction and complete radius ladder.

The full action recovers the complete route ordering. At the smallest radius, the PDE yields the complete ordering $c00 > c02 > c01 > c03 > c05 > c04$, matching the full action exactly at every audited resolution. The truncation instead predicts $c03 > c02 > c05 > c01 > c04 > c00$, reversing both the winning route and the global rank structure. The PDE winner exceeds its nearest competitor by 80.557 nats, nearly three orders of magnitude above the combined deterministic numerical margin of 0.081 nats. For $c00$, the two zero-radius extrapolations, 140.283 and 142.703, approach the full-action limit 144.434 and remain sharply separated from the truncated prediction -0.837; the $c03$ extrapolations show the same agreement. For the prespecified six-path family at fixed T and D, the divergence correction therefore determines which candidate has the largest small-tube probability.

4.3 Topology and Barrier Differences Govern Next-State Choice

Topology-aware intervention and sampling. We finally test how direct-exit topology and relative barriers combine into normalized next-state choice. In the five-well Gaussian-mixture landscape, saddle verification gives exactly three direct neighbors of the source basin, $\mathcal{C}(S) = \{E, U, L\}$; state R is adjacent only to E and serves as a nonadjacent negative control for exits from S. We compare the baseline landscape $G0$ with $G1$, obtained by multiplying only the unnormalized mixture weight of U by 1.5 before renormalization; all component centers and covariances remain fixed. A component-center distance rule is invariant to this intervention, whereas the landscape model predicts a barrier-mediated redistribution of next-state probability. At each of five noise levels per landscape, five independently evolved approximate quasi-stationary populations each seed a batch of $3{,}200$ releases. We record the first departure from the source basin and retain right-censoring, invalid events, and nonadjacent departures in the complete event accounting.

Barrier-difference log-odds and intervention response. Of $160{,}000$ releases, $159{,}983$ yield a valid direct exit, 17 are right-censored, and none exits directly from S to the nonadjacent state R. Unconstrained fits of log odds against $1/D$ for U/E and L/E in $G0$, and for E/U and L/U in $G1$, recover the corresponding negative barrier differences to within 4.5%; all four theoretical slopes lie inside their whole-batch bootstrap intervals. The common escape scale cancels from these

Table 1: **Barrier differences predict normalized log-odds slopes.** Reference exits are E in $G0$ and U in $G1$. Theory is $-[\Delta\Phi_{Sb} - \Delta\Phi_{Sr}]$; unconstrained fits use the five shared noise levels. All four predictions lie within the 95% whole-batch bootstrap intervals (maximum relative error 4.42%).

Landscape	Log-odds	Theory	Estimate (95% CI)	Rel. error
$G0$	U/E	-0.17419	-0.17892 $[-0.20592, -0.15789]$	2.72%
$G0$	L/E	-0.53342	-0.53940 $[-0.55790, -0.52009]$	1.12%
$G1$	E/U	-0.04990	-0.05210 $[-0.07060, -0.03352]$	4.42%
$G1$	L/U	-0.58330	-0.57443 $[-0.61508, -0.52596]$	1.52%

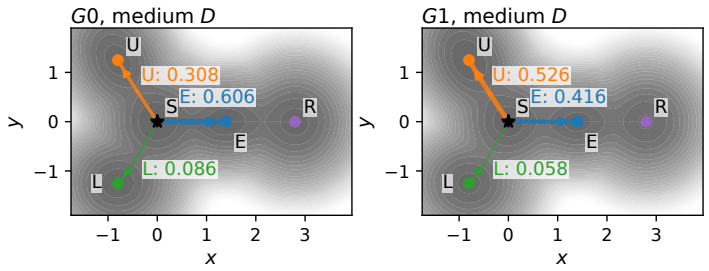

Figure 3: **A fixed-center energy intervention reverses normalized next-state choice.** Contours show $G0$ (left) and $G1$ (right), with source S, direct basin neighbors E, U, L, and R adjacent only to E. Arrow widths and labels give middle-noise first-exit probabilities along saddle-verified edges. Multiplying only U's unnormalized mixture weight by 1.5 reverses the dominant exit from E (0.606) to U (0.526), with centers and covariances fixed. Among $160{,}000$ releases, none exits directly from S to R, consistent with the verified topology.

ratios, so the slope agreement directly tests the normalized choice law independently of absolute-rate calibration.

Intervention, topology, and finite-noise calibration. The fixed-center intervention reverses the dominant exit from E $(0.606$ at the middle noise level) to U (0.526) and increases $P_D(U \mid S)$ by 0.132–0.301 across the noise grid, with a mean increase of 0.218 (95% CI $[0.212, 0.222]$). Because component centers and covariances remain fixed, this response directly separates landscape-mediated choice from center distance. A separate topology control normalizes the distance score over $\{E, U, L, R\}$ instead of the verified direct-exit set $\{E, U, L\}$; including the nonadjacent state worsens per-event negative log-likelihood by approximately 0.0875 in all ten landscape–noise cells. At the representative middle noise level, independently computed Eyring–Kramers prefactors reduce Jensen–Shannon divergence from 1.095×10^{-3} to 4.997×10^{-5} in $G0$ and from 6.343×10^{-4} to 9.381×10^{-5} in $G1$, reductions of 95.4% and 85.2%. Together, these results identify complementary roles for the three theoretical quantities: basin adjacency defines the competitor set, barrier differences determine weak-noise log-odds slopes, and curvature prefactors sharpen finite-noise calibration. Appendix D reports the time-step, initialization, and source-core sensitivity analyses.

5 Discussion and Scope

Together, the experiments trace a progression from landscape geometry to stochastic association. Free-time rates recover barrier-controlled exponential scaling; on the equal-barrier landscape, an action-independent tube PDE recovers the complete ordering predicted by the full Onsager–Machlup action; under competing exits, adjacency fixes the competitor set, barrier differences determine log-odds slopes, and curvature prefactors sharpen finite-noise calibration. The central insight is that a shared potential generates three probability objects—fixed-time path probabilities, free-time escape rates, and normalized next-state choice—with different limits and normalizations. Controlled gradient landscapes make each layer independently testable and yield a protocol for learned representations: identify nondegenerate attractors, verify saddle-mediated adjacency, and evaluate path and choice predictions on held-out data.

Reproducibility Statement

Appendices A–D specify the proofs, potentials, event definitions, numerical methods, sampling budgets, frozen inference protocols, and complete outcomes. Saved artifacts include seeds, per-trajectory events, PDE cells, path coefficients, environment manifests, and source-result hashes.

AI Use Disclosure

Large language model tools assisted with literature discovery, research ideation, feedback on experimental design and numerical analyses, formulation and checking of mathematical arguments and proof steps, and drafting and editing the manuscript. The authors independently verified the cited sources, theoretical claims, code, experiments, and reported numerical results and take full responsibility for the manuscript.

References

Joshua T. Abbott, Joseph L. Austerweil, and Thomas L. Griffiths. Random walks on semantic networks can resemble optimal foraging. *Psychological Review*, 122(3):558–569, 2015. doi: 10.1037/a0038693.

Joseph L. Austerweil, Joshua T. Abbott, and Thomas L. Griffiths. Human memory search as a random walk in a semantic network. In *Advances in Neural Information Processing Systems*, volume 25, pp. 3050–3058, 2012. URL https://proceedings.neurips.cc/paper/2012/hash/14d9e8007c9b41f57891c48e07c23f57-Abstract.html.

A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes i: Sharp asymptotics for capacities and exit times. *Journal of the European Mathematical Society*, 6(4):399–424, 2004. doi: 10.4171/JEMS/14.

Andrew Cattle and Xiaojuan Ma. Predicting word association strengths. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1283–1288. Association for Computational Linguistics, 2017. doi: 10.18653/v1/D17-1132. URL https://aclanthology.org/D17-1132/.

Allan M. Collins and Elizabeth F. Loftus. A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407–428, 1975. doi: 10.1037/0033-295X.82.6.407.

L. F. Cugliandolo and V. Lecomte. Rules of calculus in the path integral representation of white noise langevin equations: the onsager–machlup approach. *Journal of Physics A: Mathematical and Theoretical*, 50(34):345001, 2017. doi: 10.1088/1751-8121/aa7dd6.

Simon De Deyne, Danielle J. Navarro, Amy Perfors, Marc Brysbaert, and Gert Storms. The small world of words english word association norms for over 12,000 cue words. *Behavior Research Methods*, 51(3):987–1006, 2019. doi: 10.3758/s13428-018-1115-7.

Dimas Abreu Dutra, Bruno Otavio Soares Teixeira, and Luis Antonio Aguirre. Maximum a posteriori state path estimation: Discretization limits and their interpretation. *Automatica*, 50(5): 1360–1368, 2014. doi: 10.1016/j.automatica.2014.03.003.

Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Minimum action method for the study of rare events. *Communications on Pure and Applied Mathematics*, 57(5):637–656, 2004. doi: 10.1002/cpa.20005.

M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, 2 edition, 1998. doi: 10.1007/978-1-4612-0611-8.

Michael Gutmann and Aapo Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pp. 297–304. PMLR, 2010. URL https://proceedings.mlr.press/v9/gutmann10a.html.

Thomas T. Hills, Michael N. Jones, and Peter M. Todd. Optimal foraging in semantic memory. *Psychological Review*, 119(2):431–440, 2012. doi: 10.1037/a0027373.

John J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982. doi: 10.1073/pnas.79.8.2554.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6:695–709, 2005. URL https://www.jmlr.org/papers/v6/hyvarinen05a.html.

Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic Differential Equations and Diffusion Processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland, 2 edition, 1989.

Michael N. Jones, Thomas T. Hills, and Peter M. Todd. Hidden processes in structural representations: A reply to Abbott, Austerweil, and Griffiths (2015). *Psychological Review*, 122(3):570–574, 2015. doi: 10.1037/a0039248.

Dmitry Krotov and John J. Hopfield. Dense associative memory for pattern recognition. In *Advances in Neural Information Processing Systems*, volume 29, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/eaae339c4d89fc102edd9dbdb6a28915-Abstract.html.

Yann LeCun, Sumit Chopra, Raia Hadsell, Marc'Aurelio Ranzato, and Fu Jie Huang. A tutorial on energy-based learning. In Gökhan Bakir, Thomas Hofmann, Bernhard Schölkopf, Alexander J. Smola, Ben Taskar, and S. V. N. Vishwanathan (eds.), *Predicting Structured Data*, pp. 191–246. MIT Press, 2007.

Tony Lelièvre, Dorian Le Peutrec, and Boris Nectoux. Eyring–kramers exit rates for the overdamped langevin dynamics: The case with saddle points on the boundary. *Journal de l'École polytechnique — Mathématiques*, 12:881–982, 2025. doi: 10.5802/jep.303.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, volume 26, pp. 3111–3119, 2013.

Douglas L. Nelson, Cathy L. McEvoy, and Thomas A. Schreiber. The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3):402–407, 2004. doi: 10.3758/BF03195588.

Lars Onsager and S. Machlup. Fluctuations and irreversible processes. *Physical Review*, 91(6):1505–1512, 1953. doi: 10.1103/PhysRev.91.1505.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763, 2021.

Sanjeev Raja, Martin Šípka, Michael Psenka, Tobias Kreiman, Michal Pavelka, and Aditi S. Krishnapriyan. Action-minimization meets generative modeling: Efficient transition path sampling with the onsager–machlup functional. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 50972–51008, 2025. URL https://proceedings.mlr.press/v267/raja25a.html.

Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Thomas Adler, Lukas Gruber, Markus Holzleitner, David Kreil, Michael Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. Hopfield networks is all you need. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=tL89RnzIiCd.

Maria A. Rodriguez and Paola Merlo. Word associations and the distance properties of context-aware word embeddings. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pp. 376–385. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.conll-1.30. URL https://aclanthology.org/2020.conll-1.30/.

Lawrence A. Shepp and Ofer Zeitouni. A note on conditional exponential moments and onsager–machlup functionals. *The Annals of Probability*, 20(2):652–654, 1992. doi: 10.1214/aop/1176989796.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/3001ef257407d5a371a96dcd947c7d93-Abstract.html.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Kai Ueltzhöffer, Diana J. N. Armbruster-Genç, and Christian J. Fiebach. Stochastic dynamics underlying cognitive stability and flexibility. *PLOS Computational Biology*, 11(6):e1004331, 2015. doi: 10.1371/journal.pcbi.1004331.

Leijun Ye, Jianfeng Feng, and Chunhe Li. Controlling brain dynamics: Landscape and transition path for working memory. *PLOS Computational Biology*, 19(9):e1011446, 2023. doi: 10.1371/journal.pcbi.1011446.

Ofer Zeitouni. On the onsager–machlup functional of diffusion processes around non-c^2 curves. *The Annals of Probability*, 17(3):1037–1054, 1989. doi: 10.1214/aop/1176991255.

A Technical Details and Proofs

A.1 Proof of the Onsager–Machlup Characterization

Proof of Theorem 2. Write $\mu = -\nabla\Phi$. We invoke the fixed-initial Onsager–Machlup small-tube ratio theorem for the additive isotropic diffusion

$$dX_t = \mu(X_t)\,dt + \sqrt{2D}\,dW_t.$$

The classical result for smooth reference curves is given by Ikeda & Watanabe (1989, Chapter VI, Section 9); the extension to Cameron–Martin reference paths used here is established in Zeitouni (1989); Shepp & Zeitouni (1992). In the globally bounded setting $\mu \in C_b^2(\mathbb{R}^d;\mathbb{R}^d)$, for any $\gamma, \eta \in H^1([0,T];\mathbb{R}^d)$ with $\gamma(0) = \eta(0) = a$, this theorem yields, for uniform tubes,

$$\lim_{\varepsilon\downarrow 0} \frac{\mathbb{P}_a(B_\varepsilon(\gamma))}{\mathbb{P}_a(B_\varepsilon(\eta))} = \exp\{-S^T_{\mu,D}[\gamma] + S^T_{\mu,D}[\eta]\}, \tag{18}$$

where

$$S^T_{\mu,D}[\gamma] = \int_0^T \left[\frac{1}{4D}\|\dot\gamma(t) - \mu(\gamma(t))\|^2 + \frac{1}{2}\nabla\cdot\mu(\gamma(t))\right] dt. \tag{19}$$

This is the continuous-time state-path objective selected by shrinking tube probabilities (Dutra et al., 2014).

Assumption 1 implies $\mu \in C^2(\mathbb{R}^d;\mathbb{R}^d)$, but it does not require μ or its first two derivatives to be globally bounded. This does not restrict the present tube comparison, because the small-tube events around two fixed reference paths are localized to a compact region. Indeed, the images of γ and η are compact. Choose $\varepsilon_0 > 0$ and a compact set K whose interior contains their closed ε_0-neighborhoods. A smooth cutoff then produces $\widetilde\mu \in C_b^2(\mathbb{R}^d;\mathbb{R}^d)$ that agrees with μ on a neighborhood of K.

Couple the diffusions driven by μ and $\widetilde\mu$ using the same Brownian motion and the same initial condition. Pathwise uniqueness implies that the coupled processes agree until they leave the region

on which the two drifts coincide. For every $0 < \varepsilon < \varepsilon_0$, each of the tube events $B_\varepsilon(\gamma)$ and $B_\varepsilon(\eta)$ remains inside this region. Hence the corresponding tube-event indicators, and therefore their probabilities, are identical for the two diffusions. Applying the cited C_b^2 small-tube ratio theorem to $\widetilde{\mu}$ proves Eq. (18) for the original diffusion, while the two action functionals coincide along γ and η.

For $\mu = -\nabla\Phi$,

$$\nabla\cdot\mu = -\Delta\Phi, \qquad \dot\gamma - \mu(\gamma) = \dot\gamma + \nabla\Phi(\gamma).$$

Thus Eq. (19) is exactly $S^T_{\mathrm{OM},D}$ in Eq. (4), and Eq. (18) proves Eq. (3).

Finally, if

$$\gamma^*_{\mathrm{OM}} \in \arg\min_{\gamma\in\Gamma^T_{a,b}} S^T_{\mathrm{OM},D}[\gamma],$$

then for any admissible η,

$$-S^T_{\mathrm{OM},D}[\gamma^*_{\mathrm{OM}}] + S^T_{\mathrm{OM},D}[\eta] \geq 0.$$

Therefore

$$\lim_{\varepsilon\downarrow 0} \frac{\mathbb{P}_a(B_\varepsilon(\gamma^*_{\mathrm{OM}}))}{\mathbb{P}_a(B_\varepsilon(\eta))} \geq 1,$$

so every action minimizer is a most-probable reference path in the small-tube sense. □

Independent coefficient check via Girsanov. Although the theorem above already proves the required OM formula, it is useful to independently verify the coefficient and sign of the divergence term.

Under the stronger bounded-drift condition above, Novikov's condition holds on every finite time interval. Girsanov's theorem gives, relative to the zero-drift process with the same diffusion coefficient,

$$\frac{dP^\mu}{dP^W} = \exp\left[\frac{1}{2D}\int_0^T \mu(X_t)\cdot dX_t^{\mathrm{It\hat{o}}} - \frac{1}{4D}\int_0^T \|\mu(X_t)\|^2 dt\right].$$

For constant diffusion, the Itô–Stratonovich conversion follows from quadratic covariation:

$$\int_0^T \mu(X_t)\circ dX_t = \int_0^T \mu(X_t)\cdot dX_t^{\mathrm{It\hat{o}}} + \frac{1}{2}\sum_{i=1}^d [\mu_i(X), X_i]_T.$$

Since

$$d[X_i, X_j]_t = 2D\,\delta_{ij}\,dt,$$

we have

$$\sum_{i=1}^d [\mu_i(X), X_i]_T = 2D\int_0^T \nabla\cdot\mu(X_t)\,dt.$$

Consequently,

$$\int_0^T \mu(X_t)\cdot dX_t^{\mathrm{It\hat{o}}} = \int_0^T \mu(X_t)\circ dX_t - D\int_0^T \nabla\cdot\mu(X_t)\,dt. \tag{20}$$

For $\mu = -\nabla\Phi$, the Stratonovich chain rule gives

$$\int_0^T \mu(X_t)\circ dX_t = -\big(\Phi(X_T) - \Phi(X_0)\big),$$

while

$$\nabla\cdot\mu = -\Delta\Phi.$$

Substitution into the Girsanov exponent gives

$$-\frac{\Phi(X_T) - \Phi(X_0)}{2D} + \frac{1}{2}\int_0^T \Delta\Phi(X_t)\,dt - \frac{1}{4D}\int_0^T \|\nabla\Phi(X_t)\|^2 dt.$$

Combining this with the zero-drift Cameron–Martin contribution

$$-\frac{1}{4D}\int_0^T \|\dot\gamma(t)\|^2 dt$$

and using

$$\Phi(\gamma(T)) - \Phi(\gamma(0)) = \int_0^T \dot\gamma(t)\cdot\nabla\Phi(\gamma(t))\,dt$$

gives

$$-\frac{1}{4D}\int_0^T \|\dot\gamma + \nabla\Phi(\gamma)\|^2 dt + \frac{1}{2}\int_0^T \Delta\Phi(\gamma)\,dt = -S^T_{\mathrm{OM},D}[\gamma].$$

This independently confirms both the coefficient $1/2$ and the negative sign of the Laplacian term in the action.

A.2 Relation to the Fixed-Endpoint Action of Raja et al.

To compare Eq. (4) directly with Raja et al. (2025), consider the drift

$$\mu(x) = -\frac{1}{\zeta}\nabla\phi(x).$$

The OM action becomes

$$S^T_{\mathrm{OM},D}[\gamma] = \frac{1}{4D}\int_0^T \left\|\dot\gamma + \frac{1}{\zeta}\nabla\phi(\gamma)\right\|^2 dt - \frac{1}{2\zeta}\int_0^T \Delta\phi(\gamma)\,dt.$$

Expanding the square and using

$$\int_0^T \dot\gamma(t)\cdot\nabla\phi(\gamma(t))\,dt = \phi(b) - \phi(a)$$

gives

$$S^T_{\mathrm{OM},D}[\gamma] = \frac{1}{4D}\int_0^T \left[\|\dot\gamma\|^2 + \frac{1}{\zeta^2}\|\nabla\phi(\gamma)\|^2 - \frac{2D}{\zeta}\Delta\phi(\gamma)\right] dt + \frac{\phi(b)-\phi(a)}{2D\zeta}.$$

The integral term is exactly the fixed-endpoint form used in Appendix D of Raja et al. (2025). The remaining term depends only on the endpoints a and b. Hence, for fixed endpoints and fixed T, the two forms differ only by an additive constant and therefore have identical path minimizers.

A.3 Discretization and the Minimum-Energy Functional

There is an important distinction between discretizing the Onsager–Machlup functional and maximizing the joint transition density of an Euler–Maruyama discretization.

The Euler–Maruyama scheme for Eq. (1) is

$$X_{k+1} = X_k - \nabla\Phi(X_k)\Delta t + \sqrt{2D\Delta t}\,\xi_k, \qquad \xi_k \sim \mathcal{N}(0, I_d).$$

Multiplying its Gaussian transition densities yields the discrete objective

$$S^{(L)}_{\mathrm{E}} = \sum_{k=0}^{L-1} \frac{\Delta t}{4D}\left\|\frac{x_{k+1}-x_k}{\Delta t} + \nabla\Phi(x_k)\right\|^2.$$

This is the discrete analogue of the minimum-energy functional and does not contain the drift-divergence correction.

It should not, in general, be identified with the continuous-time Onsager–Machlup path objective (Dutra et al., 2014). In particular, when $\nabla\cdot\mu$ depends on the state, the minimum-energy and Onsager–Machlup functionals may have different minimizers. For fixed T, they are guaranteed to have the same path minimizers whenever the integrated divergence correction is path-independent; a simple sufficient case is a state-independent drift divergence.

For the full OM objective, we instead discretize Eq. (4) itself. A midpoint approximation is

$$S_{\mathrm{OM}}^{(L)} = \sum_{k=0}^{L-1} \Delta t \left[\frac{1}{4D} \left\| \frac{x_{k+1} - x_k}{\Delta t} + \nabla\Phi(\bar{x}_k) \right\|^2 - \frac{1}{2}\Delta\Phi(\bar{x}_k) \right],$$

where

$$\bar{x}_k = \frac{x_k + x_{k+1}}{2}.$$

The Euler minimum-energy objective and the full OM objective are therefore treated as distinct path objectives in our experiments rather than as two equivalent discretization conventions.

A.4 Freidlin–Wentzell Scaling and Transition-Cost Properties

Why the full OM action is not a free-time cost. Let m be a non-degenerate local minimum of Φ. Then

$$\nabla\Phi(m) = 0, \qquad \nabla^2\Phi(m) \succ 0, \qquad \Delta\Phi(m) > 0.$$

Fix arbitrary endpoints $a, b \in \mathbb{R}^d$. Choose any H^1 path γ_1 from a to m with duration T_1 and any H^1 path γ_2 from m to b with duration T_2. Since these paths have compact images and $\Phi \in C^2$, both have finite OM action.

For any $\tau > 0$, insert between them the constant path

$$\gamma_m(t) \equiv m, \qquad t \in [0, \tau].$$

By additivity of the action under concatenation,

$$S_{\mathrm{OM},D}^{T_1+\tau+T_2}[\gamma_1 * \gamma_m * \gamma_2] = S_{\mathrm{OM},D}^{T_1}[\gamma_1] + S_{\mathrm{OM},D}^{\tau}[\gamma_m] + S_{\mathrm{OM},D}^{T_2}[\gamma_2].$$

At the minimum m,

$$S_{\mathrm{OM},D}^{\tau}[\gamma_m] = -\frac{\tau}{2}\Delta\Phi(m).$$

Hence

$$S_{\mathrm{OM},D}^{T_1+\tau+T_2}[\gamma_1 * \gamma_m * \gamma_2] = C - \frac{\tau}{2}\Delta\Phi(m),$$

where C is independent of τ. Since $\Delta\Phi(m) > 0$,

$$S_{\mathrm{OM},D}^{T_1+\tau+T_2}[\gamma_1 * \gamma_m * \gamma_2] \longrightarrow -\infty \qquad \text{as } \tau \to \infty.$$

Therefore the unrestricted free-time infimum of the full OM action is $-\infty$ whenever such a stable minimum is available. This justifies restricting the full OM functional to fixed time horizons.

Freidlin–Wentzell scaling. Set

$$\varepsilon = 2D.$$

Then Eq. (1) takes the standard small-noise form

$$dX_t = -\nabla\Phi(X_t)\,dt + \sqrt{\varepsilon}\,dW_t.$$

The Freidlin–Wentzell path-space large-deviation principle has speed $\varepsilon^{-1} = (2D)^{-1}$ and rate functional

$$I_T[\gamma] = \frac{1}{2}\int_0^T \|\dot{\gamma}(t) + \nabla\Phi(\gamma(t))\|^2 dt$$

for absolutely continuous paths with square-integrable derivative. Thus the logarithmic probability of a path neighborhood is governed by

$$\exp\left(-\frac{I_T[\gamma]}{2D}\right)$$

at leading weak-noise order, and minimizing I_T over both paths and durations yields the quasipotential in Eq. (7) (Freidlin & Wentzell, 1998).

Proof of Proposition 3. We prove the claims in order.

(i) Nonnegativity and $V(a,a)=0$. For every admissible path,

$$I_T[\gamma] = \frac{1}{2}\int_0^T \|\dot\gamma + \nabla\Phi(\gamma)\|^2 dt \geq 0.$$

Taking infima preserves nonnegativity, hence

$$V(a,b) \geq 0.$$

To show $V(a,a)=0$, consider the constant path $\gamma_T(t) \equiv a$ on $[0,T]$. Its action is

$$I_T[\gamma_T] = \frac{T}{2}\|\nabla\Phi(a)\|^2.$$

Therefore

$$0 \leq V(a,a) \leq \frac{T}{2}\|\nabla\Phi(a)\|^2$$

for every $T>0$. Letting $T \downarrow 0$ gives

$$V(a,a) = 0.$$

(ii) Triangle inequality. Let $\epsilon > 0$. By the definition of the infimum, there exist durations $T_1, T_2 > 0$ and admissible paths $\gamma_1 \in \Gamma_{a,b}^{T_1}$ and $\gamma_2 \in \Gamma_{b,c}^{T_2}$ such that

$$I_{T_1}[\gamma_1] \leq V(a,b) + \epsilon,$$

and

$$I_{T_2}[\gamma_2] \leq V(b,c) + \epsilon.$$

Their concatenation

$$\gamma_1 * \gamma_2 \in \Gamma_{a,c}^{T_1+T_2}$$

is absolutely continuous. Since the action is an integral over time,

$$I_{T_1+T_2}[\gamma_1 * \gamma_2] = I_{T_1}[\gamma_1] + I_{T_2}[\gamma_2].$$

Hence

$$V(a,c) \leq V(a,b) + V(b,c) + 2\epsilon.$$

Since $\epsilon > 0$ is arbitrary,

$$V(a,c) \leq V(a,b) + V(b,c).$$

(iii) Zero cost along deterministic relaxation. Suppose the deterministic trajectory φ initialized at a satisfies

$$\dot\varphi(t) = -\nabla\Phi(\varphi(t)), \qquad \varphi(0) = a, \qquad \varphi(t) \to b \quad \text{as } t \to \infty.$$

Continuity of the vector field implies that b is an equilibrium, so

$$\nabla\Phi(b) = 0.$$

For any $T>0$, following φ on $[0,T]$ incurs zero action:

$$I_T[\varphi] = 0.$$

Let

$$y_T := \varphi(T).$$

Connect y_T to b over one unit of time by

$$\eta_T(s) = (1-s)y_T + sb, \qquad s \in [0,1].$$

Since $\nabla\Phi$ is locally Lipschitz near b and $\nabla\Phi(b)=0$, there exists $L>0$ such that, for all sufficiently large T,

$$\|\nabla\Phi(\eta_T(s))\| \leq L\|y_T - b\|.$$

Also,

$$\|\dot\eta_T(s)\| = \|b - y_T\|.$$

Therefore

$$I_1[\eta_T] = \frac{1}{2}\int_0^1 \|\dot{\eta}_T + \nabla\Phi(\eta_T)\|^2 ds \leq \frac{1}{2}(1+L)^2\|y_T - b\|^2.$$

Because $y_T \to b$,

$$I_1[\eta_T] \to 0.$$

Concatenating $\varphi|_{[0,T]}$ with η_T gives

$$0 \leq V(a,b) \leq I_1[\eta_T] \longrightarrow 0,$$

and hence

$$V(a,b) = 0.$$

□

Proof of Proposition 4. For any absolutely continuous path γ, expanding the square in the FW action gives

$$\begin{aligned} I_T[\gamma] &= \frac{1}{2}\int_0^T \|\dot{\gamma} - \nabla\Phi(\gamma)\|^2\, dt + 2\int_0^T \dot{\gamma}\cdot\nabla\Phi(\gamma)\, dt \\ &= \frac{1}{2}\int_0^T \|\dot{\gamma} - \nabla\Phi(\gamma)\|^2\, dt + 2\big(\Phi(\gamma(T)) - \Phi(\gamma(0))\big). \end{aligned} \tag{21}$$

Every path admissible in Eq. (8) starts at a and ends on Σ_{ab}. Since s_{ab} minimizes Φ on that channel, Eq. (21) implies

$$I_T[\gamma] \geq 2\big(\Phi(\gamma(T)) - \Phi(a)\big) \geq 2\big(\Phi(s_{ab}) - \Phi(a)\big).$$

Taking the infimum yields the lower bound on $V_{\mathrm{dir}}(a,b)$.

For the matching upper bound, reverse the descending branch from s_{ab} to a. Along this uphill branch, $\dot{\gamma} = \nabla\Phi(\gamma)$, so Eq. (21) gives action equal to twice the increase in potential. Truncate the branch near its equilibrium endpoints and join it to a and s_{ab} with short segments that remain in $\mathcal{B}_a$ before the terminal time. Local regularity and the vanishing drift at both endpoints make the connectors' actions tend to zero. These admissible paths therefore have actions tending to $2(\Phi(s_{ab})-\Phi(a))$. Combining the bounds proves the first identity in Eq. (9). Applying the same argument to Σ_{ba} proves the reverse identity. □

A.5 Association Probabilities from Metastable Exit Rates

We now state the metastable conditions used in Section 3.3. Consider the overdamped gradient diffusion

$$dX_t = -\nabla\Phi(X_t)\, dt + \sqrt{2D}\, dW_t,$$

and fix a source minimum a. On the relevant sublevel set, assume that Φ is confining and Morse, the basin of a contains no additional minimum at the escape scale, and every direct boundary channel from a to a competitor $b \in \mathcal{C}(a)$ passes through one of finitely many isolated index-one saddles in a set $\mathcal{S}_{ab}$. We further assume the metastable separation under which intrabasin relaxation is asymptotically faster than interbasin escape. These conditions make the reduced competing-rate description well defined.

Let $H_a = \nabla^2\Phi(a)$, $H_s = \nabla^2\Phi(s)$, and let $\lambda_-(s) < 0$ be the unique negative eigenvalue of H_s. For the first exit from the source basin, the Eyring–Kramers expansion for the aggregate rate through channels toward b is

$$k_D(a,b) = \sum_{s\in\mathcal{S}_{ab}} \frac{|\lambda_-(s)|}{\pi}\sqrt{\frac{\det H_a}{|\det H_s|}}\exp\left[-\frac{\Phi(s)-\Phi(a)}{D}\right]\big(1+o(1)\big). \tag{22}$$

Define the lowest barrier among the direct channels to b by

$$\Delta\Phi_{ab} := \min_{s\in\mathcal{S}_{ab}}\big[\Phi(s) - \Phi(a)\big].$$

The direct-exit barrier identity gives $V_{\mathrm{dir}}(a,b) = 2\Delta\Phi_{ab}$. Factoring this scale from Eq. (22) yields

$$k_D(a,b) = \kappa_{ab}(D) \exp\left(-\frac{V_{\mathrm{dir}}(a,b)}{2D}\right), \qquad \lim_{D\to 0} 2D \log \kappa_{ab}(D) = 0.$$

The curvature factors of saddles attaining the lowest barrier add in κ_{ab}; higher-barrier channels contribute exponentially smaller terms. This establishes Eq. (10) for the aggregate direct rate to b (Bovier et al., 2004; Lelièvre et al., 2025).

Proof of Proposition 5. Substituting Eq. (10) into Eq. (11) gives Eq. (12). Set $V_* := V_{\min}^{\mathrm{dir}}(a)$ and factor $\exp[-V_*/(2D)]$ from its denominator. Then

$$\begin{aligned} -2D \log P_D^{\mathrm{assoc}}(b \mid a) = V_{\mathrm{dir}}(a,b) - V_* - 2D \log \kappa_{ab}(D) \\ + 2D \log \left[\sum_{c \in \mathcal{C}(a)} \kappa_{ac}(D) \exp\left(-\frac{V_{\mathrm{dir}}(a,c) - V_*}{2D}\right) \right]. \end{aligned}$$

Choose $c_* \in \mathcal{C}(a)$ with $V_{\mathrm{dir}}(a,c_*) = V_*$. The sum inside the logarithm is bounded below by $\kappa_{ac_*}(D)$ and above by $\sum_{c\in\mathcal{C}(a)} \kappa_{ac}(D)$. Finiteness of $\mathcal{C}(a)$ and subexponentiality of its prefactors imply that $2D$ times the logarithm of both bounds tends to zero. The prefactor term for b also vanishes on this scale, proving Eq. (13). □

If a unique $b_* \in \mathcal{C}(a)$ has the smallest direct-exit cost, then $P_D^{\mathrm{assoc}}(b_* \mid a) \to 1$ as $D \to 0$. When several exits share the minimum cost, their probabilities agree at exponential order and their relative limiting weights, when they exist, are determined by the corresponding prefactors.

A.6 Potential Learning and Association Attractors

Proof of Proposition 6. Let

$$\pi_D(z) = Z_D^{-1} \exp[-\Phi(z)/D]$$

be normalizable. The generator admits the divergence-form identity

$$\mathcal{L}_D f = -\nabla\Phi \cdot \nabla f + D\Delta f = D\, \pi_D^{-1} \nabla \cdot (\pi_D \nabla f). \tag{23}$$

For $f, g \in C_c^\infty(\mathbb{R}^d)$, integration by parts gives

$$\int f\, \mathcal{L}_D g\, \pi_D\, dz = -D \int \nabla f \cdot \nabla g\, \pi_D\, dz = \int g\, \mathcal{L}_D f\, \pi_D\, dz. \tag{24}$$

Thus the generator is symmetric in $L^2(\pi_D)$. Together with the nonexplosion assumption, which makes the Markov semigroup conservative, Eq. (24) establishes invariance and reversibility of π_D.

Now set $\Phi = \Phi_p = -D_0 \log p + C$. Direct differentiation yields

$$-\nabla\Phi_p = D_0 \nabla \log p,$$

which proves the drift identity in Proposition 6. At $D = D_0$,

$$Z_{D_0}^{-1} \exp[-\Phi_p(z)/D_0] = Z_{D_0}^{-1} e^{-C/D_0} p(z) = p(z),$$

because p is normalized. More generally,

$$\exp[-\Phi_p(z)/D] = e^{-C/D} p(z)^{D_0/D}.$$

Whenever $Z_p(D) < \infty$, normalization gives the invariant-density formula in Proposition 6.

It remains to prove uniqueness. Let $\widetilde{\Phi}$ generate a gradient diffusion at scale D_0 that is reversible with strictly positive invariant density p. For a constant scalar diffusion coefficient, reversibility is equivalent to vanishing stationary probability current. Hence

$$-p\nabla\widetilde{\Phi} - D_0 \nabla p = 0, \qquad \text{so} \qquad \nabla\widetilde{\Phi} = -D_0 \nabla \log p.$$

Since $\mathbb{R}^d$ is connected, $\widetilde{\Phi} = -D_0 \log p + C$. This proves the correspondence and its uniqueness up to an additive constant. □

Proof of Proposition 7. Since $p(z) > 0$,

$$\nabla \Phi_p(z) = -D_0 \nabla \log p(z) = -\frac{D_0}{p(z)} \nabla p(z).$$

Thus

$$\nabla \Phi_p(m) = 0 \quad \Longleftrightarrow \quad \nabla p(m) = 0.$$

Hence p and Φ_p have the same critical points.

At a critical point m, differentiating once more gives

$$\nabla^2 \Phi_p(m) = -D_0 \nabla^2 \log p(m).$$

Using

$$\nabla^2 \log p = \frac{1}{p} \nabla^2 p - \frac{1}{p^2} \nabla p \, \nabla p^\top$$

and $\nabla p(m) = 0$, we obtain

$$\nabla^2 \Phi_p(m) = -\frac{D_0}{p(m)} \nabla^2 p(m). \tag{25}$$

Because $D_0 > 0$ and $p(m) > 0$,

$$\nabla^2 p(m) \prec 0 \quad \Longleftrightarrow \quad \nabla^2 \Phi_p(m) \succ 0.$$

Therefore a non-degenerate local maximum of p is exactly a non-degenerate local minimum of Φ_p, and conversely.

It remains to establish local dynamical stability. Let

$$F(z) := -\nabla \Phi_p(z)$$

be the deterministic vector field. At a potential minimum m,

$$F(m) = 0,$$

and its Jacobian is

$$DF(m) = -\nabla^2 \Phi_p(m).$$

Since

$$\nabla^2 \Phi_p(m) \succ 0,$$

all eigenvalues of $DF(m)$ are strictly negative. The linearization criterion for autonomous ODEs therefore implies that m is a locally exponentially stable equilibrium of

$$\dot{z} = -\nabla \Phi_p(z).$$

This proves the result. □

A.7 NOTATION

Table 2: Notation used in the theoretical development.

Symbol	Meaning
$\mathcal{Z} = \mathbb{R}^d$	Ambient representation space
X_t	Stochastic association state
$\gamma(t)$	Absolutely continuous reference path
$\Phi : \mathbb{R}^d \to \mathbb{R}$	Cognitive potential
$\mu = -\nabla\Phi$	Deterministic drift field
D	Diffusion coefficient / noise strength
D_0	Fixed density-calibration / potential scale
$S^T_{\mathrm{OM},D}[\gamma]$	Fixed-time full OM action
$I_T[\gamma]$	Freidlin–Wentzell rate action
$C^T_{\mathrm{OM},D}(a,b)$	Optimal fixed-time OM cost
$\mathcal{B}_a$	Deterministic basin of minimum a
Σ_{ab}	Direct boundary exit channel from a toward b
$V(a,b)$	Unrestricted FW cost to endpoint b
$V_{\mathrm{dir}}(a,b)$	Basin-constrained direct-exit cost
$\mathcal{M}_p$	Non-degenerate modes of the representation density
$\mathcal{C}(a)$	Direct metastable competitors accessible from a
$k_D(a,b)$	Aggregate direct transition rate from a to b
$\kappa_{ab}(D)$	Subexponential transition-rate prefactor
$P^{\mathrm{assoc}}_D(b \mid a)$	Normalized next-state probability

Table 3: Critical points of the scaled Müller–Brown landscape. The Hessian spectra verify that a_d, a_c, a_r are strict minima and that s_{dc}, s_{rc} are index-one saddles. The resulting barriers are $\Delta\Phi_{dc} = 1.060347$ and $\Delta\Phi_{rc} = 0.359178$.

Point	x	y	Φ	$\mathrm{eig}(\nabla^2\Phi)$
a_d	-0.558224	1.441726	-1.466995	$(4.1053, 40.6820)$
a_c	-0.050011	0.466694	-0.807678	$(2.2104, 14.7920)$
a_r	0.623499	0.028038	-1.081667	$(5.4384, 30.0540)$
s_{dc}	-0.822002	0.624313	-0.406648	$(-7.5086, 4.9024)$
s_{rc}	0.212487	0.292988	-0.722489	$(-7.3525, 5.1089)$

B Barrier Scaling of Free-Time Escape Rates: Complete Protocol and Numerical Evidence

This appendix gives the complete protocol and numerical audits for the free-time escape experiment in Section 4.1. The experiment compares two geometrically distinct exits into the same central well, so that one numerical pipeline tests barrier scaling across both a deep and a moderate barrier. All reported coordinates, energies, and noise levels are dimensionless.

B.1 Müller–Brown Landscape and Transition Geometry

We use the scaled Müller–Brown potential

$$\Phi(x, y) = 0.01 \sum_{i=1}^{4} A_i \exp\left[u_i(x - x_i)^2 + v_i(x - x_i)(y - y_i) + w_i(y - y_i)^2\right], \tag{26}$$

with coefficients

i	A_i	u_i	v_i	w_i	x_i	y_i
1	-200	-1	0	-10	1	0
2	-100	-1	0	-10	0	0.5
3	-170	-6.5	11	-6.5	-0.5	1.5
4	15	0.7	0.6	0.7	-1	1

The scale factor 0.01 fixes the numerical energy scale and does not change the critical-point geometry. Newton solves, initialized from the standard Müller–Brown wells and saddles, identify three minima—deep a_d, central a_c, and right a_r—and the two index-one saddles connecting the outer wells to a_c (Table 3).

The two tested transitions are $a_d \to a_c$ through s_{dc} and $a_r \to a_c$ through s_{rc}. The stochastic estimand is a cause-specific rate for first arrival in the radius-0.12 core of a_c, with arrival in the other outer-well core treated as a competing event. The verified saddle connections identify the barrier associated with each target channel. Replacing separatrix crossing by arrival in the destination core adds only downhill relaxation after the saddle and therefore preserves the weak-noise exponential slope. Departures from the numerical domain are recorded separately as failures; none occurred in the reported runs.

B.2 Saddle-Resolved Minimum-Energy Paths and Barrier Audit

For each transition, let e_- be the unit eigenvector associated with the negative Hessian eigenvalue at the connecting saddle s. We initialize the two branches at $s \pm 10^{-5} e_-$ and integrate the gradient flow $\dot{z} = -\nabla\Phi(z)$ toward the adjacent minima with DOP853, relative and absolute tolerances 10^{-10} and 10^{-12}, maximum step 0.04, and endpoint radius 10^{-7}. Arc-length resampling to 201 nodes per branch, with the saddle shared, yields a 401-node reference path.

We evaluate the maximum normal-force residual

$$R_\perp = \max_j \left\| \nabla\Phi(z_j) - \left(\nabla\Phi(z_j)\cdot\widehat{t}_j\right)\widehat{t}_j \right\|_2, \tag{27}$$

Table 4: Event accounting for all free-time escape simulations. Primary rows contain seven noise levels with eight batches of 300 trajectories per level. Robustness rows contain the half-step and alternative-initialization audits described in Section B.5.

Run group	Total	Target	Competing	Horizon-censored
Deep primary	16,800	16,291	493	16
Right primary	16,800	16,771	1	28
Robustness total	14,400	14,176	195	29
All simulations	48,000	47,238	689	73

where $\widehat{t}_j$ is the centered finite-difference tangent. The acceptance threshold is $R_\perp < 10^{-6}$. A refinement audit halves the saddle displacement, maximum integration step, and endpoint radius and tightens both ODE tolerances by a factor of ten; refined paths are matched to the base paths by normalized arc length.

The deep and right paths attain residuals 1.5932×10^{-7} and 6.6250×10^{-9}, respectively. Their maximum base-to-refined displacements are 5.0786×10^{-9} and 6.1298×10^{-10}, well below the prespecified 10^{-5} threshold. The paths therefore provide numerically resolved saddle-to-minimum transition geometry; their maximum energies coincide with the connecting saddle energies and recover the barriers in Table 3.

B.3 First-Exit Sampling and Cause-Specific Rate Estimation

For each exit, we choose seven noise levels through

$$D_j = \frac{\Delta\Phi}{q_j}, \qquad q_j \in \{3.9, 4.6, 5.3, 6.0, 6.7, 7.4, 8.1\}. \tag{28}$$

The primary fit uses the four lowest-noise points $q \in \{6.0, 6.7, 7.4, 8.1\}$; the larger-noise points are reserved for fit-window and curvature diagnostics.

We simulate $dZ_t = -\nabla\Phi(Z_t)\,dt + \sqrt{2D}\,dW_t$ by Euler–Maruyama with $\Delta t = 5 \times 10^{-4}$. Primary trajectories start at the source minimum. A target event is the first entry into the radius-0.12 core of a_c; entry into the other outer-well core is a competing event. Saddle crossing alone is not counted as an exit. A trajectory is censored at the horizon or recorded as a numerical-domain failure if it leaves $[-8, 8]^2$.

The frozen horizon rule is

$$H(D) = \max\left\{8000, \frac{5}{k_{\mathrm{EK}}(D)}\right\}, \qquad k_{\mathrm{EK}}(D) = \frac{\lambda_+(s)}{2\pi}\sqrt{\frac{\det\nabla^2\Phi(a)}{|\det\nabla^2\Phi(s)|}}\exp\left(-\frac{\Delta\Phi}{D}\right), \tag{29}$$

where $\lambda_+(s)$ is the positive unstable eigenvalue of the drift linearization at s. This rule gives $H(D) = 8000$ for every sampled condition.

For trajectories $m = 1, \ldots, M$ with observed times $T_m = \min\{\tau_m, H\}$ and target indicators δ_m, we estimate the cause-specific effective rate by

$$\widehat{k}_D = \frac{\sum_{m=1}^{M}\delta_m}{\sum_{m=1}^{M}T_m}. \tag{30}$$

Competing exits contribute exposure until their stopping times but do not enter the numerator. Thus $\widehat{k}_D$ is the maximum-likelihood constant-hazard estimate of the target-specific effective rate under the stated competing-risk stopping rule.

Each transition–noise pair uses eight independently seeded batches of 300 trajectories (2,400 trajectories per point). The primary experiment therefore contains 33,600 trajectories. Table 4 reports the full event accounting, including the additional robustness runs. Every point in the four-noise primary fitting window contains at least 2,340 target events; target-event fractions exceed 97.5%, and competing-event fractions remain below 2.5%.

Table 5: Primary free-time scaling results. Theory predicts $\beta = -\Delta\Phi$. Both exits exhibit nearly linear dependence on $1/D$; the estimated slopes recover the two distinct barrier scales with relative errors of 5.45% and 0.14%.

Exit	$-\Delta\Phi$	$\widehat{\beta}$	95% interval	Rel. error	R^2
Deep	-1.06035	-1.00256	$[-1.02836, -0.97487]$	5.45%	0.999751
Right	-0.35918	-0.35866	$[-0.36578, -0.35161]$	0.14%	0.999841

The implementation used Python 3.10.20, NumPy 2.2.6, SciPy 1.15.3, and Numba 0.67.0, with base seed 20260903. A batch-consistency audit verifies the complete set of 208 batches, the assigned trajectory counts, and agreement between raw event records and aggregate summaries.

B.4 Primary Barrier-Slope Inference

On the four prespecified lowest-noise points, we fit

$$\log \widehat{k}_{D_j} = \alpha + \beta D_j^{-1} \tag{31}$$

by minimizing $\sum_j N_j[\log \widehat{k}_{D_j} - \alpha - \beta/D_j]^2$, where N_j is the target-event count. Both slope and intercept are free. Uncertainty is estimated with 2,000 bootstrap replicates: within each noise level, the eight simulation batches are resampled as whole units, aggregated, and refit. The 2.5th and 97.5th percentiles form the reported 95% interval. This preserves the batch as the independent simulation unit and propagates variation through the complete estimation pipeline.

Weighted lack-of-fit statistics with two residual degrees of freedom give $p = 0.525$ for the deep exit and $p = 0.629$ for the right exit, supporting the linear-in-$1/D$ description over the primary window. The right-exit interval contains the theoretical slope; the deep-exit estimate is within 5.45% of the barrier slope. The finite-noise diagnostics below quantify the remaining deep-exit offset without changing the prespecified primary fit.

For pointwise visualization, rate errors use the event-count approximation

$$\mathrm{SE}(\widehat{k}_D) \approx \frac{\sqrt{N_D}}{T_D}, \qquad N_D = \sum_m \delta_m, \quad T_D = \sum_m T_m. \tag{32}$$

These intervals describe sampling variation around the effective rate under the specified stopping rule; the slope intervals in Table 5 use the batch bootstrap and are the inferential quantities used in the paper.

B.5 Numerical Robustness and Finite-Noise Sensitivity

Time-step and initialization audits. We rerun both transitions with $\Delta t = 2.5 \times 10^{-4}$. For the deep exit, we additionally replace point initialization by a local harmonic proposal $\mathcal{N}(a_d, D[\nabla^2\Phi(a_d)]^{-1})$, restricted to the source side of the saddle hyperplane, below the saddle energy, and closest to a_d among the three minima. Rejection sampling is repeated for at most 50 rounds before falling back to a_d. This construction tests sensitivity to a thermally dispersed source ensemble without assuming exact quasistationary sampling. Each audit uses the four primary noise levels and eight batches of 150 trajectories per level.

For each audit, we refit the slope and form a normal-approximation interval for $\Delta\beta = \widehat{\beta}_{\mathrm{audit}} - \widehat{\beta}_{\mathrm{primary}}$ from the independently fitted slope standard errors. Table 6 shows that every 95% interval for $\Delta\beta$ lies inside the prespecified equivalence margin $[-0.1\Delta\Phi, 0.1\Delta\Phi]$ and contains zero. Both deep-exit audits also move the point estimate toward the theoretical slope.

Fit-window and finite-noise diagnostics. We fit the lowest-noise four, five, six, and seven points without selecting a window by agreement with theory. The deep-exit slopes are $-1.00256, -1.00804, -1.01373, -1.00506$; the right-exit slopes are $-0.35866, -0.35531, -0.35677, -0.35274$. Their narrow ranges show that the barrier-scale

Table 6: Numerical robustness of the fitted Arrhenius slope. Halving the time step and broadening the deep-well initialization preserve the primary estimate within the prespecified 10%-of-barrier equivalence margin.

Audit	$\widehat{\beta}$	95% interval for β	95% interval for $\Delta\beta$
Deep, half step	-1.02722	$[-1.06149, -0.99246]$	$[-0.06860, 0.01928]$
Deep, harmonic init.	-1.04756	$[-1.08580, -1.00667]$	$[-0.09303, 0.00305]$
Right, half step	-0.35934	$[-0.37112, -0.34761]$	$[-0.01451, 0.01316]$

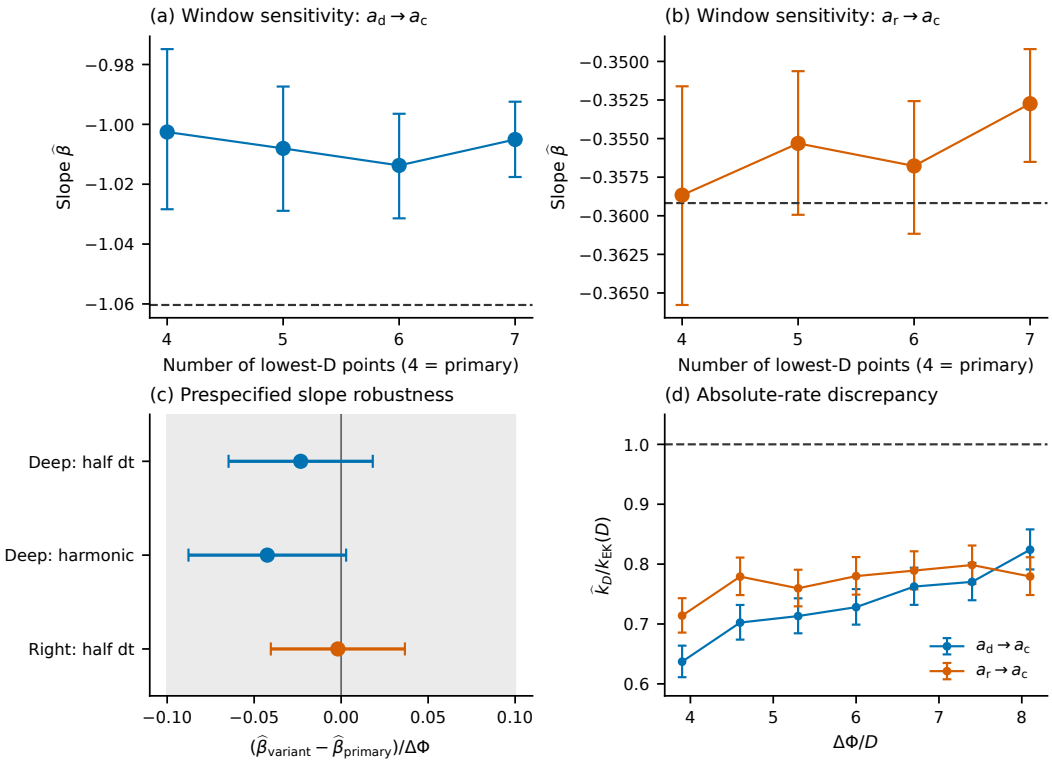

Figure 4: Complete diagnostics for free-time escape scaling. Left: fitted slopes under four nested low-noise windows, with horizontal lines at $-\Delta\Phi$. Center: primary and audit slope differences, normalized by the corresponding barrier; the shaded band is the prespecified $\pm 0.1\Delta\Phi$ equivalence region. Right: empirical-to-Eyring–Kramers rate ratios over the seven noise levels. The barrier slopes are stable under window, time-step, and initialization changes, whereas the absolute rates show a consistent finite-noise prefactor offset.

conclusion is stable to the fit window, and neither sequence is monotone toward the theoretical value.

As a curvature diagnostic, we fit

$$\log \widehat{k}_D = \alpha + \beta D^{-1} + cD \tag{33}$$

over all seven noise levels. This gives $\widehat{\beta} = -1.03324$ with 95% interval $[-1.09238, -0.97393]$ for the deep exit and $\widehat{\beta} = -0.37371$ with interval $[-0.38998, -0.35739]$ for the right exit. The deep interval now contains $-\Delta\Phi_{dc}$, while the right estimate remains close to $-\Delta\Phi_{rc}$. The scaled design matrix has condition number 9.47, so the additional term is numerically identifiable. Across the primary window, empirical-to-Eyring–Kramers rate ratios lie in 0.73–0.82 for the deep exit and 0.78–0.80 for the right exit, indicating that the dominant barrier dependence is recovered more accurately than the absolute finite-noise prefactor.

B.6 Evidence Summary

The two exits provide a stringent within-landscape comparison: their barriers differ by almost a factor of three, yet independently fitted escape-rate slopes recover the corresponding scales with $R^2 > 0.9997$ and relative errors of 5.45% and 0.14%. All prespecified checks of critical-point type, minimum-energy-path resolution, event accounting, fit linearity, time-step stability, and initialization stability are satisfied. The right-exit confidence interval contains the theoretical barrier slope. For the deeper exit, both numerical audits move the estimate toward theory while remaining within the prespecified equivalence margin around the primary result.

These results establish the first layer of the paper's decomposition: across two geometrically distinct exits whose barriers differ by almost a factor of three, free-time rates track the corresponding saddle-barrier scales. The experiment isolates the escape scale and its numerical stability; the fixed-time route ordering and normalized next-state law are tested separately in Sections 4.2 and 4.3.

C Fixed-Time Path Probabilities: Moving-Tube Protocol and Evidence

This appendix gives the equal-barrier construction, independent moving-tube calculation, small-radius extrapolation, and six-candidate numerical audits underlying the fixed-time results in Section 4.2.

C.1 Equal-Barrier Two-Channel Landscape

The fixed-time benchmark uses the analytic two-dimensional landscape

$$\Phi(x,y) = \frac{1}{4}(x^2-1)^2 + \frac{\beta}{2}y^2 + h\exp\left(-\frac{x^2}{2\sigma_x^2}\right) q(y), \tag{34}$$

$$q(y) = 1 - A_u \exp\left[-\frac{(y-y_u)^2}{2s_u^2}\right] - A_\ell \exp\left[-\frac{(y-y_\ell)^2}{2s_\ell^2}\right]. \tag{35}$$

We set $\beta = .35$, $h = .45$, $\sigma_x = .22$, $(y_u, y_\ell) = (.55, -.55)$, $(s_u, s_\ell) = (.10, .28)$, $A_\ell = .85$, and determine $A_u = .8564413124$ by a one-dimensional root solve that equalizes the two saddle energies. This balancing uses only deterministic landscape geometry, not tube probabilities. A bounded critical-point search followed by Hessian classification finds exactly two minima and two index-one saddles in the search box $[-2,2] \times [-1.8, 1.8]$. Gradient-flow descent from both sides of each saddle connects the same two minima. Table 7 gives the quantities used by the frozen experiment.

Equation (34) is analytic and its drift is dissipative. With $L(x,y) = 1 + x^2 + y^2$, the leading terms in $\mathcal{L}_D L$ are $-2x^4 - 2\beta y^2$, while the localized Gaussian perturbation contributes only bounded lower-order terms. Consequently, $\mathcal{L}_D L \le c_D L$ for some $c_D < \infty$, so the diffusion satisfies Assumption 1 despite its superlinear drift. The same compact-set localization used in Appendix A.1 applies to every candidate path and tube.

Table 7: **Equal-barrier geometry and fixed physical scales.** The positive Hessian eigenvalue gives the stable-direction curvature at each saddle.

Object	x	y	Φ or $\Delta\Phi$	Hessian eigenvalues
Source a	-1.000133	-0.000036	0.00001285	$(0.34993, 2.00601)$
Target b	1.000133	-0.000036	0.00001285	$(0.34993, 2.00601)$
Upper saddle	0	0.544978	0.36686671	$(-2.34100, 38.71092)$
Lower saddle	0	-0.512882	0.36686671	$(-2.46376, 5.10116)$
Fixed scales	$\tau = 2.857681$, $T = 3\tau = 8.573044$		$D = .32\Delta\Phi = .11739735$	

The equal barriers deliberately remove a FW route preference. At the same time, the ratio of positive saddle eigenvalues is 7.5887, creating a strong contrast in the potential Laplacian along the upper and lower channels (Fig. 5). The construction therefore isolates the divergence correction in the full finite-D action.

C.2 Smooth Candidate Construction and Basis Audit

Writing normalized time as $s = t/T$, every optimized path has the form

$$\gamma_M(s) = a + (b-a)s + \sum_{n=1}^{M} \frac{c_n}{n^2}\sin(n\pi s), \qquad c_n \in \mathbb{R}^2. \tag{36}$$

The n^{-2} scaling controls high-frequency displacement while preserving an unconstrained coefficient for every sine mode. Endpoints are imposed exactly, and position and physical-time velocity are differentiated from the same representation. We optimize the full action in Eq. (4) and the quadratic-only objective

$$S^T_{\mathrm{quad},D}[\gamma] = \frac{1}{4D}\int_0^T \|\dot\gamma(t) + \nabla\Phi(\gamma(t))\|^2 dt. \tag{37}$$

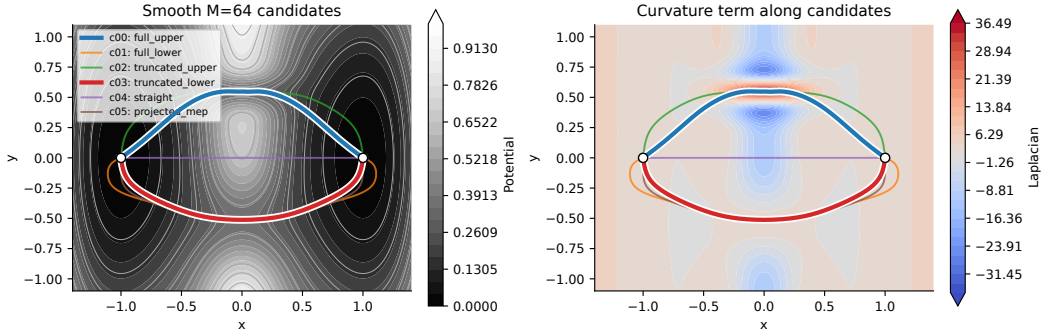

Figure 5: **Fixed-time candidate geometry.** Left: the six $M = 64$ smooth candidates on the equal-barrier potential. Right: the same paths over the Laplacian field that enters the full action through $-\frac{1}{2}\int_0^T \Delta\Phi(\gamma(t))dt$. Identical barrier heights coexist with a pronounced contraction contrast along the two channels.

For $M \in \{48, 64, 80\}$, both objectives are optimized by L-BFGS from the same six initializations spanning the two channels and the straight path. The optimization uses a 64-panel, eighth-order composite Gauss rule; every solution is rescored with a 128-panel, twelfth-order rule. All 36 optimizations satisfy a kinetically whitened gradient threshold of 5×10^{-4} and a quadrature discrepancy below 10^{-4} nats, reach both channel families, and retain the same objective winner under common-pool rescoring.

The $M = 64$ representation defines the scientific candidate set, and the $M = 80$ representation provides a nested resolution audit. The full-action winner remains in the upper channel, its action changes by 0.0372 nats, and its maximum path displacement is $0.00679\|b - a\|$. Modes above 64 contribute 0.00179 of its velocity energy. The truncated winner remains in the lower channel with action change below 5×10^{-14} nats. We retain six distinct smooth paths: the full-action upper and lower solutions ($c00, c01$), truncated-action upper and lower solutions ($c02, c03$), the exact straight path ($c04$), and a smooth projection of the saddle MEP ($c05$).

C.3 Independent Moving-Tube PDE

For a fixed candidate γ, define the offset process $Y_t = X_t - \gamma(t)$. Its drift is

$$b_\gamma(t, y) = -\nabla\Phi(\gamma(t) + y) - \dot{\gamma}(t). \tag{38}$$

If $u_\gamma(t, y)$ is the probability of surviving inside the radius ε tube until T, then

$$-\partial_t u_\gamma = D\Delta_y u_\gamma + b_\gamma(t, y) \cdot \nabla_y u_\gamma, \qquad \|y\| < \varepsilon, \tag{39}$$

$$u_\gamma(t, y) = 0, \qquad \|y\| = \varepsilon, \tag{40}$$

$$u_\gamma(T, y) = 1, \qquad \|y\| < \varepsilon, \tag{41}$$

and $Z_\varepsilon(\gamma) = u_\gamma(0, 0)$. Scaling $y = \varepsilon z$ makes the domain the fixed unit disk. We use odd Cartesian embedding grids so that $z = 0$ is an exact node, impose the true circular boundary through Shortley–Weller distances, and integrate the sparse method-of-lines system with an implicit BDF solver. A scalar transform removes the exact principal Brownian decay and is added back to $\log Z$ analytically. Second-order nonuniform central differences are accepted only when the combined operator has nonnegative off-diagonals, nonpositive row sums and strict boundary killing. No OM action is called by the PDE solver.

The calibration hierarchy in Table 8 expands with the dimensionless horizon $q = DT/\varepsilon^2$: first zero-drift Brownian survival, then constant drift, then the long-horizon envelopes required by the decreasing-radius experiment. The references are independent Bessel-series laws on the disk. Sign and rotation controls agree to at least 8×10^{-12} nats through $q = 158$ and to 10^{-10} nats at $q = 280$. Every moving-tube cell also passes the killed- generator condition; the largest observed cell Péclet number is 0.0511, below the frozen 0.25 ceiling.

Figure 6 visualizes the diffusion-core row; the drifted and long-horizon rows use the corresponding independent spectral law before the nonlinear moving-tube solver is admitted to the radius ladder.

Table 8: **Independent PDE calibration envelope.** Entries are the worst absolute log-probability errors and two-finest-grid changes within each stage; parentheses give the prespecified gate.

Calibration regime	Tested $(q, \lvert\alpha\rvert)$	Analytic error	Grid change
Brownian disk	$q = 1, 10, 40;\ \alpha = 0$	0.02010 (0.05)	0.01545 (0.03)
Constant drift	$q = 10, 40;\ \lvert\alpha\rvert \le 6$	0.04923 (0.12)	0.03816 (0.10)
Long horizon	$(70, 4), (101, 3.5), (158, 3)$	0.06981 (0.12)	0.03071 (0.10)
Radius extension	$(280, 2.25)$	0.10658 (0.12)	0.04685 (0.10)

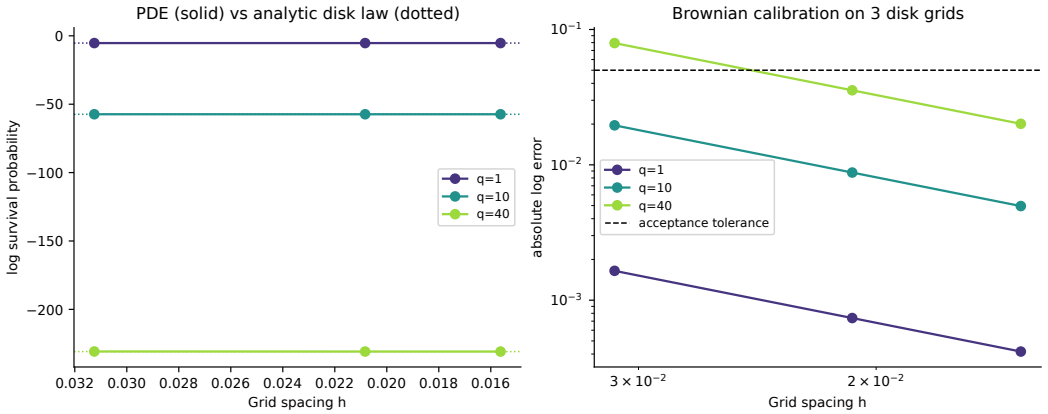

Figure 6: **Absorbing-disk diffusion-core calibration.** PDE log survival probabilities converge to the analytic Brownian disk law at $q \in \{1, 10, 40\}$ (left). Absolute log error contracts across the 65, 97 and 129 embedding grids and lies below the stage-specific gate on the finest grid (right).

Table 9: **Moving-tube radius ladder.** Ratios use the 129 embedding grid and the straight reference. δ_{grid} is the maximum paired grid change over $c00, c03, c04$ at that radius.

r	ε	q	$R_\varepsilon(c00)$	$R_\varepsilon(c03)$	δ_{grid}
.08	.160021	39.304	60.1962	17.0692	.00442
.06	.120016	69.874	85.4976	17.4380	.01036
.05	.100013	100.619	99.5318	17.5811	.01608
.04	.080011	157.217	113.3804	17.6973	.02268
.03	.060008	279.496	125.8722	17.7870	.02923

C.4 Radius Ladder and Zero-Radius Extrapolation

Let $R_\varepsilon(c) = \log Z_\varepsilon(c) - \log Z_\varepsilon(c04)$. The tube radius is $\varepsilon = r\|b - a\|$. The radius order and calibration envelopes were frozen before solving the new moving-tube cells. Table 9 reports the fine-grid ratios for the two objective winners. The ordering $c00 > c03 > c04$ holds at every radius and on both spatial grids. Although individual log probabilities reach roughly -1.6×10^3 at the smallest radius, paired log ratios remain accurately resolved; this is the rare-event regime for which direct Monte Carlo would be particularly inefficient.

Because the disk and paths are smooth, the prespecified primary extrapolation is linear in ε^2 over the three smallest radii; a quadratic fit in ε^2 over all five radii is the frozen sensitivity analysis. Table 10 and Fig. 7 show that both extrapolations approach the full OM ratios and separate sharply from the truncated ratios. The reported numerical bounds propagate paired grid and strict-repeat changes; they are deterministic sensitivity bounds, not sampling confidence intervals.

Table 10: **Zero-radius ratio audit.** Parentheses contain propagated deterministic numerical bounds on the fitted intercept.

Path	Full OM	Truncated	Linear, smallest 3	Quadratic, all 5
$c00$	144.4342	-0.8369	140.2825 (0.0603)	142.7029 (0.0624)
$c03$	17.9076	8.7963	17.9029 (0.0120)	17.9019 (0.0145)

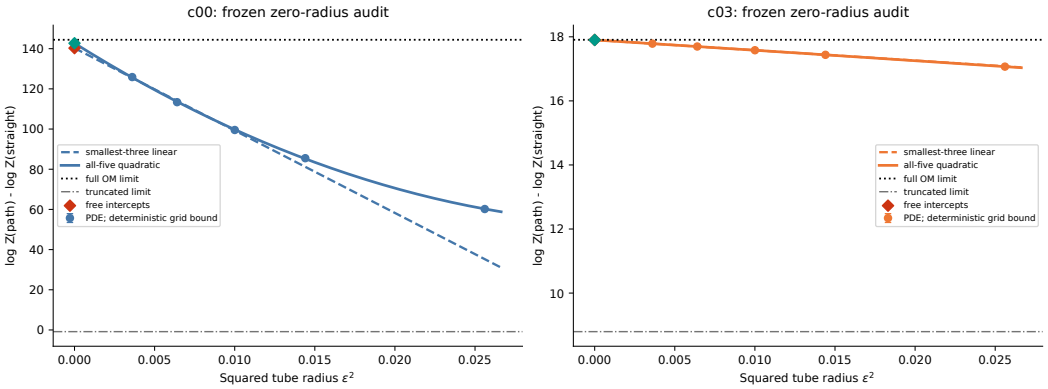

Figure 7: **Frozen small-tube extrapolation.** Fine-grid PDE ratios are shown against ε^2 with deterministic grid bounds. The dashed linear fit uses only the three smallest radii; the solid quadratic sensitivity uses all five. Their unconstrained intercepts (diamonds) converge near the full OM limits (dotted) and remain far from the truncated limits (dash-dot).

Table 11: **Complete six-path comparison at $r = .03$.** PDE entries are strict-solve log tube-probability ratios relative to $c04$; δ_{num} is the maximum of the paired grid and strict-repeat changes.

Path	Construction	PDE ratio	δ_{num}	Full OM	Truncated
$c00$	full, upper	125.8721	.02924	144.4342	$-.8369$
$c02$	truncated, upper	45.3153	.01133	49.2004	8.7485
$c01$	full, lower	18.9508	.00517	19.0933	7.7249
$c03$	truncated, lower	17.7870	.00486	17.9076	8.7963
$c05$	projected MEP	16.8534	.00461	16.9573	8.4189
$c04$	straight	0.0000	$< 10^{-11}$	0.0000	0.0000

C.5 Six-Candidate Confirmation and Numerical Stability

The final stage evaluates every candidate fixed by the smooth-basis stage at $r = .03$. Table 11 reports the strict 129-grid ratios together with the action predictions. The observed ranking

$$c00 > c02 > c01 > c03 > c05 > c04 \tag{42}$$

is unchanged on the 97 primary, 129 primary and 129 strict solves and is identical to the full-action ranking. The truncated action instead predicts $c03 > c02 > c05 > c01 > c04 > c00$. The PDE therefore recovers all 15 pairwise order relations among the six prespecified paths, including the identity of the winning route.

The strict winner–runner-up gap is 80.5568 nats. The frozen decision rule requires it to exceed the larger of 1 nat and twice the sum of the two numerical bounds; the latter is 0.0811 nats. Figure 8 shows that every candidate lies comfortably below both numerical gates. The largest 97-to-129 ratio change is $0.02924 < 0.1$ nats, and the largest primary-to-strict change is $1.57 \times 10^{-4} < 0.05$ nats.

Taken together, the nested-basis, analytic-calibration, radius, and six-candidate audits establish the fixed-time layer within the prespecified candidate family: the full OM action predicts all 15 pairwise order relations, including the winning route that the quadratic-only action reverses. This is a fixed-duration small-tube result; the free-time escape scale and normalized competing-exit law are evaluated in Appendices B and D.

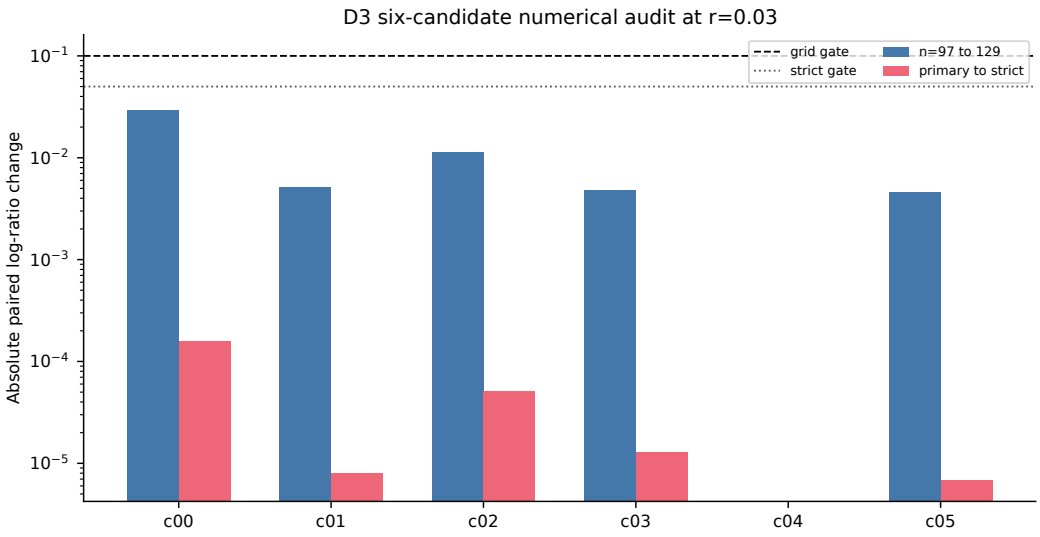

Figure 8: **Six-candidate numerical audit.** Absolute paired changes in log tube-probability ratios under grid refinement (blue) and strict solver repetition (red) are shown on a logarithmic scale. Horizontal lines are the prespecified 0.1- and 0.05-nat gates.

D NORMALIZED NEXT-STATE CHOICE: COMPETING-EXIT PROTOCOL AND EVIDENCE

This appendix gives the frozen landscape, topology audit, approximate quasi-stationary initialization, complete event accounting, inference procedure, and numerical robustness analyses underlying Section 4.3.

D.1 FROZEN LANDSCAPE AND CANDIDATE TOPOLOGY

For $G \in \{G0, G1\}$, the landscape is

$$p_G(x) = \sum_{q \in \{S,E,U,L,R\}} w_q^G \mathcal{N}(x; \mu_q, \Sigma_q), \qquad \Phi_G(x) = -\log p_G(x). \tag{43}$$

Table 12 gives the complete component specification. The intervention changes only the unnormalized U weight from 0.19 to 1.5×0.19 and then renormalizes all weights. Component centers and covariances are consequently identical across landscapes, although the modes of the mixture density can move slightly after reweighting. All predictions and basin assignments use these actual modes rather than the component centers.

Table 12: **Frozen five-component Gaussian mixture.** Covariances are written as $[\Sigma_{11}, \Sigma_{12}; \Sigma_{21}, \Sigma_{22}]$.

State	μ_q	Σ_q	w_q^{G0}	w_q^{G1}
S	$(0, 0)$	$[.1444, 0; 0, .1444]$	.24000	.21918
E	$(1.4, 0)$	$[.1764, 0; 0, .0900]$	.19000	.17352
U	$(-.8, 1.25)$	$[.1252, -.03949; -.03949, .1708]$	.19000	.26027
L	$(-.8, -1.25)$	$[.1444, 0; 0, .1156]$	.19000	.17352
R	$(2.8, 0)$	$[.1444, 0; 0, .1444]$	.19000	.17352

The component means are parameters of the density, whereas the association states are the actual modes of p_G. Their numerically refined coordinates are reported in Table 13; the small but nonzero displacements from the component means are retained in every basin, distance, and barrier calculation.

Table 13: **Actual landscape modes.** Each entry gives the refined minimum (x, y) of $\Phi_G = -\log p_G$.

State	$G0$	$G1$
S	$(0.002253,\ 0.002296)$	$(0.001324,\ 0.003571)$
E	$(1.399541,\ 0.000000)$	$(1.399541,\ 0.000000)$
U	$(-0.799371,\ 1.248969)$	$(-0.799582,\ 1.249315)$
L	$(-0.799556, -1.249445)$	$(-0.799556, -1.249445)$
R	$(2.794725,\ 0.000000)$	$(2.794725,\ 0.000000)$

Before formal sampling, we locate and refine the minima and index-one saddles, verify their Hessian signatures, and trace both descending branches of every saddle. Both landscapes have the same adjacency graph: S is adjacent precisely to E, U and L; R is adjacent to E but not to S. This topological check fixes $\mathcal{C}(S)$ independently of the later stochastic outcomes (Fig. 3). The direct barriers in Table 14 show the intended reversal: E has the lowest barrier in $G0$, whereas U has the lowest barrier in $G1$.

Table 14: **Verified direct saddles and barriers.** The two Hessian eigenvalues have opposite signs for every channel, and $\Delta\Phi_{Sb} = \Phi(s_{Sb}) - \Phi(S)$.

Landscape	Exit	Saddle (x, y)	$\mathrm{eig}(\nabla^2\Phi)$	$\Delta\Phi_{Sb}$
$G0$	E	$(0.698300,\ 0.000019)$	$(-13.0038,\ 9.2209)$	0.899097
$G0$	U	$(-0.410235,\ 0.621678)$	$(-13.8050,\ 8.5474)$	1.073286
$G0$	L	$(-0.386436, -0.673172)$	$(-23.3871,\ 7.1331)$	1.432513
$G1$	E	$(0.698315,\ 0.000029)$	$(-13.0053,\ 9.2206)$	0.900443
$G1$	U	$(-0.375284,\ 0.566692)$	$(-13.5925,\ 8.4235)$	0.850547
$G1$	L	$(-0.386465, -0.673171)$	$(-23.3894,\ 7.1320)$	1.433844

D.2 Approximate Quasi-Stationary Initialization and First-Exit Sampling

We freeze the five shared absolute noise values from the $G0$ minimum direct barrier, so comparisons between $G0$ and $G1$ never use a landscape-specific rescaling. The main Euler–Maruyama step is $\delta t = 8.0838 \times 10^{-5}$; the local relaxation scale is $\tau = 0.151106$. Table 15 reports the complete grid and fixed release horizons. Each horizon is five times the leading $G0$ EK mean residence time at that D and is used unchanged for $G1$.

Table 15: **Shared noise grid and release horizons.**

$\min_b \Delta\Phi_{Sb}^{G0}/D$	D	Horizon	Steps
2.5	0.359639	23.876	295,354
3.2	0.280968	53.442	661,101
3.9	0.230538	117.424	1,452,578
4.6	0.195456	254.137	3,143,775
5.3	0.169641	543.377	6,721,803

For each of the ten landscape–noise cells, five independent Fleming–Viot (FV) populations of 1,600 particles approximate the quasi-stationary distribution (QSD). Each population starts from a local Laplace approximation $\mathcal{N}(x_S, DH_S^{-1})$ restricted to the source basin and evolves for 40τ. Absorbed particles are cloned from survivors. Because members of one FV population are dependent, independent populations—not individual particles—are the statistical units. The prespecified gate compares the 30τ and 40τ laws in whitened means, covariance, one-dimensional CDFs and a two-dimensional histogram. All point estimates remain inside their frozen tolerances. At 40τ, the worst population still has 393 unique ancestors and lineage effective sample size 218, both above the required 100 (Fig. 10). These distributional and genealogical diagnostics qualify the 40τ populations as the approximate QSD ensembles used for release.

Each final FV population initializes 3,200 release trajectories with independent driving noise, giving five release batches per cell. The event is the first grid-observed departure from the deterministic gradient-flow basin of S; it is not nearest-center assignment or arrival inside a target core. A validated source core avoids repeated basin classifications far from the separatrix but does not alter the event definition. A half-core run below tests this acceleration. Table 16 retains direct exits, censoring, failures and R departures separately. Main probability estimates condition only on valid direct exits, but all omitted categories enter the safety criteria.

Table 16: **Complete main-release event counts.** Each row contains 16,000 trajectories. "Cens." denotes right-censoring at the fixed horizon.

Grid	Model	E	U	L	R	Cens.	Fail.
2.5	$G0$	8,012	5,406	2,577	0	5	0
2.5	$G1$	6,378	7,513	2,109	0	0	0
3.2	$G0$	8,802	5,193	2,002	0	3	0
3.2	$G1$	6,469	8,099	1,432	0	0	0
3.9	$G0$	9,697	4,930	1,371	0	2	0
3.9	$G1$	6,660	8,418	922	0	0	0
4.6	$G0$	10,422	4,624	949	0	5	0
4.6	$G1$	6,617	8,724	659	0	0	0
5.3	$G0$	11,037	4,280	681	0	2	0
5.3	$G1$	6,475	9,102	423	0	0	0

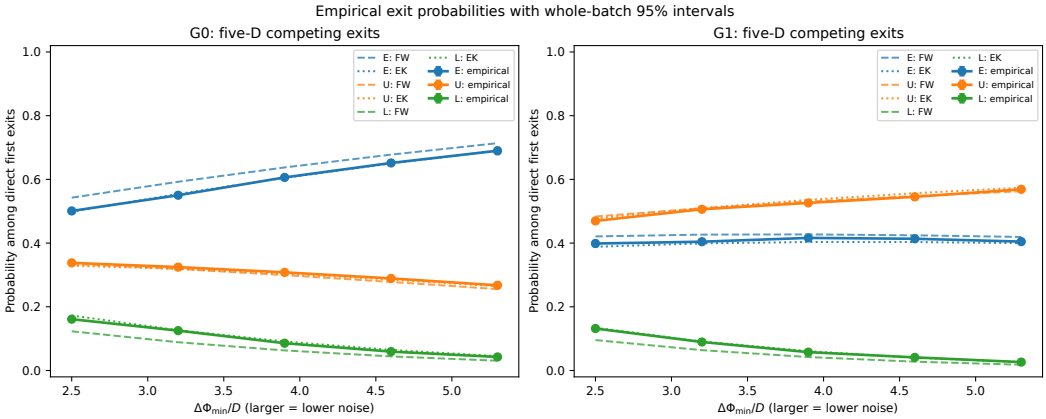

Figure 9: **Complete normalized competing-exit curves.** Empirical probabilities (solid, with whole-batch bootstrap intervals) are compared with prefactor-free Freidlin–Wentzell predictions (FW; dashed) and Eyring–Kramers predictions (EK; dotted) across the common noise grid; lower noise is to the right. The full curves establish that the ranking reversal shown at the middle noise level in Fig. 3 persists and strengthens toward lower noise, while the low-probability L channel remains resolved in every cell.

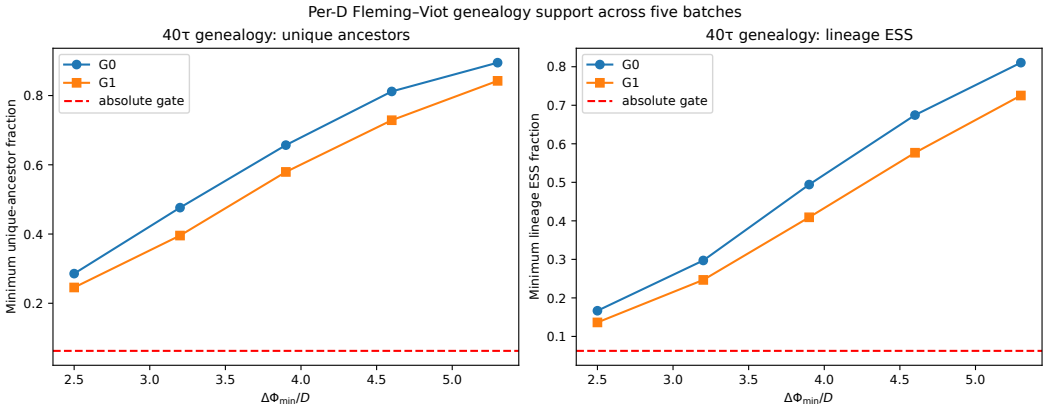

Figure 10: **Approximate-QSD genealogy audit.** Unique-ancestor and lineage-ESS fractions are shown at 40τ for every landscape–noise cell. Horizontal guides mark the absolute support threshold after conversion to counts. The lineage diagnostics retain within-population dependence, with independent FV populations serving as the statistical units.

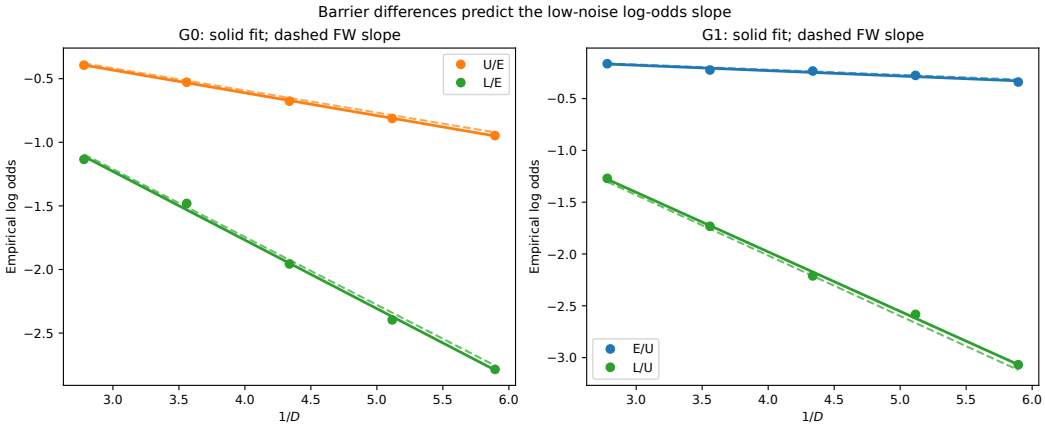

Figure 11: **Complete barrier-difference log-odds analysis.** Empirical log-odds, unconstrained fits, and barrier-difference slopes are shown for both nonreference exits in each landscape. Insets compare fitted and theoretical slopes with whole-batch bootstrap intervals.

D.3 Predictions, Inference and Negative Controls

For a direct gradient-diffusion exit, $V_{\mathrm{dir}}(S, b) = 2\Delta\Phi_{Sb}$. The prefactor-free and curvature-corrected normalized predictions are respectively

$$P_D^{\mathrm{FW}}(b \mid S) \propto \exp[-\Delta\Phi_{Sb}/D], \tag{44}$$

$$P_D^{\mathrm{EK}}(b \mid S) \propto \kappa_{Sb} \exp[-\Delta\Phi_{Sb}/D]. \tag{45}$$

For an isolated direct saddle s_{Sb}, the curvature factor used here is

$$\kappa_{Sb} = \frac{|\lambda_-(s_{Sb})|}{\pi} \sqrt{\frac{\det \nabla^2\Phi(S)}{|\det \nabla^2\Phi(s_{Sb})|}}, \tag{46}$$

with factors summed when several verified saddles lead to the same destination. Neither prediction is fitted to the release counts. For a nonreference exit b and reference exit r, the rate ratio satisfies

$$\log \frac{P_D(b \mid S)}{P_D(r \mid S)} = \log \frac{\kappa_{Sb}}{\kappa_{Sr}} - \frac{\Delta\Phi_{Sb} - \Delta\Phi_{Sr}}{D} + o(1), \tag{47}$$

so the coefficient of $1/D$ is the negative barrier difference; the normalizing denominator cancels exactly. We estimate this coefficient by exit-count inverse-variance weighted regression over all five D values. Uncertainty uses 10,000 whole-batch bootstrap replicates, jointly resampling the complete five-D curve; intervention contrasts pair the matched $G0/G1$ batch slots.

All four theoretical slopes lie inside the corresponding whole-batch bootstrap intervals (Fig. 11), covering both nonreference exits in both landscapes. As D decreases, the FW Jensen–Shannon error declines in both landscapes, with correlations -0.903 and -0.932 against the barrier-to-noise ratio. At the representative middle noise level, the independently computed EK prefactors reduce the Jensen–Shannon divergence from 1.095×10^{-3} to 4.997×10^{-5} in $G0$ and from 6.343×10^{-4} to 9.381×10^{-5} in $G1$. The barrier term therefore controls the log-odds slope, while curvature information sharpens finite-noise relative-rate calibration.

The fixed-center intervention provides a mechanism-level test beyond passive fit. Its five changes in $P_D(U \mid S)$ are 0.132, 0.182, 0.218, 0.256 and 0.301, each with a positive paired interval; their mean is 0.218 with 95% interval $[0.212, 0.222]$. The corresponding FW predictions are 0.149, 0.191, 0.231, 0.270 and 0.307, whereas the fixed component-center prediction is zero at every noise level.

The topology control uses the actual landscape modes m_b and the frozen distance scale

$$\sigma_{\mathrm{dist}} = \mathrm{median}_{b \in \{E,U,L\}} \|m_b^{G0} - m_S^{G0}\|.$$

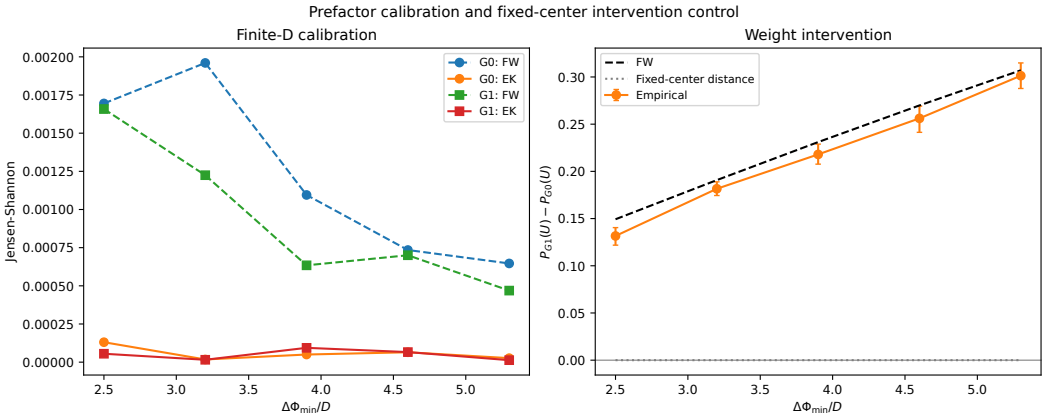

Figure 12: **Finite-noise calibration and intervention.** The top panels compare predictive divergences across D; the bottom panels show the fixed-center U-weight intervention. Barrier predictions capture the direction and strengthening of the effect, while the component-center score remains invariant.

For a candidate set A and $b \in A$, define

$$Q_A(b \mid S) = \frac{\exp[-\|m_b - m_S\|^2/(2\sigma^2_{\text{dist}})]}{\sum_{c\in A} \exp[-\|m_c - m_S\|^2/(2\sigma^2_{\text{dist}})]}. \tag{48}$$

Changing A from the verified direct set $\{E, U, L\}$ to $\{E, U, L, R\}$ leaves the distance scores and scale fixed while assigning probability mass to the nonadjacent state. This inclusion worsens per-event negative log-likelihood by approximately 0.0875 in every landscape–noise cell, directly identifying topology-aware candidate construction as part of the normalized choice law. Figure 12 displays the finite-D calibration and the intervention contrast.

D.4 Robustness of the Competing-Exit Results

At the middle noise level in $G0$, three frozen sensitivity arms each use five batches of 1,600 trajectories. Table 17 reports alternative-minus-main probability shifts. Every point estimate and interval lies within its prespecified practical-equivalence margin; all three arms also retain zero numerical failures and zero departures to R. The competing-exit probabilities are therefore stable to time-step refinement, source-core size, and initialization at the declared numerical resolution.

Table 17: **Competing-exit robustness.** Entries are probability shifts with 95% whole-batch bootstrap intervals.

Alternative	E	U	L	Margin
$\delta t/2$	$-.0043$ $[-.0138, .0056]$	$.0025$ $[-.0054, .0103]$	$.0018$ $[-.0061, .0097]$	.03
Half source core	$-.0095$ $[-.0170, -.0020]$	$.0044$ $[.0012, .0078]$	$.0051$ $[-.0013, .0112]$	.03
Source minimum	$-.0046$ $[-.0150, .0064]$	$-.0015$ $[-.0115, .0088]$	$.0061$ $[.0020, .0105]$	.05

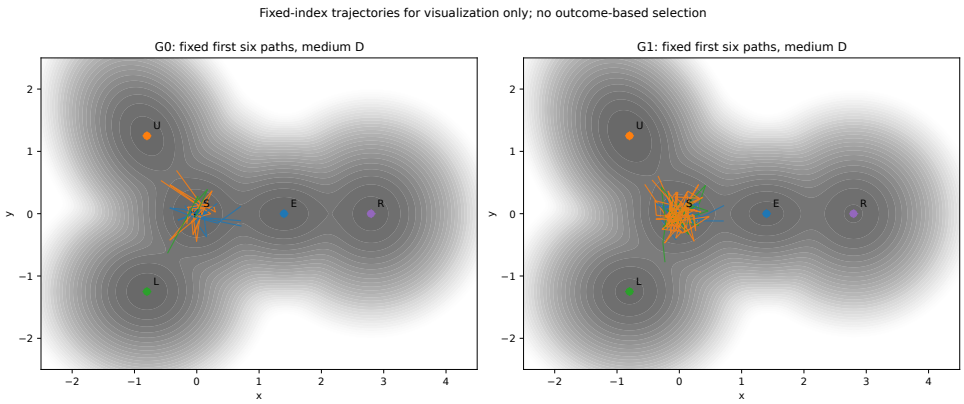

Figure 13: **Representative first-exit trajectories.** Trajectories are plotted on the frozen landscapes to illustrate the event definition and the coexistence of direct channels. Quantitative comparisons use the complete release ensemble reported in Table 16.

Together, the verified topology, approximate-QSD initialization, complete event accounting, log-odds slopes, fixed-center intervention, and numerical audits establish the normalized-choice layer on the controlled five-well landscape. First-exit probabilities follow the barrier-difference scaling in Eq. (12), while the intervention changes the dominant destination through the energy landscape even though every component center and covariance remains fixed.