What if language replaced space in Einstein's equations?

A thought experiment in information geometry, with a worked two-topic toy model

This is a model, not literal physics. The aim is to see how far the structure of general relativity (a metric, a source, curvature, geodesics) can be carried over when "places" are replaced by "states of a text", and where it breaks.

1. The dictionary

General relativity	Language counterpart
Point in spacetime	State of a text: the distribution over what can come next, given everything so far
Metric g	Fisher information metric: two states are close if a reader can barely tell them apart statistically
Mass-energy $T_{\mu\nu}$	Salience σ: usage, emotional charge, cultural centrality
Geodesic	Path of least resistance, roughly the most probable continuation
Gravity well	A mode of the distribution. A dogma or slogan is a black hole: past some point every path leads inward
Proper time τ	Distinct ideas per word. It slows near heavy concepts (long speeches that say one thing)
Light cone	Autoregressive causality: each word depends only on earlier ones
Curvature, holonomy	Untranslatability: carry a meaning around a loop of languages and it returns altered
Cosmological constant Λ	Baseline expansion of the space from neologism and drift

2. A field-equation sketch

Take the Fisher information metric on the space of next-token distributions as the geometry, and let salience play the role of matter. The field equations keep their usual form:

$$G_{\mu\nu} + \Lambda\, g_{\mu\nu} = \kappa\, T_{\mu\nu}, \quad T_{\mu\nu} = \sigma\, u_\mu u_\nu \ (+ \text{ pressure})$$

Here κ says how strongly salience warps the space (plausibly larger for a billboard slogan than for a footnote). A pressure term keeps contested concepts from collapsing into a single reading.

Weak-field limit

Let $\Phi(x) = -\log p(x)$ be the surprisal. Then:

$$d^2x/d\tau^2 = -\nabla\Phi \qquad \nabla^2\Phi = \kappa\sigma \qquad d\tau \approx (1 + \Phi)\, dt$$

Sentences fall toward high-probability regions (the score-function drift familiar from Langevin sampling), heavy concepts sit at the modes of usage, and proper time runs slowest at the heaviest concepts.

Curvature and forgetting

On a space with weighted measure $e^{-\Phi}$, the natural curvature is the Bakry–Émery tensor:

$$\mathrm{Ric}_\Phi = \mathrm{Ric} + \mathrm{Hess}\,\Phi \geq \lambda\, g \quad \Rightarrow \quad \text{convergence to equilibrium at rate } e^{-\lambda t}$$

Positive curvature means conversations forget their starting point quickly and the same few topics keep resurfacing. Flat or negative curvature lets discourse wander far without being pulled back. The stationary solution, a gradient Ricci soliton, is the “equilibrium language” in which usage and geometry are consistent.

The conservation problem

The contracted Bianchi identity forces $\nabla^{\mu} G_{\mu\nu} = 0$, so Einstein’s equations require $\nabla^{\mu} T_{\mu\nu} = 0$. Ideas are not conserved. There are two honest options:

- **Add a speaker field.** Let $\nabla^{\mu} T_{\mu\nu} = J_{\nu}$, with an equal and opposite exchange into a field of speakers and writers. Salience moves between minds and texts, and the total is conserved. This gives agency an explicit place in the physics.
- **Give up pure GR** for a modified gravity, such as f(R) or Einstein–Cartan, in which the source may leak.

3. A two-topic toy model

Put two topics, A and B, at $(-a, 0)$ and $(+a, 0)$ in a plane, and let the language distribution be a mixture of two Gaussians of width s with salience weights w_A and w_B:

$$p(x) = w_A \, \mathrm{N}(x; -a, s^2) + w_B \, \mathrm{N}(x; +a, s^2), \quad \Phi = -\log p$$

In the weak-field limit the geodesic equation has a closed form:

$$\ddot{x} = -\Sigma_k \, r_k(x) \, (x - \mu_k) \, / \, s^2, \quad r_k = w_k \, \mathrm{N}_k \, / \, p$$

Each topic acts as a spring toward its center. The weight of each spring is its *responsibility* r_k, meaning how much that topic currently explains the sentence. Near a topic one responsibility is close to 1 and the motion is a harmonic oscillator with frequency $\omega = 1/s$, so sentences circle the topic in closed rosettes.

Separability and the barrier

Because both Gaussians share the same factor in the perpendicular direction y, the potential separates: $\Phi(x, y) = y^2/(2s^2) + \Phi_x(x)$. Motion in y is a plain harmonic oscillator that never helps a sentence cross between topics. Only the energy along the axis, $E_x = \tfrac{1}{2} v_x^{\,2} + \Phi_x$, matters, and it is conserved. This separability is special to isotropic, equal-width topics.

The saddle between the topics sits close to the midpoint, which gives an approximate barrier out of each well:

$$\Delta\Phi_A \approx a^2/(2s^2) + \ln w_A, \quad \Delta\Phi_B \approx a^2/(2s^2) + \ln w_B, \quad v_{\mathrm{esc}} = \sqrt{(2\Delta\Phi)}$$

The depth difference between the wells is exactly $\ln(w_A/w_B)$: the salience ratio is the well asymmetry. With $s = 0.55$, separation $2a = 2.4$ and $(w_A, w_B) = (0.6, 0.4)$, scanning the axis numerically gives barriers of 1.90 (A) and 1.49 (B), so escape speeds along the axis of 1.95 and 1.73. The approximate formula gives 1.87 and 1.46.

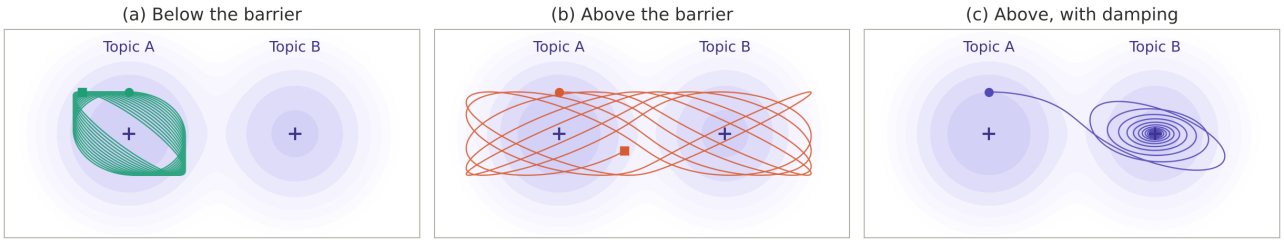


Figure 1. Three regimes for a sentence launched along the axis from topic A at (−1.2, 0.6); circle marks the start, square the end. (a) Axis speed 0.75 v_{esc}: locked in A for good. (b) 1.25 v_{esc}: energy exceeds the saddle, so the sentence wanders between both topics. (c) Same launch with damping $\gamma = 0.15$: it hops, loses energy, and is captured, here by B, the shallower topic. Shading shows Φ; darker is lower surprisal.

What the regimes say

- **Below the barrier** (axis energy under the saddle): the sentence stays in one topic forever. This is the conversation that keeps circling the same subject.
- **Above the barrier:** the sentence moves freely between topics. This is topic-hopping.
- **With damping:** pure geodesics never get captured, because energy is conserved. Capture requires dissipation, and the damping term plays the role of the speaker-field exchange J_ν. The deeper topic captures more sentences, but not all, as panel (c) shows.

Add noise to the damped dynamics and you get Langevin dynamics, whose stationary distribution is $p \propto e^{-\Phi}$. The long-run time share of the two topics is then $w_A : w_B$, and the hopping rate between them is governed by the barrier, roughly $\propto e^{-\Delta\Phi}$ (Kramers). These are standard results for this class of dynamics rather than something simulated here.

4. How it could be tested

- Estimate the Fisher metric from a language model's logits and check whether text trajectories approximate geodesics.
- Measure holonomy directly with round-trip translations and see whether the drift correlates with local curvature.
- Compare "novel ideas per token" against local salience to see whether the dilation law holds. This is the easiest to run and the most likely to fail interestingly.

5. Where the analogy strains

- Language is discrete, so differential geometry applies only approximately.
- The metric is not unique: different communities and contexts give different geometries.
- Salience is not conserved without the speaker field of section 2.
- Spacetime has one time dimension, while language has several overlapping orderings: syntax, narrative, argument, and the reader's own sequence of understanding.
- The toy model uses a flat plane, so all its curvature lives in Φ. Also, Φ is defined only up to an additive constant. Orbits do not depend on it, but the time-dilation law $d\tau \approx (1 + \Phi)dt$ does, so a real test needs a stated reference level for Φ.

Natural next step

An exactly solvable version with real curvature: the Fisher metric on the family of Gaussians (μ, σ) is the hyperbolic plane, with line element $ds^2 = (d\mu^2 + 2d\sigma^2)/\sigma^2$, and its geodesics are semicircles.