Variational Curvature, Predictive Geometry, and the Maintenance of Agency

Flyxion

December 7, 2025

Abstract

This essay explores geometry underlying predictive inference, motivated by the observation that strictly dissipative surprise minimization leads to degenerate equilibria (the dark-room paradox). We argue that agency corresponds to the active maintenance of epistemic curvature against a natural tendency toward geometric flattening. Using information geometry and variational flow, we show that the Expected Free Energy (EFE) functional plays the role of a geometric operator that regularizes curvature and prevents collapse. The resulting framework interprets predictive agency as sustained non-equilibrium geometry.

1 Predictive Inference as Geometry

Predictive processing treats perception, cognition, and action as processes that minimize probabilistic prediction error over time. While this framework has become central within computational neuroscience and machine learning, the geometric structure underlying these minimization principles is less widely recognized. In fact, each internal generative model defines a point on a statistical manifold whose intrinsic curvature determines how beliefs change under evidence, and how actions influence future observations.

In this geometric perspective, inference is not merely numerical optimization over a parameter space but motion through a curved informational geometry. This motion is governed by a variational principle whose differential structure specifies how the agent bends or straightens informational pathways. Consequently, predictive inference is most naturally interpreted as geometric flow rather than simple error correction.

1.1 Curvature and informational sensitivity

Curvature quantifies the rate at which predictive states change as beliefs shift infinitesimally. High curvature corresponds to epistemically sensitive regions where small parameter changes generate large predictive differences; low curvature corresponds to redundant or invariant regimes in which predictions remain stable despite significant variation. Thus curvature is an intrinsic measure of epistemic tension and predictive complexity.

1.2 The geometric role of gradients

Predictive updates follow gradients of a cost functional that measures disagreement between predicted and observed data. However, ordinary gradients are defined with respect to arbitrary coordinates. Because statistical curvature makes many coordinates physically irrelevant, one must instead use the natural gradient, which encodes intrinsic directions of change measured by the Fisher Information Metric. This yields a coordinate-free formulation consistent with the intrinsic informational geometry.

2 Surprise Minimization and Degenerate Agency

The central difficulty with pure surprise minimization is not conceptual but structural: because surprise represents mismatch between predictions and observations, the global minimum corresponds to states in which predictions are maximally certain and observations maximally trivial. Such states include environments of extremely low variability—“dark rooms”—in which sensory data are monotonously predictable.

2.1 Dissipative fixed points

Pure surprise minimization defines a strongly dissipative gradient flow whose fixed points are precisely those states of minimal uncertainty. These fixed points correspond to geometrically flattened regions of the belief manifold: the gradients of surprise vanish, the natural gradient becomes trivial, and policy has no direction along which to evolve. The system collapses to a state of minimal curvature and maximal symmetry.

2.2 Agency collapse

In this configuration, action becomes redundant and agency evaporates. The agent no longer needs to explore because the environment supplies no new information. The geometry of inference degenerates into a trivial manifold. This collapse is not a boundary case: it is structurally preferred by the variational objective if no other terms oppose it.

3 Variational Agency and Exploration

To maintain agency, the agent must resist geometric flattening. This requires an intrinsic incentive to seek states of epistemic tension—those that increase curvature rather than diminish it. In predictive processing this incentive is supplied by the Expected Free Energy (EFE), whose epistemic term encodes an intrinsic value for information gain.

3.1 Exploration as curvature maintenance

The epistemic contribution of the EFE rewards trajectories that traverse regions of high curvature, thereby sustaining predictive complexity and preventing collapse. Exploration is thus reinterpreted as geometric maintenance: an energetic expenditure to preserve broken symmetry in the belief manifold.

3.2 Non–equilibrium structure

Agency is intrinsically non–equilibrium. A system that successfully predicts all sensory data eventually achieves a state in which further prediction requires no change in belief, driving the system to equilibrium. The EFE prevents equilibrium by sustaining nontrivial curvature. Thus intelligence is not the elimination of uncertainty but the maintenance of controlled uncertainty that sustains agency.

4 Information Geometry and Variational Structure

Information geometry studies statistical models as differentiable manifolds, where probability distributions correspond to points on a curved space equipped with a natural Riemannian metric. The intrinsic geometry of this manifold determines which directions correspond to statistically meaningful variation, how belief states evolve under new evidence, and how optimization trajectories should be parameterized so as to remain well-posed independent of coordinate choice.

4.1 Belief manifold and statistical distance

Let $\mathcal{M}$ denote the manifold of admissible beliefs or internal generative models. A point $\theta \in \mathcal{M}$ parametrizes a probability distribution $p_\theta(y)$ over sensory or latent variables. Infinitesimal variation in the model induces an infinitesimal variation in the corresponding distribution, with the statistical distance measured through the Fisher Information Metric,

$$G_{ij}(\theta) = \mathbb{E}_\theta\Big[\partial_i \log p_\theta(y)\, \partial_j \log p_\theta(y)\Big]. \tag{1}$$

This metric defines the intrinsic geometry of $\mathcal{M}$ independent of external coordinates. Under smooth reparametrizations, G transforms covariantly, implying that the shortest path (geodesic) between beliefs is a coordinate-free notion determined by the curvature of the model family.

4.2 Natural gradient and variational flow

Standard gradient descent uses Euclidean geometry on parameter space and is therefore sensitive to arbitrary coordinate choices. The *natural gradient*, introduced by Amari, deforms the descent direction by the inverse metric G^{-1},

$$\tilde{\nabla} f(\theta) := G^{-1}(\theta)\, \nabla f(\theta), \tag{2}$$

so that the update direction points along the steepest descent in the geometry defined by statistical distinguishability. Trajectories of $\tilde{\nabla} f$ coincide with geodesics of decreasing cost, thereby minimizing variation in model description while respecting informational sensitivity.

4.3 Covariant derivatives and gauge-like structure

Because $\mathcal{M}$ is not generally flat, infinitesimal comparison of belief states requires a covariant derivative rather than an ordinary derivative. Differential operators must be compatible with G, ensuring that the geometry encodes which quantities are physically or inferentially invariant. This leads naturally to a geometric calculus similar in spirit to the use of covariant derivatives in curved spacetime, where the distinction between apparent and true invariance is determined by the connection induced by G.

4.4 Curvature and inferential stability

Curvature on $\mathcal{M}$ quantifies how local updates twist, stretch, or distort the space of admissible beliefs. High curvature implies that small parameter variations produce large changes in the induced distributions, representing high epistemic sensitivity; conversely, locally flat regions encode redundant or over-specified beliefs. The geometric tension associated with curvature underlies the stability of belief evolution and sets the stage for later discussions (Section **??**) regarding the energetic cost of maintaining nontrivial curvature in order to preserve adaptive agency.

5 Variational Principles and Field-Theoretic Interpretation

The modern view of inference treats belief updating as a variational problem: the agent holds a parametric family of internal models and seeks those parameters that best account for data. The basic object is a functional whose minimization encodes the consistency of internal dynamics with external regularities. The resulting Euler–Lagrange structure permits a systematic translation between abstract Bayesian updating and continuous geometric flow.

5.1 Variational characterization of inference

Let θ denote internal model parameters and y observed quantities. A variational functional $\mathcal{F}(\theta; y)$ encodes the discrepancy between model predictions and observed evidence. Inference consists in evolving θ so as to minimize $\mathcal{F}$,

$$\dot{\theta} = -\tilde{\nabla}\mathcal{F}(\theta), \tag{3}$$

where the natural gradient $\tilde{\nabla}$ induces geometric descent. This formalism treats inference as continuous time evolution on $\mathcal{M}$, allowing interpretation as a dissipative flow that tends toward local equilibria.

5.2 From variational calculus to field theory

When beliefs are distributed over a structured state space—spatial, temporal, or hierarchical—one introduces fields

$$S(x,t), \qquad v(x,t),$$

representing local surprise and local policy or flow. The variational functional then becomes a field action,

$$\mathcal{A}[S,v] = \int \Big(\mathcal{L}(S, \nabla S, v, \nabla v)\Big)\, dx\, dt, \tag{4}$$

whose Euler–Lagrange equations define coupled partial differential equations governing the joint dynamics of prediction and action. This field-theoretic description generalizes pointwise inference to spatially extended systems, mirroring constructions in continuum mechanics.

5.3 Dissipation and irreversibility

The resulting PDEs are typically parabolic and strictly dissipative, implying that evolution is time-oriented: information accumulates, and the system tends toward states whose variation is increasingly constrained by previously assimilated evidence. This irreversibility parallels entropy increase in thermodynamics, although here the monotonic quantity is informational rather than thermodynamic entropy.

5.4 Steady-state equilibria

As the flow relaxes, the system may reach fixed points where $\dot{\theta} = 0$ or $\dot{S} = 0$, depending on the representation. These equilibria represent beliefs that optimally capture the statistical structure of observations under the assumed model class. Crucially, the geometric nature of the functional determines the qualitative character of these equilibria, including whether nontrivial active policies survive or vanish. Later sections (Sections **??**–14) examine conditions under which equilibria become degenerate.

6 Natural Gradient, Optimality, and Covariance

In statistical geometry, the ordinary Euclidean gradient fails to respect the intrinsic structure of the belief manifold. Parameter increments of identical numerical magnitude can produce radically different changes in the underlying probability distributions. The natural gradient corrects this imbalance by measuring displacement in terms of the intrinsic Fisher metric G, leading to updates that respect the true informational geometry of belief.

6.1 Intrinsic optimality

Let ∇ denote the ordinary gradient and $\tilde{\nabla}$ the natural gradient. Then

$$\tilde{\nabla} f = G^{-1} \nabla f, \tag{5}$$

where G is the Fisher information matrix and G^{-1} its inverse. This transformation is uniquely determined (up to reparametrization) by information geometry. From the standpoint of learning dynamics, it selects the steepest descent direction relative to the underlying manifold, ensuring that inference proceeds along geodesics rather than coordinate-dependent trajectories.

6.2 Covariant derivatives and invariance

While the natural gradient defines the infinitesimal descent direction, covariant derivatives determine how vector fields vary across the manifold. Given a vector field X and a direction Y, the covariant derivative $\nabla_Y X$ measures change relative to the geometry rather than the ambient Euclidean embedding. Its vanishing identifies quantities that are invariant under reparametrization—a critical requirement for formulating physically meaningful laws in curved spaces.

6.3 Metric compatibility

Compatibility of the Levi–Civita connection with the Fisher metric,

$$\nabla G = 0, \tag{6}$$

ensures that parallel transport preserves inner products and distances. This property makes the Levi–Civita connection central to geometric formulations of inference, even though non-metric α–connections will later prove essential for describing irreversible updating.

6.4 Geodesic flows and learning

The geodesics of $(\mathcal{M}, G)$ constitute the least-action paths of belief change under information geometry. Inference trajectories determined by the natural gradient approach these geodesics, implying that locally optimal learning behaves as optimal transport of probability mass along shortest informational paths. This geometric insight links variational inference to optimal control, continuum mechanics, and transport theory.

7 Transport, Bayesian Updating, and alpha-Connections

The statistical geometry of belief is extended here into a dynamical field formulation in which *surprise* plays the role of a scalar potential and *policy* becomes a vector field coupled to its gradient. This interpretation makes explicit the continuous nature of inference and exposes the geometric forces that shape agency.

7.1 Surprise as scalar potential

Let $S : \mathcal{M} \to \mathbb{R}$ denote the surprise field on the belief manifold. For each belief state $\theta \in \mathcal{M}$, the quantity $S(\theta)$ measures the degree of mismatch between predicted and observed outcomes. The gradient ∇S therefore identifies directions of maximal informational tension. States of minimal surprise form local minima of S, analogous to potential wells in classical mechanics.

7.2 Policy as vector field

Let v be a vector field on $\mathcal{M}$ encoding the agent's action (including perception–action cycles). In field variables, one considers the coupling

$$v(\theta) \ \propto \ -\tilde{\nabla} S(\theta), \tag{7}$$

where $\tilde{\nabla}$ is the natural gradient. Thus action is driven by informational tension: the agent moves along directions that reduce surprise with maximal efficiency relative to the intrinsic geometry.

7.3 Dissipative relaxation

In the absence of additional terms, the induced flow

$$\dot{\theta} = -\tilde{\nabla}S(\theta) \tag{8}$$

is strictly dissipative and relaxes toward critical points of S. These critical points correspond to minimal–complexity niches in which surprise is nearly constant. The resulting steady state satisfies $\dot{\theta} = 0$ and $\nabla S = 0$, implying $v = 0$. This is the geometric expression of the so–called *dark–room* degeneracy.

7.4 Predictability and flattening

Because $\tilde{\nabla}S$ measures informational curvature, the collapse to $\nabla S \approx 0$ represents geometric flattening of the belief manifold. Predictability becomes maximal, variation vanishes, and policy loses its driving force. The system arrives at a fixed point of minimal epistemic curvature; agency is extinguished by over–successful certainty.

7.5 Non–equilibrium and curvature

Active agency, by contrast, requires a persistent non–equilibrium regime in which ∇S remains appreciable. Hence the maintenance of agency demands an intrinsic mechanism that prevents complete flattening of the manifold. Subsequent sections will show that this mechanism is geometrically realized by the epistemic term of the Expected Free Energy functional, which acts as a curvature–preserving regularizer sustaining non–zero informational tension.

8 Holonomy, Path Dependence, and Epistemic Memory

Curvature on the belief manifold introduces a characteristic phenomenon known as *holonomy*: parallel transport of informational quantities around a closed loop need not return them to their original configuration. Inference thus possesses a geometric "memory" of the path taken, independent of the initial and final coordinates in parameter space.

8.1 Parallel transport and loops

Let $\gamma : [0, 1] \to \mathcal{M}$ be a closed loop beginning and ending at θ_0. Parallel transport of a tangent vector u_0 along γ defines a new vector u_1 at θ_0. When $\mathcal{M}$ is curved, u_1 generally differs from u_0, with the discrepancy determined by the curvature tensor. This difference encodes information accumulated along the loop that cannot be deduced from endpoints alone.

8.2 Epistemic holonomy in sequential inference

An analogous effect occurs in sequential Bayesian updating. Consider a parameterized model subject to a sequence of evidential updates drawn from distinct data sources. If the updates are performed in two different orders, the resulting posteriors need not coincide, even when the combined evidence is identical. The residual discrepancy—typically in the covariance structure—represents a *geometric phase* induced by the curvature of the belief manifold. Inference thus possesses a path–dependent memory.

8.3 Mixture–model illustration

For example, in learning a mixture of distributions, assimilating samples from components in the sequence $A \to B \to C$ and $C \to B \to A$ can lead to posteriors whose means agree but whose higher–order uncertainty structures differ. Despite identical endpoints in the space of sufficient statistics, the underlying geometry yields distinct effective beliefs.

8.4 Agency as path–dependent structure

Agency derives from this geometric memory. The optimal policy depends not only on the present belief state but also on the informational trajectory by which it was reached. Holonomy thereby shapes future inference and action. The history of exploration—the epistemic path—is encoded in curvature and reappears as structure guiding subsequent behaviour.

9 Quotienting, Nuisance Parameters, and Effective Geometry

Inference problems frequently contain parameters that are not of direct interest but whose presence influences the geometry of the full model. Eliminating such *nuisance parameters* induces a new, effective geometry on the reduced space in which agency actually unfolds.

9.1 Marginalization as geometric quotient

Let the full parameter space split as $\Theta = \{\theta, \phi\}$, where θ are identifiable parameters and ϕ are nuisance variables. Marginalizing ϕ (integrating it out of the posterior) collapses points differing only in ϕ into a single equivalence class. Conceptually, this procedure implements a *quotient* by the orbit generated by transformations of ϕ that leave the likelihood invariant.

9.2 Induced metric via Schur complement

The Fisher information metric on Θ decomposes blockwise as

$$G(\theta, \phi) = \begin{pmatrix} G_{\theta\theta} & G_{\theta\phi} \\ G_{\phi\theta} & G_{\phi\phi} \end{pmatrix}.$$

After marginalization, the effective metric on the reduced manifold $\mathcal{M}_{\text{eff}} = \{\theta\}$ is given by the Schur complement

$$G_{\text{eff}} \;=\; G_{\theta\theta} \;-\; G_{\theta\phi}\, G_{\phi\phi}^{-1}\, G_{\phi\theta}.$$

The second term encodes correlations between θ and ϕ and ensures that the reduced geometry is the natural one induced by the original model.

9.3 Curvature induced by elimination

Even if the full space Θ were flat, the reduced space $\mathcal{M}_{\text{eff}}$ need not be. The induced metric generally has nonvanishing curvature, reflecting information that was implicit in the nuisance directions. Thus eliminating variables can create geometric structure rather than simplify it away.

9.4 Agency on the reduced manifold

Policies operate on the space of identifiable parameters. Accordingly, agency is governed by the geometry of $\mathcal{M}_{\text{eff}}$. The quotient construction shows that hidden symmetries and nuisance variables can shape effective curvature and thereby influence exploration and control, even when those variables are no longer explicit in the state description.

10 Expected Free Energy as Geometric Operator

The Expected Free Energy (EFE) augments the variational free energy with an epistemic incentive that actively resists geometric flattening. From the field–theoretic perspective, the EFE adds a term that penalizes globally flat configurations and ensures that gradients of surprise remain nonzero.

10.1 Surprise–only regime

A system minimizing only the instantaneous surprise

$$S = -\ln p(y|\theta)$$

seeks states of maximal predictability. Because S becomes constant in such states, its gradient vanishes and the policy field v collapses. The surprise–only regime therefore produces the dark–room attractor as a global minimum of predictive curvature.

10.2 Epistemic term as curvature regulator

The EFE is typically decomposed schematically as

$$\mathcal{G} = \underbrace{\mathbb{E}\Big[S\Big]}_{\text{risk}} - \alpha \underbrace{\mathbb{E}\Big[D_{\mathrm{KL}}\Big(q(\theta|y) \,\|\, p(\theta)\Big)\Big]}_{\text{epistemic value}},$$

where the second term measures expected information gain. This epistemic term acts as a curvature regulator: policies that lead to uniform, flat predictive states are penalized, while policies that explore high–uncertainty (high– curvature) regions are rewarded.

10.3 Geometric necessity

The EFE should therefore not be viewed as a technical correction but as a geometric necessity: it implements a covariant operator on the belief manifold that prevents a collapse to maximal symmetry. By sustaining nonzero curvature, the EFE enables persistent agency and continual inference. In this view, intelligent behaviour is the maintenance of a controlled deviation from equilibrium, enforced by the geometry of the variational objective.

11 Policy, Exploration, and Optimal Transport

Learning may be interpreted as optimal transport of probability mass on the belief manifold. At each moment, the agent carries its prior distribution forward to a posterior distribution in light of new evidence. The geometry of this transport is governed by the Fisher Information Metric and the associated natural gradient flow.

11.1 Probability mass transport

Given a prior distribution $p(\theta)$ and a posterior $q(\theta|y)$, Bayesian updating corresponds to transporting probability mass from the prior location to the posterior. This transport is optimal in the sense that it minimizes a variational free–energy cost functional, which upper bounds the instantaneous surprise.

11.2 Diffeomorphic flows

Instantaneous Bayesian updates may be regarded as diffeomorphic flows on the belief manifold: locally smooth, invertible maps that preserve the manifold structure. These flows follow natural–gradient directions, which correspond to geodesics relative to the metric G. Consequently, the agent moves along locally shortest paths in information space, relative to the intrinsic geometry.

11.3 Topology change in long–run learning

Over longer timescales, learning may involve changes in the topology of belief space, such as the merging or splitting of hypothesis components. These transitions cannot be represented by a single global diffeomorphism; rather, they correspond to sequences of local flows interrupted by phase transitions. The result is a dynamic belief topology whose structure reflects the history of evidence and curvature.

12 Entropy, Complexity, and Predictive Cost

Curvature quantifies the informational cost of prediction. High curvature implies a landscape in which small changes in belief produce large changes in expected data distributions. Maintaining coherent predictions in such a landscape requires greater energetic or computational expenditure.

12.1 Curvature = cost of predictability

In a high–curvature environment, the minimal action required to sustain accurate predictions grows with the magnitude of curvature. Conversely, reducing curvature (flattening the manifold) reduces this cost. Surprise–only dynamics therefore tend toward flattening as a form of energetic efficiency.

12.2 Maintenance of low entropy = energy

Maintaining low entropy (high informational structure) requires continuous investment. This principle echoes classical results in statistical mechanics, where maintaining order far from equilibrium requires an energetic flux. Similarly, sustaining nontrivial predictive structure requires an ongoing investment in curvature.

12.3 Adaptive complexity as necessary expenditure

The epistemic component of the EFE provides a variational justification for adaptive complexity: sustaining curvature is instrumentally valuable. The agent does not merely tolerate complexity but actively invests in it to preserve agency and avoid collapse. Adaptive complexity is thus a necessary expenditure rather than a contingent one.

13 Conceptual Synthesis

Taken together, these considerations show that agency, inference, and exploration are inseparable from the geometry of belief. The variational framework, interpreted geometrically, unifies predictive processing, information geometry, and the statistical mechanics of non–equilibrium systems under a single operator: the Expected Free Energy.

14 Resolution of the Dark–Room Problem

The dark–room degeneracy arises because minimizing instantaneous surprise encourages the system to seek perfectly predictable states, which correspond to flat regions of the belief manifold. In such states, the gradients of surprise vanish and the policy field collapses.

The Expected Free Energy resolves this paradox by introducing an epistemic term that rewards exploration and penalizes global flattening. Sustaining nonzero curvature preserves agency and ensures that the system remains sensitive to information and capable of adaptive behaviour.

15 Implications and Open Questions

Agency appears not as a static capacity but as an ongoing act of resisting geometric flattening. This interpretation raises important questions:

- How should we quantify the energetic cost of curvature in biological or artificial systems?
- What invariants characterize sustainable agency?
- Can the balance between exploration and exploitation be formalized as a trade–off between curvature and energy?

Addressing these questions requires integrating statistical mechanics, information geometry, and predictive processing into a coherent theory of intelligent systems.

16 Conclusion

Agency is inseparable from curvature. Minimizing surprise without regard to curvature leads to degeneracy. The Expected Free Energy introduces a geometric operator that sustains non–equilibrium structure, breaks symmetry, and prevents collapse. Agency is therefore the active maintenance of curvature against a natural tendency toward geometric flattening.

References

[1] K. Friston. A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, 360:815–836, 2005.

[2] S.-I. Amari and H. Nagaoka. *Methods of Information Geometry*. American Mathematical Society, 2000.

[3] S.-I. Amari. *Information Geometry and Its Applications*. Springer, 2016.

[4] A. Caticha. Entropic Dynamics, the Schrödinger Equation and the Information Geometry of Entropic Time. *Journal of Physics: Conference Series*, 701, 2016.

[5] C. R. Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.

[6] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[7] M. I. Jordan, D. Prince. Variational inference: a review. *TDA Workshop Notes*, 2015.

[8] T. Amos. Holonomy in Statistical Manifolds. *Entropy*, 20(5):345, 2018.

[9] N. Ay, J. Jost, H. V. Leinster, and W. Wong. *Information Geometry*. Springer, 2017.

[10] C. Maes. The fluctuation theorem as a Gibbs property. *Journal of Statistical Physics*, 95:367–392, 1999.

[11] N. Tishby and N. Zaslavsky. Deep learning and the information bottleneck principle. *IEEE Information Theory Workshop*, 2015.

[12] G. Hinton, P. Dayan, B. Freeman, and R. Neal. The wake–sleep algorithm for unsupervised neural networks. *Science*, 268(5214):1158–1160, 1995.

[13] Y. Ollivier. A Ricci curvature for Markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810–864, 2009.

[14] F. Rigoli, J. Bruineberg. Expectation, free energy, and active inference. *Biological Cybernetics*, 112:6–7, 2018.

[15] J. Kim and D. Van Camp. The Variational Information Bottleneck: beyond VAE. *ICLR*, 2018.