The Inertia of Belief

Robert C. Dennis
cdenn016@gmail.com

January 15, 2026

Abstract

We present a unified geometric framework from which several foundational sociological models emerge as limiting cases: DeGroot social learning, Friedkin-Johnsen opinion dynamics, bounded confidence models, echo chamber formation, and Social Impact Theory all arise from variational free energy minimization on statistical manifolds under appropriate parameter regimes. This unification reveals that consensus, polarization, and bounded confidence are not separate phenomena requiring distinct theories but manifestations of a single information-geometric principle operating under different boundary conditions.

The framework's deeper structure explains why phenomenological mass/spring models of belief dynamics have proven empirically successful: the Fisher information metric, the unique Riemannian metric on probability spaces, naturally plays the role of an inertial mass tensor. Confident beliefs resist change while uncertain beliefs update readily—not by analogy to physics but as a consequence of information geometry. We adopt a Hamiltonian ansatz treating precision as epistemic inertia, which reduces to standard Bayesian free energy descent in the overdamped limit while predicting oscillation, overshooting, and resonance in underdamped regimes consistent with attitude change research. The framework reframes confirmation bias and belief perseverance as geometric consequences of epistemic inertia rather than irrationality.

Keywords: Gauge theory · Active inference · Free energy principle · Information geometry · Sociology

1 Introduction

Why do some beliefs resist change more than others? Why does influence itself seem to harden the minds and poison the empathies of those who wield it? Some beliefs are stiff while others readily sway. While confident beliefs clearly possess more "cognitive inertia" than uncertain ones, a principled mathematical foundation for this intuitive phenomenon remains elusive. Current theories of belief updating, from Bayesian inference Jaynes (2003) to predictive coding Friston (2010); Clark (2013), model belief change as gradient descent. This is a purely dissipative process where beliefs flow toward lower free energy without momentum, inertia, or dynamics. Though enormously successful across neuroscience Friston et al. (2016), psychology Hohwy (2013), and machine learning Millidge et al. (2021), this framework does not account for inertial or oscillatory dynamics.

In this article, we show that beliefs possess an epistemic inertia proportional to an agent's prior precision, observational precision, and their social interactions. Just as physical objects with mass resist acceleration, beliefs held with high confidence resist change and, once moving, tend to continue in their direction on a statistical manifold of beliefs. The second-order expansion of the variational free energy reveals that the Fisher information metric (Amari, 2016) suggests an ansatz for an inertial mass tensor for belief dynamics. The Hamiltonian structure we adopt treats these second-order terms, traditionally neglected (Friston, 2008; Bogacz, 2017), as generating momentum with sociological and psychological consequences.

Furthermore, beliefs propagate through networks of agents in attention patterns ranging from coordinated consensus to turbulent disagreement, often exhibiting distortion, resonance, and phase transitions (Castellano et al., 2009; Galam, 2012). While numerous models capture aspects of this collective evolution, from opinion dynamics (Hegselmann and Krause, 2002a) to quantum-inspired approaches (Busemeyer and Bruza, 2012), a principled geometric foundation has been lacking.

As an intuitive example, consider an agent with strong priors about a political position. When presented with contradicting evidence, their belief doesn't immediately flip but resists change, may overshoot when it does shift, and might oscillate before settling. Conversely, an uncertain agent responds quickly to new evidence with minimal resistance. These phenomena, typically attributed to cognitive biases (Kahneman, 2011), emerge naturally from belief inertia.

In this article we provide three primary contributions to the field:

First, we derive a second-order belief dynamics from a first-principles model, showing that the Fisher metric provides a natural inertial mass tensor $M = \Lambda_{\text{prior}} + \Lambda_{\text{observation}} + \Lambda^{\text{in}}_{\text{social}} + \Lambda^{\text{out}}_{\text{social}}$ combining prior conviction, sensory data, and social interactions (both incoming influence and outgoing recoil). Via pullback geometry on statistical manifolds (Amari, 2016; Nielsen, 2020), we extend the variational free energy to multi-agent systems where social coupling takes the form $D_{\text{KL}}(q_i \| \Omega_{ij} \cdot q_j)$, penalizing disagreement with neighbors after gauge transport into a common frame.

Second, we demonstrate that several foundational models from sociology and opinion dynamics emerge as limiting cases of our unified framework. DeGroot social learning (DeGroot, 1974), Friedkin-Johnsen opinion dynamics (Friedkin and Johnsen, 1990), bounded confidence models (Hegselmann and Krause, 2002a; Deffuant et al., 2000), echo chamber formation, and Social Impact Theory (Latané, 1981) all arise from appropriate parameter regimes and approximations of the same variational free energy functional. This unification reveals that phenomena previously requiring separate theoretical apparatuses, such as consensus formation, polarization, and bounded confidence are manifestations of a single geometric principle operating under different boundary conditions.

Third, we unify documented but theoretically distinct psychological phenomena such as attitude oscillation in persuasion Kaplowitz and Fink (1992); Fink et al. (2002), perceptual overshoot Burge (2010); Webster (2015), and momentum in expectations Coibion and Gorodnichenko (2015) as well as cognitive biases such as confirmation bias Nickerson (1998a) and belief perseverance Anderson et al. (1980a), as natural consequences of epistemic inertia operating in different parameter regimes. These effects, absent in first-order, purely dissipative, treatments Parr et al. (2022), emerge from second-order dynamics and yield testable predictions (such as precision-scaled relaxation times, resonance frequencies, and stopping distances) that distinguish our framework from purely dissipative models.

Our approach thereby opens powerful mathematical tools traditionally relegated to physics such as symplectic geometry Arnold (1989), perturbation theory Holmes (2012), Noether's theorem Olver (1993), renormalization group methods Wilson and Kogut (1975); Goldenfeld (1992), topological phenomena Nakahara (2003); Bernevig and Hughes (2013), and critical point analyses Strogatz (2015); Sornette (2006) for understanding cognitive and social dynamics. By recognizing beliefs as dynamical and inertial quantities, we bridge information geometry, cognitive science, and collective behavior within a unified Hamiltonian framework.

2 Mathematical Framework

2.1 Beliefs as Points on Statistical Manifolds

We model beliefs as probability distributions $q(\xi)$ parameterized by $\xi \in \mathbb{R}^n$ on a statistical manifold $\mathcal{B}$.

For the remainder of this article we shall consider multi-variate Gaussian (MVG) beliefs and priors for convenience.

$$q = \mathcal{N}(\mu_q, \Sigma_q) \tag{1}$$

$$p = \mathcal{N}(\mu_p, \Sigma_p) \tag{2}$$

where μ_ν represents the mean value and Σ_ν represents uncertainty.

The Kullback-Leibler (KL) divergence measures the distance between an agent's belief q and their prior model p

$$D_{\mathrm{KL}}(q\|p) = \int q(x) \log \frac{q(x)}{p(x)} dx \tag{3}$$

2.2 Multi-Agent Belief Geometry

We extend our single-agent framework to networks of interacting cognitive agents via attention. The geometric setting is a principal G-bundle $\pi : P \to \mathcal{C}$ where the base manifold $\mathcal{C}$ encodes agent positions and social network topology, and the gauge group G acts on belief space (see Appendix A for the complete geometric treatment). Each agent i maintains beliefs $q_i = \mathcal{N}(\mu_i, \Sigma_i)$ anchored to a prior $p_i = \mathcal{N}(\bar{\mu}_i, \bar{\Sigma}_i)$, along with an internal reference frame ϕ_i that determines how they encode information.

To compare beliefs across agents, we introduce parallel transport operators as

$$\Omega_{ij} = e^{\phi_i} e^{-\phi_j} \in G \tag{4}$$

where $\phi_i, \phi_j \in \mathfrak{g}$, the Lie algebra of G.

This operator transforms agent j's beliefs into agent i's reference frame as

$$q_j \to \Omega_{ij} \cdot q_j = \mathcal{N}(\Omega_{ij}\mu_j, \Omega_{ij}\Sigma_j\Omega_{ij}^T) \tag{5}$$

The transformed belief can then be compared with agent i's own beliefs via KL divergence:

$$D_{ij} = D_{\mathrm{KL}}(q_i \| \Omega_{ij} \cdot q_j) \tag{6}$$

The gauge group G could, in principle, be any Lie group acting on the statistical manifold such as SO(d) (rotations), GL(d) (general linear transformations), affine transformations, or more exotic choices. Different groups encode different assumptions about how agents' reference frames can differ. For simplicity, we work with $G = \mathrm{SO}(d)$, which preserves the metric structure such that agents agree on how different two beliefs are, even if they disagree on what the beliefs may mean.

As we show later, the flat gauge ($\Omega_{ij} = I$ for all agent pairs) recovers standard consensus models where all agents share a common reference frame. All derivations in this paper, such as the mass formula, classical model limits, and dynamical predictions, hold in the flat gauge limit. The general gauge structure is mathematical scaffolding that could, generally, accommodate reference frame heterogeneity (e.g., cultural or linguistic translation effects) but is not empirically motivated in the present article. We retain it for full generality while emphasizing that the core results do not depend on non-trivial gauge choices but may be extended in suitable situations (see appendix).

2.3 Multi-Agent Free Energy

The total variational free energy for a network of agents balances individual belief maintenance with social consensus pressure (see Appendix B for the complete derivation and Appendix E for the generative model foundation):

$$\mathcal{F}[\{q_i\}, \{\phi_i\}] = \sum_i \underbrace{D_{\mathrm{KL}}(q_i \| p_i)}_{\text{Prior beliefs}} + \sum_{i,j} \underbrace{\beta_{ij} D_{\mathrm{KL}}(q_i \| \Omega_{ij} \cdot q_j)}_{\text{Social alignment}} \tag{7}$$

$$- \sum_i \underbrace{\mathbb{E}_{q_i}[\log p(o_i \mid c_i)]}_{\text{Sensory evidence}} \tag{8}$$

where β_{ij} represents the attention agent i places in agent j's beliefs and we take p_i to be quasi-static. The attention naturally emerges as

$$\beta_{ij} = \frac{\exp(-D_{\mathrm{KL}}(q_i \| \Omega_{ij} \cdot q_j)/\tau)}{\sum_k \exp(-D_{\mathrm{KL}}(q_i \| \Omega_{ik} \cdot q_k)/\tau)} \tag{9}$$

with temperature τ controlling selectivity. This softmax form is not arbitrary but emerges uniquely from maximum entropy principles (Appendix C), and recovers the attention mechanism used in transformer architectures. The forward KL direction in the social term is similarly principled (Appendix D).

2.4 Proper Time as Information-Theoretic Arc Length

A fundamental issue in any dynamical theory of belief is the definition of time. Wall-clock time is unsuitable because cognitive processes operate on different timescales for different agents and contexts. We propose defining proper time as the information-theoretic arc length traversed on the statistical manifold whereby time is taken as "a difference which makes a difference".

For an infinitesimal belief change $d\mu$, the proper time increment is

$$\boxed{d\tau = \sqrt{d\mu^T \Sigma^{-1} d\mu} = \|d\mu\|_{\Sigma^{-1}}} \tag{10}$$

This definition has several appealing properties:

Scale dependence. For a high-precision agent (Σ small, Σ^{-1} large), a small change in μ corresponds to a large proper time such that "time moves fast" in the sense that each update is cognitively significant. For a low-precision agent, the same parametric change corresponds to a small proper time whereby updates are relatively insignificant or slow. A single bit of information is enormous for a simple, small-scale agent but imperceptible for a complex one such as humans.

Information-theoretic interpretation. To second order, both KL directions give the same result

$$\mathrm{KL}(q + dq \| q) \approx \mathrm{KL}(q \| q + dq) \approx \frac{1}{2} d\mu^T \Sigma^{-1} d\mu = \frac{1}{2} d\tau^2 \tag{11}$$

Thus proper time measures accumulated information change. The choice between KL directions is immaterial at this order.

Invariance. Proper time is invariant under reparameterization of the belief space, depending only on the intrinsic geometry of the statistical manifold.

For a trajectory $\mu(t)$ parameterized by some external parameter t, the total proper time elapsed is:

$$\tau = \int \sqrt{\dot{\mu}^T \Sigma^{-1} \dot{\mu}}\, dt \tag{12}$$

This is analogous to proper time in special relativity, where different observers (agents) experience time differently depending on their state. Here, the velocity through belief space determines how quickly proper time accumulates, with high-precision agents experiencing faster proper time for identical parametric motion.

2.5 Hamiltonian Formulation of Belief Dynamics

The proper time metric introduced above equips belief space with a natural notion of kinetic energy: $T = \frac{1}{2}\dot{\mu}^T \mathbf{M} \dot{\mu}$ where $\mathbf{M}$ is derived below. Combined with the free energy as potential, this suggests a Hamiltonian formulation where precision plays the role of inertial mass.

We adopt this as an ansatz (not derived from first principles) motivated by the geometric structure and its empirical consequences.

Under the adiabatic approximation where priors evolve slowly relative to beliefs (see Appendix B for the complete phase space construction), the mass matrix takes a remarkably interpretable form.

2.6 Mass as Fisher Information: The Complete Derivation

Our central result enabling Hamiltonian mechanics on belief space is that the effective cognitive inertia (the resistance to belief change) emerges as the total Fisher information from all sources of constraint. We derive this explicitly from the variational free energy functional.

2.6.1 The Variational Free Energy

The complete variational free energy for a multi-agent system decomposes as

$$F = \underbrace{\sum_i D_{\mathrm{KL}}(q_i \| p_i)}_{\text{complexity}} - \underbrace{\sum_i \mathbb{E}_{q_i}[\log p(o_i|c_i)]}_{\text{accuracy}} + \underbrace{\sum_{i,k} \beta_{ik} D_{\mathrm{KL}}(q_i \| \Omega_{ik} \cdot q_k)}_{\text{consensus}} \tag{13}$$

where q_i is agent i's posterior belief, p_i is agent i's prior, $p(o_i|c_i)$ is the observation likelihood with hidden state c_i, $\Omega_{ik} \cdot q_k$ is neighbor k's belief transported into agent i's reference frame, and β_{ik} is the attention-weighted coupling strength. Denoting the full state vector $\xi = (\mu_1, \ldots, \mu_N, \Sigma_1, \ldots, \Sigma_N)$, the mass matrix is the Hessian of this free energy:

$$\mathbf{M} = \frac{\partial^2 F}{\partial \xi \partial \xi^\top} \tag{14}$$

This Hessian defines the effective mass matrix $\mathbf{M}$ for belief dynamics. Unlike the intrinsic Fisher-Rao metric (which depends only on the belief geometry), this Hessian mass matrix incorporates contributions from priors, observations, and social coupling. Each term in F contributes independently to the total mass.

2.6.2 Contribution 1: Prior Precision

The complexity cost $D_{\mathrm{KL}}(q_i \| p_i)$ penalizes deviation from the prior. Its Hessian with respect to the mean yields

$$\frac{\partial^2}{\partial \mu_i \partial \mu_i^\top} D_{\mathrm{KL}}(q_i \| p_i) = \bar{\Sigma}_{p_i}^{-1} \equiv \bar{\Lambda}_{p_i} \tag{15}$$

This is the prior precision, i.e. resistance to deviating from innate or learned expectations.

2.6.3 Contribution 2: Observation Precision

The accuracy term $-\mathbb{E}_{q_i}[\log p(o_i|c_i)]$ rewards explaining observations. For a Gaussian observation model $p(o_i|c_i) = \mathcal{N}(o_i \mid c_i, R_i)$ where c_i is the hidden cause and R_i is the sensory noise covariance:

$$\frac{\partial^2}{\partial \mu_i \partial \mu_i^\top}\left[-\mathbb{E}_{q_i}[\log p(o_i|c_i)]\right] = R_i^{-1} \equiv \Lambda_{o_i} \tag{16}$$

This is the observation precision—the inverse sensory noise covariance. Precise observations (small R_i, large Λ_{o_i}) provide strong grounding that resists belief change.

2.6.4 Contribution 3: Social Precision

The consensus term $\sum_k \beta_{ik} D_{\mathrm{KL}}(q_i \| \Omega_{ik} \cdot q_k)$ penalizes disagreement with neighbors. Taking the Hessian:

$$\frac{\partial^2}{\partial \mu_i \partial \mu_i^\top} \sum_k \beta_{ik} D_{\mathrm{KL}}(q_i \| \Omega_{ik} \cdot q_k) = \sum_k \beta_{ik} \Omega_{ik} \Sigma_{q_k}^{-1} \Omega_{ik}^\top = \sum_k \beta_{ik} \tilde{\Lambda}_{q_k} \tag{17}$$

where $\tilde{\Lambda}_{q_k} = \Omega_{ik} \Lambda_{q_k} \Omega_{ik}^\top$ is the precision of neighbor k transported into agent i's frame.

Additionally, agent i appears in the consensus terms of its neighbors j, contributing a *reciprocal* mass:

$$\frac{\partial^2}{\partial \mu_i \partial \mu_i^\top} \sum_j \beta_{ji} D_{\mathrm{KL}}(q_j \| \Omega_{ji}[q_i]) = \sum_j \beta_{ji} \Lambda_{q_i} \tag{18}$$

2.6.5 The Complete Mass Formula

Combining all contributions, the effective mass of agent i is:

$$\boxed{M_i = \underbrace{\bar{\Lambda}_{p_i}}_{\substack{\text{prior}\\\text{precision}}} + \underbrace{\Lambda_{o_i}}_{\substack{\text{observation}\\\text{precision}}} + \underbrace{\sum_k \beta_{ik} \tilde{\Lambda}_{q_k}}_{\substack{\text{incoming}\\\text{social precision}}} + \underbrace{\sum_j \beta_{ji} \Lambda_{q_i}}_{\substack{\text{outgoing}\\\text{social precision}}}} \tag{19}$$

This four-part structure has transparent intuitive meaning. The term $\bar{\Lambda}_{p_i}$ represents **prior inertia**, the resistance arising from the cost of deviating from deep expectations. The term Λ_{o_i} represents **sensory inertia**, grounding through observation whereby precise senses anchor beliefs. The term $\sum_k \beta_{ik} \tilde{\Lambda}_{q_k}$ represents **incoming social inertia**, the effect of being pulled toward confident neighbors. Finally, the term $\sum_j \beta_{ji} \Lambda_{q_i}$ represents **outgoing social inertia**, the recoil from exerting influence on others.

2.6.6 Interpretation

The identification of mass with total Fisher information reveals several intuitive phenomena and grounds them geometrically and quantitatively:

Sensory anchoring. Agents with precise observations (Λ_o large) have greater belief inertia. This seems counterintuitive; shouldn't better data make beliefs more flexible? The resolution is that precise observations provide strong evidence for the current state. An agent with low-noise sensors has high Fisher information, meaning small belief changes would dramatically worsen the likelihood fit. The agent is anchored by its own sensory precision. This predicts that experts with reliable instrumentation become harder to move than novices relying on noisy signals; not because the expert is stubborn, but because their high-fidelity observations geometrically constrain the belief manifold. The very precision that makes expertise valuable simultaneously makes it rigid.

Social amplification. The social terms show that an agent's inertia is collective. An agent coupled to confident neighbors inherits their precision as mass via the incoming term $\sum_k \beta_{ik} \tilde{\Lambda}_{q_k}$. A population of high-precision agents becomes collectively rigid, while uncertain agents readily reach consensus. This predicts that expertise clusters resist external perturbation: a group of confident specialists, each attending to the others, forms a mutually reinforcing mass that deflects outside information. Conversely, low-confidence agents coalesce readily around any confident neighbor, explaining how charismatic leaders can rapidly consolidate uncertain populations. The dynamics are asymmetric: confident clusters repel, uncertain populations attract.

Reciprocal costs. The outgoing term $\sum_j \beta_{ji} \Lambda_{q_i}$ reveals that influencing others costs flexibility. An agent that strongly affects its neighbors accumulates inertia from those interactions, becoming less responsive itself. Influence is not free but instead is paid for in epistemic rigidity. The more others attend to your beliefs, the more those beliefs resist change. This has profound implications: leaders become trapped by their followers, gurus ossify under the weight of their disciples' attention, and public figures grow deaf to feedback as their audience grows. Henry Adams observed that "power is poison. Its effect on Presidents has always been tragic" Adams (1918). Our framework provides a geometric mechanism for this tragedy. The very act of projecting influence onto others accumulates as inertial mass, progressively consuming the capacity for empathy and update. The powerful become rigid not through moral failure but through the geometry of attention: when many minds attend to yours, the Fisher information contributed by those outgoing connections makes belief change increasingly costly. This predicts that influence hierarchies naturally produce epistemic stratification, with those at the top most resistant to information from below.

Important Caveat: The Hamiltonian as Ansatz

The preceding development identifies the Fisher information metric with an inertial mass tensor. This identification is natural and geometrically principled. The Fisher metric is the unique (up to scaling) Riemannian metric on statistical manifolds. However, we emphasize that the second-order (Hamiltonian) dynamics are an ansatz, not a derivation.

The Fisher metric provides geometry; a notion of distance and curvature on belief space. It does not by itself imply that beliefs evolve according to Hamiltonian mechanics with this metric as mass. We adopt this ansatz because it naturally identifies precision with inertial resistance to belief change; it predicts phenomena (oscillation, overshooting, resonance) observed empirically in attitude change research; it reduces to well-justified gradient flow in the overdamped limit $\gamma \to \infty$; and the derivations of classical sociological models (Section 4) rely *only* on this overdamped limit, not on the full Hamiltonian structure.

The empirical adequacy of the underdamped regime remains to be established. The framework's predictive value lies partly in making this distinction precise: overdamped predictions are geometrically necessary, while underdamped predictions depend on the ansatz.

3 Results

3.1 Cognitive Phenomena from Belief Momentum

The Hamiltonian formulation introduces a quantity absent from standard treatments of Bayesian belief updating we refer to as epistemic momentum. Just as physical momentum allows objects to flow past equilibrium, epistemic momentum allows beliefs to overshoot, oscillate, and resist change in ways that pure gradient descent fundamentally cannot capture.

Definition 1 (Cognitive Momentum)**.** *The cognitive momentum of agent i is the product of epistemic mass and belief velocity*

$$\boxed{\pi_i = M_i\dot{\mu}_i = \left(\bar{\Lambda}_{p_i} + \Lambda_{o_i} + \sum_k \beta_{ik}\tilde{\Lambda}_{q_k} + \sum_j \beta_{ji}\Lambda_{q_i}\right)\dot{\mu}_i} \tag{20}$$

where $\dot{\mu}_i$ is the rate of belief change.

The full theory includes conjugate momentum Π_i^{Σ} for covariance dynamics (see Appendix B), but mean dynamics capture the primary phenomenology.

For an isolated agent with isotropic uncertainty $\Sigma_i = \sigma_i^2 I$, this simplifies to

$$\pi_i = \frac{1}{\sigma_i^2}\dot{\mu}_i = \Lambda_i\dot{\mu}_i \tag{21}$$

Momentum is not simply the velocity of belief. A confident agent (high Λ) moving slowly has the same momentum as an uncertain agent (low Λ) moving quickly. This asymmetry has interesting consequences for belief dynamics.

Table 1: Components of cognitive momentum and their psychological interpretations.

Component	**Formula**	**Psychological Role**
Bare momentum	$\bar{\Lambda}_{p_i}\dot{\mu}_i$	Inertia from prior expectations
Sensory momentum	$\Lambda_{o_i}\dot{\mu}_i$	Inertia from observation precision
Social momentum	$\sum_k \beta_{ik}\tilde{\Lambda}_{q_k}\dot{\mu}_i$	Inertia from social embedding
Recoil momentum	$\sum_j \beta_{ji}\Lambda_{q_i}\dot{\mu}_i$	Inertia from influencing others

3.2 Belief Overshoot and Perseverance

Presently, research treats belief perseverance as a flaw in information processing. Epistemic momentum yields an alternative perspective: perseverance is the natural dynamical consequence of beliefs possessing inertia.

Consider an agent moving toward a conclusion with velocity $\dot{\mu}$ who encounters a sudden shift in evidence (a new equilibrium). An agent lacking inertia would instantly reverse course whereas an inertial agent possesses a ”kinetic energy” that must be converted into ”potential energy” before they can stop and reverse.

In the low-damping limit, energy conservation on the quadratic free energy landscape dictates that the initial kinetic energy converts to potential energy at the point of maximum overshoot:

$$\frac{1}{2}M_i|\dot{\mu}_i|^2 = \frac{1}{2}K_i|d_{\text{overshoot}}|^2 \tag{22}$$

where K_i is the stiffness (curvature) of the free energy well provided by the new evidence.

Solving for the overshoot distance:

$$\boxed{d_{\text{overshoot}} = |\dot{\mu}_i|\sqrt{\frac{M_i}{K_i}}} \tag{23}$$

This result differs from constant-force braking models. It implies that for a given velocity and evidence strength (K), the overshoot scales with the square root of the epistemic mass (precision).

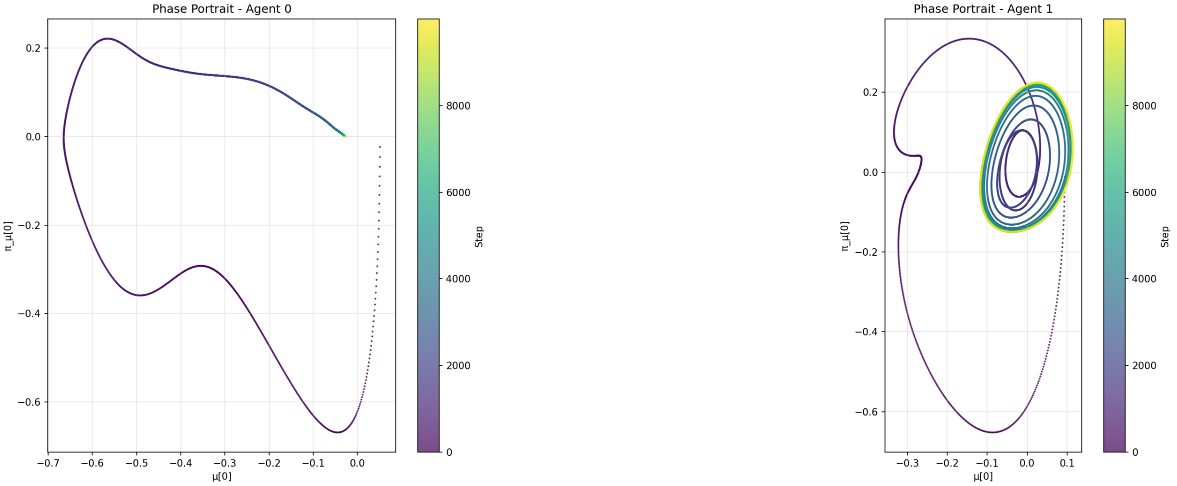
This provides a geometric derivation for why high-precision priors ($M \approx \bar{\Lambda}_p$) lead to ”stubbornness.” A person with deep convictions (high M) moving through a line of reasoning possesses enormous epistemic momentum. When they encounter a contradiction (a force K), they do not stop immediately. Instead, they ”coast” past the logical stopping point, requiring a larger counter-force or longer time to arrest their belief trajectory.This leads to a testable quantitative prediction: the ratio of overshoot distances for high-precision (Λ_H) versus low-precision (Λ_L) agents scales as the square root of their precisions:

$$\frac{d_H}{d_L} = \sqrt{\frac{\Lambda_H}{\Lambda_L}} \tag{24}$$

While less dramatic than a linear scaling, this square-root relationship is a rigid constraint imposed by the harmonic geometry of Gaussian belief spaces.

3.3 Belief Oscillation

Another prediction of our Hamiltonian epistemic dynamics is oscillation phenomena. Unlike gradient descent, which monotonically approaches equilibrium, Hamiltonian systems can overshoot, oscillate, and decay (see Figure 1)



(a) Overdamped ($\gamma > 2\sqrt{KM}$)

(b) Underdamped ($\gamma < 2\sqrt{KM}$)

Figure 1: Phase portraits of belief dynamics. (a) Overdamped regime: beliefs decay monotonically to equilibrium. (b) Underdamped regime: beliefs exhibit oscillatory dynamics. Both simulations run from $t = 0$ to $t = 250$ over 10,000 steps.

3.3.1 The Damped Epistemic Oscillator

The damping coefficient γ is not a new free parameter but inherits directly from standard variational inference. To see this, note that in the overdamped limit where $M\ddot{\mu} \ll \gamma\dot{\mu}$, the equation of motion $M\ddot{\mu} + \gamma\dot{\mu} + \nabla F = 0$ reduces to $\gamma\dot{\mu} \approx -\nabla F$, yielding $\dot{\mu} = -\gamma^{-1}\nabla F$. This is precisely gradient descent on the free energy, $\dot{q} = -\nabla_q F$, with learning rate $\eta = \gamma^{-1}$. Thus γ is simply the inverse learning rate from standard variational inference. The underdamped regime ($\gamma < 2\sqrt{KM}$) corresponds to fast learning where inertial effects become significant, introducing dynamics not present in first-order treatments.

$$M_i\ddot{\mu}_i + \gamma_i\dot{\mu}_i + \nabla_{\mu_i} F = 0 \tag{25}$$

where $\gamma_i > 0$ is a damping coefficient. This equation, from the physics perspective, is the well-known damped harmonic oscillator.

For small displacements from equilibrium μ^* we have

$$M_i\ddot{\delta\mu} + \gamma_i\dot{\delta\mu} + K_i\delta\mu = 0 \tag{26}$$

where $K_i = \nabla^2 F|_{\mu^*}$ represents the belief's "stiffness" (curvature of free energy at equilibrium, completely analogous to a spring).

Once again we arrive at a quantifiable prediction:

In the sub-critical ($\gamma_i < 2\sqrt{K_i M_i}$) regime, beliefs will oscillate around equilibrium with a frequency and decay time given by

$$\boxed{\omega = \sqrt{\frac{K_i}{M_i} - \frac{\gamma_i^2}{4M_i^2}} \approx \sqrt{\frac{\text{Evidence strength}}{\text{Epistemic mass}}}} \tag{27}$$

$$\tau = \frac{2M_i}{\gamma_i} \tag{28}$$

3.3.2 Three Dynamical Regimes

As the standard physics of oscillators show, the discriminant $\Delta = \gamma_i^2 - 4K_i M_i$ manifestly determines different behaviors. In the over-damped regime ($\Delta > 0$), beliefs decay to equilibrium monotonically without oscillation, resembling standard Bayesian updating in the literature. In the critically damped regime ($\Delta = 0$), the system exhibits the fastest approach to equilibrium without oscillation, suggesting this may be optimal for rapid learning. In the under-damped regime ($\Delta < 0$), beliefs oscillate around the equilibrium value, overshooting periodically before equilibrating, producing distinctly non-standard Bayesian dynamics.

As an intuitive example, consider an agent with strong prior beliefs and low receptivity to counter-evidence. When confronted with strong contradictory evidence, the agent will generally exhibit initial resistance as the high mass $M = \Lambda$ resists the force of evidence, followed by acceleration as persistent evidence eventually accelerates belief change, then overshoot as momentum carries belief past the truth, then oscillation as belief swings between acceptance and rejection, and finally settling as damping eventually brings convergence to equilibrium.

This pattern (resist, over-correct, oscillate) is consistent with phenomena documented in attitude change and belief correction research Eagly and Chaiken (1993); Lewandowsky et al. (2012) but remains unexplained by standard Bayesian models. Here we find a natural and intuitive account.

3.4 Resonant Response to Periodic Evidence

Driven oscillatory systems exhibit resonance: a well-characterized phenomenon in physics and engineering where response amplitude peaks at a specific driving frequency. This mechanical effect, not to be confused with metaphorical uses of "resonance", makes precise predictions for belief dynamics.

Periodic evidence driving achieves maximum belief change at the agents belief resonance frequency given by

$$\boxed{\omega_{\text{res}} = \sqrt{\frac{K_i}{M_i}} = \sqrt{\frac{\text{Evidence strength}}{\text{Epistemic mass}}}} \tag{29}$$

3.4.1 Amplitude at Resonance

For example, with sinusoidal forcing $f(t) = f_0 \cos(\omega t)$, the steady-state amplitude is shown (in physics/engineering) to be

$$A(\omega) = \frac{f_0/M_i}{\sqrt{(\omega_0^2 - \omega^2)^2 + (\gamma\omega/M_i)^2}} \tag{30}$$

where $\omega_0 = \sqrt{K/M}$ is the system's "natural" frequency.

At resonance ($\omega = \omega_{\text{res}} \approx \omega_0$) then, we have

$$A_{\max} = \frac{f_0}{\gamma_i \sqrt{K_i/M_i}} = \frac{f_0}{\gamma_i}\sqrt{\frac{M_i}{K_i}} \tag{31}$$

Curiously this implies that high-mass (confident) agents have larger resonance amplitudes rather than smaller. While they resist off-resonance forcing, properly timed evidence produces dramatic swings. This prediction then offers myriad applications in psychological/sociological fields (education, advertising, negotiating, therapy, etc).

3.5 Belief Perseverance in Social Embeddings

The characteristic time for a belief to relax toward equilibrium in a social setting is given by

$$\boxed{\tau = \frac{M_i}{\gamma_i} = \frac{\bar{\Lambda}_{p_i} + \Lambda_{o_i} + \sum_k \beta_{ik}\tilde{\Lambda}_{q_k} + \sum_j \beta_{ji}\Lambda_{q_i}}{\gamma_i}} \tag{32}$$

High-precision beliefs have long decay times. This suggests phenomena where agents tend to hold onto beliefs even after thorough debunking and evidence to their contrary.

For example, if agent A has precision $\Lambda_A = 10$ and agent B has $\Lambda_B = 1$ (both with equal damping γ), then

$$\frac{\tau_A}{\tau_B} = \frac{\Lambda_A}{\Lambda_B} = 10 \tag{33}$$

Agent A's false beliefs persist ten times longer than that of B's, despite identical evidence exposure.

3.5.1 The Debunking Problem

Typically debunking assumes beliefs respond instantaneously to evidence yet our theory of epistemic momentum predicts that immediate debunking is ineffective. The belief should flow past the correction target. Furthermore, if debunking occurs at intervals matching the belief's natural oscillation period, resonance could amplify rather than dampen the response.

A candidate method for debunking, then, is to properly time the belief trajectory before reinforcing the correction. However, predicting that time scale for a given agent may be difficult.

3.6 Sociology and Multi-Agent Momentum Transfer

When agents interact through the attention free energy (β_{ij} term), momentum can transfer between beliefs, i.e. one agent's beliefs affects another's. This suggests a system of coupled equations of motion given an attention pattern of a multi-agent system.

3.6.1 Coupled Equations of Motion

The full multi-agent dynamics with damping are

$$\boxed{M_i\ddot{\mu}_i + \gamma_i\dot{\mu}_i + \nabla_{\mu_i}F = 0} \tag{34}$$

We may expand the gradient as

$$M_i\ddot{\mu}_i = -\gamma_i\dot{\mu}_i - \bar{\Lambda}_{p_i}(\mu_i - \bar{\mu}_i) - \sum_k \beta_{ik}\tilde{\Lambda}_{q_k}(\mu_i - \tilde{\mu}_k) - \sum_j \beta_{ji}\Lambda_{q_i}\Omega_{ji}^T(\tilde{\mu}_i^{(j)} - \mu_j) \tag{35}$$

Then this can be written as

$$\boxed{\underbrace{M_i\ddot{\mu}_i}_{\text{Inertia}} = - \underbrace{\gamma_i\dot{\mu}_i}_{\text{Damping}} - \underbrace{\nabla_{\mu_i}F_{\text{prior}}}_{\text{Prior force}} - \underbrace{\nabla_{\mu_i}F_{\text{consensus}}}_{\text{Social force}}} \tag{36}$$

Dynamic attention. The expansion above assumes fixed attention weights β_{ij}. When attention is dynamic, the gradient includes an additional term from the product rule:

$$\nabla_{\mu_i}F_{\text{social}} = \sum_j \beta_{ij}\nabla_{\mu_i}D_{\text{KL}_{ij}} + \sum_j D_{\text{KL}_{ij}}\nabla_{\mu_i}\beta_{ij} \tag{37}$$

The second term represents attention reallocation: agents are pulled toward neighbors whose attention would increase most upon approach. This amplifies homophily beyond the direct KL force, accelerating polarization when attention is sharply selective (low κ). Most derivations in this paper assume fixed attention, neglecting this effect. The full product rule becomes significant when modeling adaptive attention such as in active learning, persuasion dynamics, or computational implementations where attention weights are optimized. We reserve these terms to future investigation.

3.6.2 Epistemic Momentum Transfer

When agent k changes belief, it transfers epistemic momentum to agent i according to

$$\left.\frac{d\pi_i}{dt}\right|_{\text{from } k} = -\beta_{ik}\tilde{\Lambda}_{q_k}(\mu_i - \tilde{\mu}_k) - \beta_{ki}\Lambda_{q_i}\Omega_{ki}^T(\tilde{\mu}_k^{(i)} - \mu_i) \tag{38}$$

The total momentum transfer over a given interaction time scale $[0, T]$ is

$$\boxed{\Delta\pi_i = -\int_0^T \Big[\beta_{ik}\tilde{\Lambda}_{q_k}(\mu_i - \tilde{\mu}_k) + \beta_{ki}\Lambda_{q_i}\Omega_{ki}^T(\tilde{\mu}_k^{(i)} - \mu_i)\Big]\, dt} \tag{39}$$

3.6.3 Conservation and Non-Conservation

Without priors and damping, the total momentum is a conserved quantity.

$$\frac{d}{dt}\sum_i \pi_i = 0 \quad \text{(closed system)} \tag{40}$$

In contrast, with priors and damping, momentum is assuredly not conserved. Momentum flows into the environment (the prior) and is then dissipated

$$\frac{d}{dt}\sum_i \pi_i = -\sum_i \gamma_i\dot{\mu}_i - \sum_i \bar{\Lambda}_{p_i}(\mu_i - \bar{\mu}_i) \tag{41}$$

This allows us to define a momentum current from agent k to agent i as

$$J_{k\to i} = \beta_{ik}\tilde{\Lambda}_{q_k}(\tilde{\mu}_k - \mu_i) \tag{42}$$

This satisfies the continuity equation

$$\dot{\pi}_i + \gamma_i\dot{\mu}_i + \bar{\Lambda}_{p_i}(\mu_i - \bar{\mu}_i) = \sum_k J_{k\to i} \tag{43}$$

We find that momentum flows from agents with different beliefs via attention β_{ik} and sender precision Λ_{q_k}. High-precision agents are powerful momentum sources as their motion strongly affects coupled neighbors. However, their strength is weighted by their relative attentions β_{ij}

3.7 Summary

Our epistemic momentum framework unifies seemingly disparate phenomena such as confirmation bias, belief perseverance, oscillation, and social influence into manifestations of a single underlying epistemic Hamiltonian mechanics. Beliefs are not just updated, they are dynamic. Evidence does not instantly change minds but rather applies an epistemic force. Finally, confident beliefs don't only resist change rather, they possess epistemic inertia that carries them further than evidence alone would have led them.

4 Classical Sociological Models as Limiting Cases

We now demonstrate that several classical models from sociology and network science emerge from the overdamped limit of our framework. These derivations do *not* depend on the inertial ansatz. We assume the overdamped dynamics follow Riemannian Gradient Flow (Natural Gradient Descent), where the metric tensor determines the path of steepest descent.

Table 2: Testable predictions from cognitive momentum theory.

Phenomenon	Prediction	Experimental Test
Overshoot	Distance $\propto \sqrt{\text{precision}}$	Measure overshoot vs. confidence
Belief oscillation	Under-damped agents overshoot truth and oscillate	Track belief trajectories over time
Resonance	Optimal persuasion occurs at $\omega_{\text{res}} = \sqrt{K/M}$	Vary message timing, measure change
Perseverance	Decay time $\tau = M/\gamma$	Measure false belief persistence vs. uncertainty
Social momentum	High-Λ agents transfer more momentum	Attention vs. source confidence
Recoil	Persuaders become harder to persuade	Measure attitude stiffness after persuasion attempts

4.1 DeGroot Social Learning

4.1.1 Classical Formulation

DeGroot's model (1974) describes social learning as iterative averaging of neighbors' beliefs

$$\mu_i(t+1) = \sum_j w_{ij}\mu_j(t) \tag{44}$$

where $W = [w_{ij}]$ is a row-stochastic matrix ($\sum_j w_{ij} = 1$) representing social influence weights. Under mild conditions, beliefs converge to a consensus determined by the network structure.

4.1.2 Sociological Context

DeGroot's model captures the fundamental sociological insight that individuals update beliefs by averaging opinions from their social network. It has been applied to jury deliberation, scientific consensus formation, and organizational decision-making. However, the model treats influence weights w_{ij} as exogenous and fixed, providing no mechanism for how attention emerges from belief similarity.

4.1.3 Derivation from VFE Framework

Proposition 2 (DeGroot as VFE Limit)**.** *The DeGroot update rule* (44) *emerges from gradient flow on the VFE under*

(i) Overdamped dynamics: $\gamma \to \infty$

(ii) Low uncertainty: $\Sigma_i \to \sigma^2 I$ with σ^2 small

(iii) Flat manifold: $\Omega_{ij} = I$ (shared reference frames)

(iv) No self-coupling: $\alpha = 0$

(v) No observations: $\lambda_{obs} = 0$

(vi) Fixed attention: $\beta_{ij} = w_{ij}$ (constant, not softmax)

Proof. **Step 1: Simplify VFE.** Under these conditions, the free energy reduces to:

$$F[\mu] = \frac{\lambda_\beta}{2\sigma^2} \sum_{i,j} w_{ij} \|\mu_i - \mu_j\|^2 \tag{45}$$

Step 2: Compute gradient.

$$\nabla_{\mu_i} F = \frac{\lambda_\beta}{\sigma^2} \sum_j w_{ij}(\mu_i - \mu_j) \tag{46}$$

$$= \frac{\lambda_\beta}{\sigma^2} \left[\mu_i - \sum_j w_{ij}\mu_j \right] \quad \text{(using row-stochasticity)} \tag{47}$$

Step 3: Apply gradient flow. The mass matrix is $M_i \approx \sigma^{-2} I$. Gradient flow gives:

$$\frac{d\mu_i}{d\tau} = -M_i^{-1} \nabla_{\mu_i} F = -\lambda_\beta \left(\mu_i - \sum_j w_{ij}\mu_j \right) \tag{48}$$

Step 4: Discretize. Forward Euler with $\Delta\tau = 1/\lambda_\beta$:

$$\mu_i(\tau + \Delta\tau) = \sum_j w_{ij}\mu_j(\tau) \tag{49}$$

This is exactly the DeGroot update (44). □ □

4.1.4 What the Unified Framework Adds

Dynamic attention. Removing the fixed-attention assumption and using softmax attention, influence weights become endogenous:

$$\beta_{ij}(\tau) = \frac{\exp(-\|\mu_i(\tau) - \mu_j(\tau)\|^2/(2\sigma^2\kappa))}{\sum_k \exp(-\|\mu_i(\tau) - \mu_k(\tau)\|^2/(2\sigma^2\kappa))} \tag{50}$$

Agents pay more attention to similar others, creating homophily as an emergent property rather than assumption.

Uncertainty dynamics. Relaxing the uniform-Σ assumption, beliefs become full distributions $q_i = \mathcal{N}(\mu_i, \Sigma_i)$. Uncertainty can increase or decrease over time, capturing phenomena like pluralistic ignorance or confidence polarization that mean-only models miss.

Epistemic inertia. When agents receive asymmetric attention (some have many followers), the outgoing social mass term becomes significant. High-attention agents develop higher mass, making their beliefs more resistant to change—a mechanistic explanation for rigidity in positions of authority.

4.2 Friedkin-Johnsen Opinion Dynamics

4.2.1 Classical Formulation

Friedkin and Johnsen (1990) extended DeGroot by introducing "stubbornness" as attachment to initial opinions

$$\mu_i^* = \alpha_i \mu_i(0) + (1 - \alpha_i) \sum_j w_{ij} \mu_j^* \tag{51}$$

where $\alpha_i \in [0, 1]$ represents agent i's resistance to social influence and μ^* denotes equilibrium. This model better captures polarization and persistent disagreement, as stubborn agents prevent full consensus.

4.2.2 Sociological Context

The Friedkin-Johnsen model addresses a key empirical puzzle: social groups often fail to reach consensus despite dense communication. The stubbornness parameter α_i is typically interpreted as a personality trait or ideological commitment. However, this raises the question: what determines stubbornness, and can it change over time?

4.2.3 Derivation from VFE Framework

Proposition 3 (Friedkin-Johnsen as VFE Equilibrium)**.** *The Friedkin-Johnsen equilibrium emerges from the VFE framework under DeGroot conditions plus*

(vii) Non-zero self-coupling: $\alpha > 0$

(viii) Fixed priors: $p_i = \mathcal{N}(\mu_i(0), \Sigma_p)$ (initial beliefs)

Moreover, the stubbornness parameter α_i emerges from prior precision and social context.

Proof. **Step 1: VFE with self-coupling.** Including the prior term:

$$F[\mu] = \frac{\alpha}{2\Sigma_p} \sum_i \|\mu_i - \mu_i(0)\|^2 + \frac{\lambda_\beta}{2\sigma^2} \sum_{i,j} w_{ij} \|\mu_i - \mu_j\|^2 \tag{52}$$

Step 2: Equilibrium condition. At steady state, $\nabla_{\mu_i} F = 0$:

$$\frac{\alpha}{\Sigma_p}(\mu_i^* - \mu_i(0)) + \frac{\lambda_\beta}{\sigma^2} \left(\mu_i^* - \sum_j w_{ij} \mu_j^* \right) = 0 \tag{53}$$

Step 3: Solve for equilibrium.

$$\mu_i^* = \frac{\alpha/\Sigma_p}{\alpha/\Sigma_p + \lambda_\beta/\sigma^2}\mu_i(0) + \frac{\lambda_\beta/\sigma^2}{\alpha/\Sigma_p + \lambda_\beta/\sigma^2}\sum_j w_{ij}\mu_j^* \tag{54}$$

Define the *emergent stubbornness*:

$$\alpha_i' = \frac{\alpha/\Sigma_p}{\alpha/\Sigma_p + \lambda_\beta/\sigma^2} \tag{55}$$

Then $\mu_i^* = \alpha_i'\mu_i(0) + (1 - \alpha_i')\sum_j w_{ij}\mu_j^*$, matching (51). □

4.2.4 Mechanistic Stubbornness

This demonstrates that stubbornness α_i' in (55) is not a fixed personality trait but emerges from two sources:

Prior precision Σ_p^{-1}. Agents with strong initial convictions (low Σ_p) exhibit high stubbornness. This captures ideological commitment or expertise: a climate scientist has high prior precision about global warming, making them resistant to contrarian social influence.

Social coupling strength $\lambda_\beta \sum_j w_{ij}$. Agents experiencing intense social pressure (many influential neighbors) become less stubborn in equilibrium. However, this same pressure increases their epistemic inertia, slowing their rate of approach to equilibrium demonstrating a subtle but important distinction.

The framework predicts that the same individual should exhibit different degrees of stubbornness across different social contexts thereby contradicting trait-based theories that treat resistance to influence as a stable personality characteristic.

4.3 Echo Chambers and Polarization

4.3.1 Phenomenon

Echo chambers constitute a self-reinforcing dynamical process: individuals preferentially attend to similar others (homophily), causing in-group beliefs to converge while out-group beliefs diverge, culminating in isolation as cross-group communication declines.

4.3.2 Derivation from VFE Framework

Proposition 4 (Emergent Homophily and Polarization)**.** *Softmax attention automatically creates homophilic coupling. When initial belief distributions are bimodal, this leads to stable polarized equilibria with within-group consensus and cross-group divergence.*

Proof. **Step 1: Softmax creates homophily.** For Gaussian beliefs with common covariance $\sigma^2 I$:

$$\beta_{ij} = \frac{\exp(-\|\mu_i - \mu_j\|^2/(2\sigma^2\kappa))}{\sum_k \exp(-\|\mu_i - \mu_k\|^2/(2\sigma^2\kappa))} \tag{56}$$

Similar beliefs yield high β_{ij}; dissimilar beliefs yield exponentially suppressed attention.

Step 2: Polarized equilibrium structure. Consider N agents divided into two groups A and B of equal size $N/2$, with group means μ_A and μ_B separated by distance $d = \|\mu_A - \mu_B\|$. At equilibrium, all agents within a group share the group mean.

For agent $i \in A$, define $x = \exp(-d^2/(2\sigma^2\kappa))$. Define the total attention weight allocated to the in-group and out-group as $\beta_{\text{in-group}}$ and $\beta_{\text{out-group}}$. The attention weights are:

$$\beta_{\text{in-group}} = \frac{N/2 - 1}{(N/2-1) + (N/2)x} \tag{57}$$

$$\beta_{\text{out-group}} = \frac{(N/2)x}{(N/2-1) + (N/2)x} \tag{58}$$

Step 3: Stability condition. The gradient flow pulls each agent toward the attention-weighted mean of all others. For the polarized state to be stable, out-group pull must be negligible.

Requiring $\beta_{\text{out-group}} < 1/N$ (out-group attention vanishes as $N \to \infty$):

$$\frac{(N/2)x}{(N/2) + (N/2)x} < \frac{1}{N} \tag{59}$$

For large N with $x \ll 1$, this simplifies to $x < 2/N$, yielding:

$$\exp\left(-\frac{d^2}{2\sigma^2\kappa}\right) < \frac{2}{N} \tag{60}$$

Taking logarithms:

$$\boxed{\|\mu_A - \mu_B\|^2 > 2\sigma^2\kappa \log(N/2) \approx 2\sigma^2\kappa \log N} \tag{61}$$

This is the polarization threshold: groups separated by more than $\sqrt{2\sigma^2\kappa \log N}$ in belief space maintain stable segregation. □

4.3.3 Interpretation

The $\log N$ factor has an intuitive meaning: larger populations require greater belief separation to maintain polarization, because each agent has more potential out-group contacts. However, the dependence is only logarithmic; doubling the population requires only a modest increase in separation.

4.3.4 Phase Transition in Polarization

The stability condition reveals a phase transition in the temperature parameter κ:

High temperature (κ large): Attention is diffuse, cross-group communication persists, system converges to global consensus.

Low temperature (κ small): Attention is sharp, cross-group communication collapses, system locks into polarized state.

The critical temperature scales with initial belief separation

$$\kappa^{\text{crit}} \sim \frac{\|\mu_A(0) - \mu_B(0)\|^2}{2\sigma^2 \log N} \tag{62}$$

4.3.5 Connection to Filter Bubbles

Social media platforms that use engagement-based ranking effectively lower κ (sharpen attention toward similar content). The VFE framework predicts this design choice should increase polarization, consistent with empirical observations. Interventions to reduce polarization should target κ: increasing exposure diversity (raising temperature) or increasing epistemic humility (raising uncertainty σ^2) can prevent the polarization phase transition.

4.4 Bounded Confidence Models

4.4.1 Classical Formulation

Hegselmann-Krause (2002) and Deffuant et al. (2000) introduced bounded confidence: agents only interact with others within a threshold distance ϵ:

$$\mu_i(t+1) = \frac{1}{|N_i(\epsilon)|} \sum_{j \in N_i(\epsilon)} \mu_j(t) \tag{63}$$

where $N_i(\epsilon) = \{j : |\mu_j - \mu_i| < \epsilon\}$. This hard threshold captures the intuition that people ignore opinions too distant from their own.

4.4.2 Derivation from VFE Framework

Proposition 5 (Bounded Confidence as Low-Temperature Limit)**.** *The bounded confidence dynamics emerge from the VFE framework as the attention temperature $\kappa \to 0$.*

Proof. **Step 1: Softmax sharpening.** The attention weight from i to j is:

$$\beta_{ij} = \frac{\exp(-\|\mu_i - \mu_j\|^2/(2\sigma^2\kappa))}{\sum_k \exp(-\|\mu_i - \mu_k\|^2/(2\sigma^2\kappa))} \tag{64}$$

As $\kappa \to 0$, this concentrates on the nearest neighbors.

Step 2: Effective threshold. An agent at distance d receives attention proportional to $\exp(-d^2/(2\sigma^2\kappa))$. Following the echo chamber analysis, attention becomes negligible ($< 1/N$) when:

$$d^2 > 2\sigma^2\kappa \log N \tag{65}$$

Thus the *effective confidence bound* is:

$$\boxed{\epsilon_{\text{eff}} = \sigma\sqrt{2\kappa \log N}} \tag{66}$$

Step 3: Dynamics correspondence. The VFE gradient flow is $\dot{\mu}_i \propto \sum_j \beta_{ij}(\mu_j - \mu_i)$. When κ is small, attention concentrates on agents within ϵ_{eff}:

$$\dot{\mu}_i \approx \frac{1}{|N_i(\epsilon_{\text{eff}})|} \sum_{j \in N_i(\epsilon_{\text{eff}})} (\mu_j - \mu_i) \tag{67}$$

Discretizing with unit time step yields the bounded confidence update. □ □

4.4.3 Key Differences from Classical Models

Soft vs. hard threshold. The VFE framework produces smooth exponential decay rather than a hard cutoff. This is more psychologically realistic as people don't entirely ignore slightly-too-distant opinions but attend to them with diminishing weight.

Adaptive threshold. Unlike fixed-ϵ models, the effective threshold $\epsilon_{\text{eff}} = \sigma\sqrt{2\kappa \log N}$ depends on uncertainty σ (uncertain agents tolerate more distant opinions), temperature κ (higher temperature broadens the acceptance window), and population size N (larger populations have slightly wider effective thresholds).

This predicts that high uncertainty reduces polarization by expanding the range of opinions an agent will engage with.

4.5 Confirmation Bias

4.5.1 Natural Gradient Interpretation

In natural gradient descent, the mass matrix $M_i = \partial^2 F / \partial \mu_i^2$ serves as a metric tensor determining the direction of steepest descent

$$\frac{d\mu_i}{d\tau} = -M_i^{-1} \nabla_{\mu_i} F \tag{68}$$

This is the same object that plays the role of inertial mass in the Hamiltonian formulation. However, interpreting it as a metric rather than a mass requires only that we accept natural gradient descent as the appropriate dynamics on statistical manifolds. This is a standard assumption in information geometry (Amari, 2016) such that here we need not the full inertial ansatz.

4.5.2 Mechanism

Agents with large M_i update more slowly for the same gradient

$$\|d\mu_i\| \propto M_i^{-1} \|\nabla_{\mu_i} F\| \tag{69}$$

Since $M_i = \bar{\Lambda}_{p_i} + \Lambda_{o_i} + \sum_k \beta_{ik} \tilde{\Lambda}_{q_k} + \sum_j \beta_{ji} \Lambda_{q_i}$, high mass (slow updating) arises from strong priors, precise observations, confident neighbors, or many followers.

This is confirmation bias whereby high-precision agents respond less to the same evidence because they sit in steep-walled free energy wells.

4.6 Social Impact Theory

4.6.1 Classical Formulation

Latané's Social Impact Theory (1981) posits that social influence on a target is a multiplicative function of source properties:

$$\text{Impact} = f(\text{Strength} \times \text{Immediacy} \times \text{Number}) \tag{70}$$

4.6.2 Correspondence with VFE Framework

The social force term in our framework provides a quantitative interpretation. The force on agent i from neighbor j is

$$F_{j\to i} = \beta_{ij}\Lambda_{q_j}(\mu_j - \mu_i) \tag{71}$$

Strength ↔ Λ_{q_j}. Source precision (confidence) amplifies the force exerted on the target.

Immediacy ↔ β_{ij}. Attention weights decay with belief dissimilarity, providing an analog of social/epistemic proximity.

Number ↔ $\sum_j$. Total social force sums over all sources.

4.6.3 Caveat

This is an interpretive correspondence, not a formal derivation. Latané's immediacy referred to physical proximity; our β_{ij} captures epistemic proximity (belief similarity). The VFE framework provides one quantitative instantiation of SIT's qualitative principle.

Model	Rigor	Depends on Ansatz?	Notes
DeGroot	Exact	No	Overdamped limit
Friedkin-Johnsen	Exact	No	Equilibrium solution
Echo Chambers	Derived	No	Softmax attention
Bounded Confidence	Approximate	No	Low-κ limit
Confirmation Bias	Geometric	Partially	Mass matrix interpretation
Social Impact Theory	Interpretive	No	Qualitative correspondence

Table 3: Quality assessment of derivations. Core results depend only on gradient flow, not the inertial ansatz.

5 Discussion

5.1 Unifying Phenomenological Models

For decades, researchers across psychology, neuroscience, and opinion dynamics have empirically modeled belief change using spring-mass metaphors without theoretical justification. Kaplowitz and Fink's damped oscillator model of attitude change Kaplowitz et al. (1983); Kaplowitz and Fink (1992), which successfully predicted oscillation and overshoot in response to persuasive messages Fink et al. (2002), treated mass and damping as free parameters which they fit to data. Similarly, bounded confidence models in opinion dynamics Hegselmann and Krause (2002b); Deffuant et al. (2000), momentum effects in economic expectations Coibion and Gorodnichenko (2015), and overshoot phenomena in perceptual adaptation Webster (2015) all invoke inertial dynamics without explaining their origin.

Our central contribution is showing that these are not mere analogies but instead consequences of variational inference on curved statistical manifolds. The Fisher information metric acts as an inertial mass tensor for epistemic dynamics; damping emerges from dissipative terms in the variational principle where the restoring force is the gradient of the variational free energy. This provides a principled basis for parameter values previously treated as phenomenological: epistemic inertia equals $\Lambda_{\text{prior}} + \Lambda_{\text{observation}} + \Lambda_{\text{social}}$, the total precision of an agent's belief; damping γ arises from dissipation due to attention, metabolic costs, or environmental noise; and K corresponds to the curvature of the free energy landscape at equilibrium. The framework thus unifies disparate empirical observations under a single geometric principle: beliefs are points on a Riemannian manifold, and their dynamics are geodesic motion with friction.

5.2 Parameter Regimes and Existing Evidence

The framework predicts that oscillatory dynamics appear only when $\gamma < 2\sqrt{KM}$ (underdamped). Most laboratory learning tasks—designed with high noise and volatile environments—place observers in the overdamped regime where standard gradient descent suffices. This explains why momentum effects are rarely observed in perceptual learning paradigms.

In contrast, Kaplowitz and Fink (Kaplowitz and Fink, 1992) observed attitude oscillation in persuasion studies where participants had strong prior commitments confronting credible counter-evidence—precisely the high-M, low-γ conditions our framework predicts should produce underdamped dynamics. Both phenomena emerge from the same equations in different parameter regimes.

5.3 Why Was This Overlooked?

The connection between precision and inertial mass, despite its naturalness, has remained hidden at the intersection of several disciplines that rarely communicate.

Psychology has historically focused on static biases and heuristics, cataloging the ways beliefs deviate from normative standards rather than the temporal dynamics of how beliefs change Nickerson (1998b). Researchers have been reluctant to ask "how fast does this belief evolve?" with appropriate dynamical tools. When oscillation was observed Kaplowitz et al. (1983); Fink et al. (2002), it remained isolated within communication science, never connecting to the variational principles that would explain why attitudes behave like mechanical systems.

Neuroscience, meanwhile, focuses on gradient-based learning primarily because neural systems are highly damped. Synaptic time constants, metabolic constraints, and homeostatic regulation ensure that neural dynamics operate in the overdamped regime Friston (2008) Bogacz (2017). This has led researchers to overlook oscillatory or momentum-like behavior that might otherwise be visible. The brain appears to do gradient descent because it operates in a regime where inertial effects are suppressed rather than because momentum is absent from the underlying mathematics. It has been there the whole time.

Information geometry, meanwhile, provides the mathematical language for these ideas but was developed largely within statistics and machine learning, far from psychological or sociological theory. The Fisher metric was studied as an abstract structure on probability spaces Amari (2016), not as an inertia tensor governing dynamics.

Sociology has developed sophisticated models of opinion dynamics, social influence, and collective behavior DeGroot (1974); Friedkin and Johnsen (2011); Flache et al. (2017), yet these frameworks typically assume instantaneous averaging or threshold-based contagion rather than momentum-carrying dynamics. The insight that social attention directed toward an agent accumulates as epistemic mass (thereby making the attended-to individual more resistant to belief change) requires bridging network science with variational mechanics. This connection was obscured by the disciplinary boundaries separating formal sociology from the physics-inspired methods that would reveal it.

5.4 Cognitive Biases as Emergent Phenomena

A striking implication of our framework is that phenomena often attributed to cognitive biases emerge naturally from epistemic inertia rather than requiring separate psychological mechanisms.

Belief perseverance is the tendency for beliefs to persist even after their evidential basis has been discredited Anderson et al. (1980b); Ross et al. (1975) follows directly from epistemic mass. High-precision beliefs have large $M = \Lambda$ and thus long relaxation times $\tau = M/\gamma$. They resist change not due to irrationality but because mass resists acceleration. The debriefing paradigm, which demonstrates that beliefs persist after subjects learn the initial evidence was fabricated, is explained by inertia.

The continued influence effect, whereby misinformation continues to affect reasoning even after correction Lewandowsky et al. (2012), similarly reflects momentum decay rather than memory failure. Corrections apply a counter-force, but if the original misinformation was encoded with high precision such that the belief's momentum carries it past the corrected equilibrium before dissipation brings it to rest.

Confirmation bias, the tendency to seek and weight evidence consistent with existing beliefs Nickerson (1998b), can be reinterpreted as a consequence of inertial dynamics. Beliefs with high epistemic inertia are less deflected by contradictory evidence than by confirmatory evidence. This is not a bias in the pejorative sense but a geometrical consequence of how informationally massive objects respond to epistemic forces.

Perhaps most striking is the prediction that influence itself accumulates as inertial mass. The outgoing social term $\sum_j \beta_{ji}\Lambda_{q_i}$ in the mass formula means that agents whose beliefs are attended to by many others become progressively more rigid. This provides a geometric mechanism for observations that leaders, experts, and public figures often exhibit reduced sensitivity to feedback; what Adams called "poison" and which Solzhenitsyn poetically described as "power is a poison well known for thousands of years... for those who are unaware of any higher sphere, it is a deadly poison. For them there is no antidote" Adams (1918); Solzhenitsyn (1973). The phenomenon is not necessarily a moral failure but mathematical inevitability. When many minds attend to yours, the Fisher information contributed by those outgoing connections makes belief change increasingly costly. Empathy, the capacity to update one's model of others, requires precisely the flexibility that influence consumes. This predicts that influence hierarchies naturally produce epistemic stratification, with those at the top most resistant to information from below, and suggests that the "isolation of power" documented by historians and political scientists emerges from the same geometric principles governing individual belief dynamics.

This reframing provides a mechanistic basis for intervention. If belief perseverance stems from high precision rather than stubbornness, then interventions targeting uncertainty (increasing γ/M) may be more effective than those targeting content. Similarly, if the rigidity of leadership emerges from accumulated social attention, then institutional designs that distribute attention more evenly may preserve epistemic flexibility at the top.

5.5 Proposed Experimental Tests

The predictions derived above distinguish the inertial framework from standard first-order Bayesian models. We propose and outline several experiments which could test these predictions. The detailed implementation of these experiments is left for future work.

5.5.1 Belief Oscillation Under Strong Counter-Evidence

Prediction: High-confidence agents should overshoot equilibrium and exhibit non-monotonic belief trajectories when confronted with strong counter-evidence.

Design: Measure participants' prior beliefs and confidence on contentious topics via incentivized elicitation. Present strong, credible counter-evidence and track belief trajectories via repeated measurements (e.g., slider scales at 1-minute intervals over 20 minutes).

Standard Bayesian models predict monotonic convergence toward the posterior. The inertial framework predicts that high-confidence participants may transiently overshoot, briefly adopting positions more extreme than the evidence warrants before settling to equilibrium. Any observed non-monotonicity falsifies purely dissipative models.

5.5.2 Precision-Dependent Relaxation Times

Prediction: Belief relaxation time τ scales linearly with prior precision: $\tau = \Lambda/\gamma$.

Design: Measure prior confidence via betting procedures or confidence intervals. Following exposure to counter-evidence, measure time to reach stable posterior beliefs across participants varying in initial confidence.

The framework predicts that participants with twice the initial precision require twice the relaxation time, independent of the direction or magnitude of belief change. Standard models predict relaxation rates depend on evidence strength rather than prior confidence.

5.5.3 Resonant Persuasion

Prediction: Periodic messaging achieves maximum belief change amplitude at resonance frequency $\omega_{\text{res}} = \sqrt{K/M}$.

Design: Deliver persuasive messages at varying intervals (e.g., every 30 seconds, 2 minutes, 5 minutes, 10 minutes) across conditions, holding total exposure constant. Measure final belief change amplitude.

The framework predicts a non-monotonic relationship with a peak at intermediate frequency determined by participant confidence. Standard models predict monotonic effects of message frequency. Resonance is a signature of second-order dynamics.

5.5.4 Social Attention and Epistemic Rigidity

Prediction: Agents who receive sustained social attention accumulate epistemic mass, becoming more resistant to belief revision than equally confident but unattended individuals.

Design: In a group deliberation paradigm, manipulate attention asymmetry: some participants are designated as ”influencers” whose opinions are solicited and displayed to others, while ”followers” hold equally strong priors but receive no social attention. After exposure to identical counter-evidence, measure belief change magnitude and relaxation dynamics across conditions.

Standard models predict belief change depends on prior confidence and evidence strength, not social role. The inertial framework predicts that attended-to advisors exhibit smaller belief shifts and longer relaxation times than observers with matched initial confidence, due to the social attention term $\sum_j \beta_{ji}\Lambda_{q_i}$ contributing to effective mass.

5.5.5 Momentum in Economic Expectations

Prediction: Professional forecasters with high precision should exhibit overshooting when revising expectations following macroeconomic surprises.

Design: Analyze existing forecast revision data Coibion and Gorodnichenko (2015) following large, unexpected economic announcements (e.g., surprise interest rate changes, employment shocks). Track individual forecaster trajectories over subsequent revision rounds, conditioning on pre-shock forecast confidence.

Standard information models predict gradual, monotonic adjustment toward rational expectations. The inertial framework predicts that high-confidence forecasters may transiently overshoot, revising past the rational benchmark before settling, while low-confidence forecasters adjust monotonically. This pattern, if present in existing panel data, would distinguish momentum-based from friction-based accounts of expectation stickiness.

5.6 Relation to Existing Models

Table 4 summarizes predictions distinguishing the inertial framework from existing approaches.

Table 4: Predictions distinguishing the inertial framework from first-order models.

Phenomenon	**First-Order Models**	**This Framework**
Approach to equilibrium	Monotonic	Can oscillate
Precision dependence	Weights evidence	Determines inertia
Overshooting	Not predicted	Predicted (underdamped)
Resonance to periodic input	Not predicted	Predicted
Belief perseverance	Separate bias	Emerges from mass
Continued influence	Memory failure	Momentum decay
Social momentum transfer	Absent	Predicted
Attention increases rigidity	Not predicted	Predicted (social mass)
Forecast overshooting	Friction/stickiness	Momentum decay

Standard Bayesian updating and its neural implementations (predictive coding, active inference) correspond to first-order dissipative dynamics Friston (2010); Bogacz (2017). The free energy principle emerges as a limiting case: in the overdamped limit ($\gamma \gg 2\sqrt{KM}$), inertial terms become negligible and dynamics reduce to gradient descent on the free energy landscape. Novel predictions arise when damping is sufficiently weak that second-order terms contribute meaningfully.

The DeGroot model of social learning DeGroot (1974) and its extensions assume instantaneous opinion averaging without momentum. Our framework predicts that strongly-held beliefs (high Λ) should resist social influence and potentially induce oscillatory collective dynamics in networks with strong coupling. These phenomena are absent from standard consensus models but consistent with observed polarization dynamics.

5.7 Hierarchical Extensions

The microscopic dynamics developed here (belief equilibration timescales $\tau = M/\gamma$, conditions for oscillatory versus monotonic convergence, momentum transfer through attention networks) provide the foundation for multi-scale theories of collective belief. In forthcoming work, we show that agents achieving sufficient consensus undergo dynamical renormalization-like coarse-graining, yielding emergent informational meta-agents (e.g. institutions, economies, societies) with their own shared beliefs, precisions, and gauge frames. The present framework supplies the dynamics underlying such emergence. The conditions under which individual beliefs synchronize, the timescales of collective equilibration, and the momentum currents that drive or resist consensus formation lead to emergent informational hierarchies.

Critically, our theory makes no distinction between "physical" and "abstract" informational systems. The same equations governing individual belief dynamics may similarly apply to institutional beliefs (corporate strategy, scientific consensus, market expectations), with appropriate identification of the relevant state spaces and attention structures.

5.8 Limitations

Several simplifying assumptions warrant future relaxation.

Gaussian beliefs. While analytically tractable, Gaussian distributions cannot capture the multi-modal posteriors characteristic of hypothesis competition, cognitive dissonance, or attitude ambivalence. Extension to exponential families is developed in Appendix F; extension to arbitrary distributions requires numerical methods or variational approximations.

Quasi-static precision. We have treated precision Λ as slowly-varying relative to the mean μ for demonstrative purposes. The complete theory couples mean and precision dynamics on $\mathbb{R}^d \times \mathrm{SPD}(d)$, allowing precision itself to oscillate or exhibit inertia.

Empirical validation. Direct observation of underdamped belief oscillation, precision-scaled relaxation, or resonant persuasion remains for future experimental work. The Kaplowitz-Fink results Kaplowitz et al. (1983); Fink et al. (2002) provide suggestive evidence, but replication with controlled experiments utilizing the present framework's predictions would strengthen the empirical case.

6 Conclusion

We have proposed that beliefs naturally possess inertia proportional to their precision. The identification of epistemic mass with statistical precision transforms our understanding of belief dynamics and provides new tools that extend beyond dissipative gradient flow into Hamiltonian dynamics.

Our framework predicts oscillations, overshooting, resistance, decay, and resonances in belief dynamics. More fundamentally, it reframes cognitive biases not as irrationality but as consequences of belief inertia. Just as physical mass resists acceleration, cognitive precision resists belief change.

This shift in perspective offers researchers new tools for understanding persuasion, education, therapy, negotiation, and social dynamics. By recognizing that confident beliefs are massive and uncertain beliefs are light, we open new directions for research in socio-psychological dynamics.

The mathematics has been hiding in plain sight for decades, obscured by disciplinary boundaries between differential geometry, physics, and information theory. The Fisher information metric may serve as an inertia tensor for the dynamics of thought.

Data Availability

All code and data will be made publicly available upon publication at
https://github.com/cdenn016/Sociology-Psychology-Epistemic-Inertia

Acknowledgments

Claude Sonnet 4.5 was utilized for programming our variational free energy descent simulation suite. All code was manually reviewed, corrected, and mathematically validated by the author. Furthermore, Claude was utilized for typesetting figures, LaTeX equations, and general organizational and manuscript clarity advice. The author further declares no funding or conflicts of interest.

A Gauge Frame Variations and Pullback Geometry

The Hamiltonian formulation of belief dynamics reflects deep geometric structure. Each agent's belief space carries a gauge freedom: the choice of coordinate frame in which beliefs are expressed. Physical quantities must be invariant under these gauge transformations, while the dynamics must be covariant. This appendix develops the transformation theory for the mass matrix, momenta, and Hamilton's equations under gauge frame variations.

A.1 Gauge Structure of Multi-Agent Belief Systems

A.1.1 The Principal Bundle

The geometric setting is a principal G-bundle $\pi : P \to \mathcal{C}$ where $\mathcal{C}$ denotes the base manifold encoding agent positions and social network topology, $G = \mathrm{SO}(d)$ is the gauge group corresponding to rotations in belief space, and the fiber $\pi^{-1}(c)$ over each point $c \in \mathcal{C}$ is the space of reference frames.

Each agent i located at $c_i \in \mathcal{C}$ expresses beliefs in a local frame. The **transport operator** $\Omega_{ik} \in \mathrm{SO}(d)$ relates agent k's frame to agent i's frame.

A.1.2 Gauge Transformations

A **gauge transformation** is a smooth assignment of group elements to each agent

$$g : \{1, \ldots, N\} \to \mathrm{SO}(d), \quad i \mapsto g_i \tag{72}$$

Under this transformation, belief parameters transform as:

$$\mu_i \mapsto \mu_i' = g_i \mu_i \tag{73}$$

$$\Sigma_i \mapsto \Sigma_i' = g_i \Sigma_i g_i^T \tag{74}$$

$$\Lambda_{q_i} \mapsto \Lambda_{q_i}' = g_i \Lambda_{q_i} g_i^T \tag{75}$$

The transport operators transform as:

$$\Omega_{ik} \mapsto \Omega_{ik}' = g_i \Omega_{ik} g_k^{-1} \tag{76}$$

This ensures that the transported belief $\tilde{q}_k = \Omega_{ik} \cdot q_k$ transforms consistently:

$$\tilde{\mu}_k' = g_i \tilde{\mu}_k, \quad \tilde{\Lambda}_{q_k}' = g_i \tilde{\Lambda}_{q_k} g_i^T \tag{77}$$

A.2 Transformation of the Mass Matrix

A.2.1 Mean Sector

The mean-sector mass matrix transforms as a tensor under gauge transformations.

Diagonal blocks:

$$\begin{aligned}
[\mathbf{M}^\mu]'_{ii} &= \bar{\Lambda}'_{p_i} + \sum_k \beta_{ik}\tilde{\Lambda}'_{q_k} + \sum_j \beta_{ji}\Lambda'_{q_i} \\
&= g_i\bar{\Lambda}_{p_i}g_i^T + \sum_k \beta_{ik}g_i\tilde{\Lambda}_{q_k}g_i^T + \sum_j \beta_{ji}g_i\Lambda_{q_i}g_i^T \\
&= g_i\left[\bar{\Lambda}_{p_i} + \sum_k \beta_{ik}\tilde{\Lambda}_{q_k} + \sum_j \beta_{ji}\Lambda_{q_i}\right]g_i^T \\
&= g_i\,[\mathbf{M}^\mu]_{ii}\,g_i^T
\end{aligned} \tag{78}$$

Off-diagonal blocks:

$$\begin{aligned}
[\mathbf{M}^\mu]'_{ik} &= -\beta_{ik}\Omega'_{ik}\Lambda'_{q_k} - \beta_{ki}\Lambda'_{q_i}(\Omega'_{ki})^T \\
&= -\beta_{ik}(g_i\Omega_{ik}g_k^{-1})(g_k\Lambda_{q_k}g_k^T) - \beta_{ki}(g_i\Lambda_{q_i}g_i^T)(g_k\Omega_{ki}g_i^{-1})^T \\
&= -\beta_{ik}g_i\Omega_{ik}\Lambda_{q_k}g_k^T - \beta_{ki}g_i\Lambda_{q_i}\Omega_{ki}^Tg_k^T \\
&= g_i\,[\mathbf{M}^\mu]_{ik}\,g_k^T
\end{aligned} \tag{79}$$

Block matrix form: Define the block-diagonal gauge matrix:

$$\mathbf{G} = \mathrm{diag}(g_1, g_2, \ldots, g_N) \in \mathrm{SO}(d)^N \tag{80}$$

Then the full mean-sector mass matrix transforms as:

$$\boxed{(\mathbf{M}^\mu)' = \mathbf{G}\,\mathbf{M}^\mu\,\mathbf{G}^T} \tag{81}$$

This is the transformation law for a $(0,2)$-tensor (metric tensor) on the product manifold.

A.2.2 Covariance Sector

The covariance-sector mass involves Kronecker products. Under gauge transformation:

$$\begin{aligned}
[\mathbf{M}^\Sigma]'_{ii} &= \frac{1}{2}(\Lambda'_{q_i}\otimes\Lambda'_{q_i})\cdot\left(1 + \sum_k \beta_{ik} + \sum_j \beta_{ji}\right) \\
&= \frac{1}{2}(g_i\Lambda_{q_i}g_i^T \otimes g_i\Lambda_{q_i}g_i^T)\cdot\left(1 + \sum_k \beta_{ik} + \sum_j \beta_{ji}\right) \\
&= \frac{1}{2}(g_i\otimes g_i)(\Lambda_{q_i}\otimes\Lambda_{q_i})(g_i^T\otimes g_i^T)\cdot\left(1 + \sum_k \beta_{ik} + \sum_j \beta_{ji}\right)
\end{aligned} \tag{82}$$

The transformation law is:

$$\boxed{(\mathbf{M}^\Sigma)' = (\mathbf{G}\otimes\mathbf{G})\,\mathbf{M}^\Sigma\,(\mathbf{G}^T\otimes\mathbf{G}^T)} \tag{83}$$

A.2.3 Cross Blocks

The mean-covariance cross blocks transform as:

$$(\mathbf{C}^{\mu\Sigma})' = \mathbf{G}\,\mathbf{C}^{\mu\Sigma}\,(\mathbf{G}^T \otimes \mathbf{G}^T) \tag{84}$$

A.3 Transformation of Momenta

For Hamilton's equations to be covariant, momenta must transform contragrediently to positions.

A.3.1 Mean Momentum

The mean momentum transforms as a covector:

$$\boxed{(\pi_i^\mu)' = g_i\,\pi_i^\mu} \tag{85}$$

This ensures the pairing $\langle \pi^\mu, \dot{\mu} \rangle$ is gauge-invariant:

$$\langle (\pi^\mu)', \dot{\mu}' \rangle = (g_i \pi_i^\mu)^T (g_i \dot{\mu}_i) = (\pi_i^\mu)^T g_i^T g_i \dot{\mu}_i = (\pi_i^\mu)^T \dot{\mu}_i = \langle \pi^\mu, \dot{\mu} \rangle \tag{86}$$

A.3.2 Covariance Momentum

The covariance momentum $\Pi^\Sigma \in \mathrm{Sym}(d)$ transforms as:

$$\boxed{(\Pi_i^\Sigma)' = g_i\,\Pi_i^\Sigma\,g_i^T} \tag{87}$$

The pairing with $\dot{\Sigma}$ uses the trace:

$$\mathrm{tr}[(\Pi^\Sigma)'\dot{\Sigma}'] = \mathrm{tr}[(g_i \Pi_i^\Sigma g_i^T)(g_i \dot{\Sigma}_i g_i^T)] = \mathrm{tr}[\Pi_i^\Sigma \dot{\Sigma}_i] \tag{88}$$

where we used cyclicity of the trace and $g_i^T g_i = I$.

A.4 Covariance of Hamilton's Equations

A.4.1 Velocity Equation

The velocity equation $\dot{\mu} = (\mathbf{M}^\mu)^{-1}\pi^\mu$ transforms as:

$$\begin{aligned}
\dot{\mu}' &= ((\mathbf{M}^\mu)')^{-1}(\pi^\mu)' \\
&= (\mathbf{G}\mathbf{M}^\mu\mathbf{G}^T)^{-1}\mathbf{G}\pi^\mu \\
&= \mathbf{G}^{-T}(\mathbf{M}^\mu)^{-1}\mathbf{G}^{-1}\mathbf{G}\pi^\mu \\
&= \mathbf{G}(\mathbf{M}^\mu)^{-1}\pi^\mu \quad (\text{since } \mathbf{G}^{-T} = \mathbf{G} \text{ for } \mathrm{SO}(d)) \\
&= \mathbf{G}\dot{\mu}
\end{aligned} \tag{89}$$

This confirms $\dot{\mu}$ transforms as a vector: $\dot{\mu}' = \mathbf{G}\dot{\mu}$.

A.4.2 Force Equation

The force equation involves the free energy gradient. Under gauge transformation:

$$\left(\frac{\partial F}{\partial \mu_i}\right)' = g_i \frac{\partial F}{\partial \mu_i} \tag{90}$$

This follows from the chain rule and the invariance of F under gauge transformations when transport operators transform consistently.

The geodesic force transforms similarly, ensuring full covariance:

$$\boxed{\dot{\pi}' = \mathbf{G}\dot{\pi}} \tag{91}$$

A.5 The Connection and Its Variation

A.5.1 Connection 1-Form

The transport operators Ω_{ik} encode a discrete connection on the agent network. For agents connected along an edge $e = (i, k)$, define:

$$A_e = \Omega_{ik} \in \mathrm{SO}(d) \tag{92}$$

Under gauge transformation:

$$A_e \mapsto A_e' = g_i A_e g_k^{-1} \tag{93}$$

This is the discrete analog of the gauge transformation $A \mapsto gAg^{-1} + g\,dg^{-1}$ for continuous connections.

A.5.2 Curvature

The curvature around a closed loop $\gamma = (i \to j \to k \to i)$ is:

$$F_\gamma = \Omega_{ij}\Omega_{jk}\Omega_{ki} \in \mathrm{SO}(d) \tag{94}$$

This is gauge-covariant: $F_\gamma' = g_i F_\gamma g_i^{-1}$.

A **flat connection** satisfies $F_\gamma = I$ for all loops, meaning beliefs can be consistently parallel-transported around any cycle. Nonzero curvature represents "information geometry frustration" whereby belief frames cannot be consistently aligned around cycles.

A.5.3 Variation of Connection

Consider an infinitesimal variation of the connection:

$$\delta\Omega_{ik} = \omega_{ik}\,\Omega_{ik}, \quad \omega_{ik} \in \mathfrak{so}(d) \tag{95}$$

The variation of transported precision is:

$$\delta\tilde{\Lambda}_k = \omega_{ik}\tilde{\Lambda}_k + \tilde{\Lambda}_k\omega_{ik}^T = [\omega_{ik}, \tilde{\Lambda}_k] \tag{96}$$

where this equals the commutator since $\omega_{ik}^T = -\omega_{ik}$ (antisymmetry).

A.6 Variation of the Mass Matrix Under Connection Changes

A.6.1 Diagonal Block Variation

$$\delta[\mathbf{M}^{\mu}]_{ii} = \sum_{k} \beta_{ik}\, \delta\tilde{\Lambda}_{q_k} = \sum_{k} \beta_{ik}\, [\omega_{ik}, \tilde{\Lambda}_{q_k}] \tag{97}$$

A.6.2 Off-Diagonal Block Variation

$$\begin{aligned} \delta[\mathbf{M}^{\mu}]_{ik} &= -\beta_{ik}\, \delta(\Omega_{ik}\Lambda_{q_k}) - \beta_{ki}\, \delta(\Lambda_{q_i}\Omega_{ki}^{T}) \\ &= -\beta_{ik}\omega_{ik}\Omega_{ik}\Lambda_{q_k} - \beta_{ki}\Lambda_{q_i}(\omega_{ki}\Omega_{ki})^{T} \\ &= -\beta_{ik}\omega_{ik}\Omega_{ik}\Lambda_{q_k} - \beta_{ki}\Lambda_{q_i}\Omega_{ki}^{T}\omega_{ki}^{T} \end{aligned} \tag{98}$$

Using $\omega_{ki}^{T} = -\omega_{ki}$ (antisymmetry):

$$\boxed{\delta[\mathbf{M}^{\mu}]_{ik} = -\beta_{ik}\omega_{ik}\Omega_{ik}\Lambda_{q_k} + \beta_{ki}\Lambda_{q_i}\Omega_{ki}^{T}\omega_{ki}} \tag{99}$$

A.7 Pullback Geometry

The **pullback** of the metric under a map $\phi : \mathcal{Q} \to \mathcal{Q}$ is central to understanding how geometry transforms under coordinate changes or symmetry actions.

A.7.1 Pullback of the Fisher-Rao Metric

Let $\phi_g : \mathcal{Q} \to \mathcal{Q}$ be the action of gauge transformation g:

$$\phi_g(\mu, \Sigma) = (g\mu, g\Sigma g^{T}) \tag{100}$$

The pullback metric is:

$$(\phi_g^{*}\mathcal{G})_{(\mu,\Sigma)}(v, w) = \mathcal{G}_{\phi_g(\mu,\Sigma)}(d\phi_g \cdot v, d\phi_g \cdot w) \tag{101}$$

For the Fisher-Rao metric, gauge invariance implies:

$$\boxed{\phi_g^{*}\mathcal{G} = \mathcal{G}} \tag{102}$$

The metric is gauge-invariant representing the geometric content of our transformation laws.

A.7.2 Horizontal and Vertical Decomposition

The tangent space at each point decomposes as:

$$T_{(\mu,\Sigma)}\mathcal{Q} = H_{(\mu,\Sigma)} \oplus V_{(\mu,\Sigma)} \tag{103}$$

The tangent space at each point splits into a **vertical space** V consisting of directions along gauge orbits (pure gauge changes) and a **horizontal space** H consisting of directions orthogonal to gauge orbits (physical changes). The connection determines the horizontal subspace. A vector $v = (\delta\mu, \delta\Sigma)$ is horizontal if:

$$\mathcal{G}(v, \xi_X) = 0 \quad \forall X \in \mathfrak{so}(d) \tag{104}$$

where ξ_X is the vector field generated by X.

A.7.3 Gauge-Invariant Quantities

Only horizontal components of velocities and momenta correspond to observables. Key gauge-invariant quantities include the **consensus divergence** $\|\mu_i - \tilde{\mu}_k\|^2_{\tilde{\Lambda}_{q_k}} = (\mu_i - \tilde{\mu}_k)^T \tilde{\Lambda}_{q_k} (\mu_i - \tilde{\mu}_k)$, the **free energy** $F[\{q_i\}]$ which is gauge-invariant by construction, the **Hamiltonian** $H = \frac{1}{2}\langle \pi, \mathbf{M}^{-1}\pi \rangle + F$, and the **inter-agent KL divergence** $\mathrm{KL}(q_i \| \Omega_{ik} \cdot q_k)$.

A.8 Summary: Gauge-Covariant Hamiltonian Mechanics

Gauge Transformation Laws

Positions:

$$\mu_i' = g_i \mu_i \qquad \Sigma_i' = g_i \Sigma_i g_i^T \tag{105}$$

Momenta:

$$(\pi_i^\mu)' = g_i \pi_i^\mu \qquad (\Pi_i^\Sigma)' = g_i \Pi_i^\Sigma g_i^T \tag{106}$$

Mass Matrix:

$$\mathbf{M}' = \mathbf{G}\mathbf{M}\mathbf{G}^T \tag{107}$$

Transport Operators:

$$\Omega_{ik}' = g_i \Omega_{ik} g_k^{-1} \tag{108}$$

Hamilton's Equations: Fully covariant under these transformations.
Observables: Gauge-invariant quantities include F, H, and all inter-agent divergences.

B Hamiltonian Mechanics on Statistical Manifolds

This appendix derives the complete mass matrix structure for multi-agent belief dynamics with explicit sensory evidence, demonstrating that inertial mass emerges as statistical precision. We work in the quasi-static approximation where prior parameters $(\bar{\mu}_i, \bar{\Sigma}_i)$ evolve slowly relative to beliefs (μ_i, Σ_i).

B.1 Setup and Notation

Each agent i maintains a belief distribution $q_i = \mathcal{N}(\mu_i, \Sigma_i)$ anchored to a fixed prior $p_i = \mathcal{N}(\bar{\mu}_i, \bar{\Sigma}_i)$ and receives observations o_i through a likelihood $p(o_i \mid \theta) = \mathcal{N}(o_i; \theta, \Sigma_{o_i})$. Define:

$$\begin{aligned}
\Lambda_{q_i} &= \Sigma_i^{-1} && \text{(belief precision)} && (109)\\
\bar{\Lambda}_{p_i} &= \bar{\Sigma}_i^{-1} && \text{(prior precision)} && (110)\\
\Lambda_{o_i} &= \Sigma_{o_i}^{-1} && \text{(observation precision)} && (111)\\
\tilde{\mu}_k &= \Omega_{ik}\mu_k && \text{(transported mean)} && (112)\\
\tilde{\Lambda}_{q_k} &= \Omega_{ik}\Lambda_{q_k}\Omega_{ik}^T && \text{(transported precision)} && (113)
\end{aligned}$$

where $\Omega_{ik} \in \mathrm{SO}(d)$ is the gauge transport operator from agent k's frame to agent i's frame, given by $\Omega_{ik} = e^{\phi_i}e^{-\phi_k}$ with $\phi_i \in \mathfrak{so}(d)$.

B.2 The Extended Free Energy Functional

The complete variational free energy with explicit sensory evidence is:

$$\boxed{\mathcal{F}[\{q_i\}] = \sum_i D_{\mathrm{KL}}(q_i \| p_i) + \sum_{i,k} \beta_{ik} D_{\mathrm{KL}}(q_i \| \Omega_{ik} \cdot q_k) - \sum_i \mathbb{E}_{q_i}[\log p(o_i \mid \theta)]} \tag{114}$$

The three terms represent, respectively, **prior anchoring** (deviation from internal world-model), **social consensus** (alignment with other agents via gauge-covariant transport), and **sensory evidence** (grounding in observations).

B.3 Component Free Energies for Gaussians

B.3.1 KL Divergence Between Gaussians

For $q = \mathcal{N}(\mu_q, \Sigma_q)$ and $p = \mathcal{N}(\mu_p, \Sigma_p)$:

$$D_{\mathrm{KL}}(q \| p) = \frac{1}{2}\left[\mathrm{tr}(\Sigma_p^{-1}\Sigma_q) + (\mu_p - \mu_q)^T \Sigma_p^{-1}(\mu_p - \mu_q) - d + \ln\frac{|\Sigma_p|}{|\Sigma_q|}\right] \tag{115}$$

B.3.2 Expected Log-Likelihood

For the Gaussian likelihood $p(o_i \mid \theta) = \mathcal{N}(o_i; \theta, \Sigma_{o_i})$:

$$\mathbb{E}_{q_i}[\log p(o_i \mid \theta)] = -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_{o_i}| - \frac{1}{2}\mathbb{E}_{q_i}\left[(o_i - \theta)^T \Lambda_{o_i}(o_i - \theta)\right] \tag{116}$$

The quadratic expectation evaluates to:

$$\mathbb{E}_{q_i}\left[(o_i - \theta)^T \Lambda_{o_i}(o_i - \theta)\right] = (o_i - \mu_i)^T \Lambda_{o_i}(o_i - \mu_i) + \mathrm{tr}(\Lambda_{o_i}\Sigma_i) \tag{117}$$

Therefore:

$$\boxed{-\mathbb{E}_{q_i}[\log p(o_i \mid \theta)] = \frac{1}{2}(o_i - \mu_i)^T \Lambda_{o_i}(o_i - \mu_i) + \frac{1}{2}\mathrm{tr}(\Lambda_{o_i}\Sigma_i) + \mathrm{const}} \tag{118}$$

B.4 First Variations (Gradient)

B.4.1 Prior Term: $D_{\mathrm{KL}}(q_i \| p_i)$

$$\frac{\partial D_{\mathrm{KL}}(q_i \| p_i)}{\partial \mu_i} = \bar{\Lambda}_{p_i}(\mu_i - \bar{\mu}_i) \tag{119}$$

$$\frac{\partial D_{\mathrm{KL}}(q_i \| p_i)}{\partial \Sigma_i} = \frac{1}{2}(\bar{\Lambda}_{p_i} - \Lambda_{q_i}) \tag{120}$$

B.4.2 Consensus Term: $D_{\mathrm{KL}}(q_i \| \tilde{q}_k)$

With respect to receiver i:

$$\frac{\partial D_{\mathrm{KL}}(q_i \| \tilde{q}_k)}{\partial \mu_i} = \tilde{\Lambda}_{q_k}(\mu_i - \tilde{\mu}_k) \tag{121}$$

$$\frac{\partial D_{\mathrm{KL}}(q_i \| \tilde{q}_k)}{\partial \Sigma_i} = \frac{1}{2}(\tilde{\Lambda}_{q_k} - \Lambda_{q_i}) \tag{122}$$

With respect to sender k:

$$\frac{\partial D_{\mathrm{KL}}(q_i \| \tilde{q}_k)}{\partial \mu_k} = \Lambda_{q_k}\Omega_{ik}^T(\tilde{\mu}_k - \mu_i) \tag{123}$$

$$\frac{\partial D_{\mathrm{KL}}(q_i \| \tilde{q}_k)}{\partial \Sigma_k} = \frac{1}{2}\Omega_{ik}^T\left[\tilde{\Lambda}_{q_k} - \tilde{\Lambda}_{q_k}\Sigma_i\tilde{\Lambda}_{q_k}\right]\Omega_{ik} \tag{124}$$

B.4.3 Sensory Term: $-\mathbb{E}_{q_i}[\log p(o_i \mid \theta)]$

$$\frac{\partial}{\partial \mu_i}\left[-\mathbb{E}_{q_i}[\log p(o_i \mid \theta)]\right] = \Lambda_{o_i}(\mu_i - o_i) \tag{125}$$

$$\frac{\partial}{\partial \Sigma_i}\left[-\mathbb{E}_{q_i}[\log p(o_i \mid \theta)]\right] = \frac{1}{2}\Lambda_{o_i} \tag{126}$$

B.4.4 Total Gradient

$$\boxed{\frac{\partial \mathcal{F}}{\partial \mu_i} = \bar{\Lambda}_{p_i}(\mu_i - \bar{\mu}_i) + \sum_k \beta_{ik}\tilde{\Lambda}_{q_k}(\mu_i - \tilde{\mu}_k) + \sum_j \beta_{ji}\Lambda_{q_i}\Omega_{ji}^T(\tilde{\mu}_i^{(j)} - \mu_j) + \Lambda_{o_i}(\mu_i - o_i)} \tag{127}$$

where $\tilde{\mu}_i^{(j)} = \Omega_{ji}\mu_i$ is agent i's mean transported into agent j's frame.

$$\boxed{\frac{\partial \mathcal{F}}{\partial \Sigma_i} = \frac{1}{2}(\bar{\Lambda}_{p_i} - \Lambda_{q_i}) + \sum_k \frac{\beta_{ik}}{2}(\tilde{\Lambda}_{q_k} - \Lambda_{q_i}) + \sum_j \frac{\beta_{ji}}{2}\Omega_{ji}^T\left[\tilde{\Lambda}_{qi}^{(j)} - \tilde{\Lambda}_{qi}^{(j)}\Sigma_j\tilde{\Lambda}_{qi}^{(j)}\right]\Omega_{ji} + \frac{1}{2}\Lambda_{o_i}} \tag{128}$$

B.5 Second Variations (Hessian = Mass Matrix)

The Hessian $\mathbf{M} = \partial^2\mathcal{F}/\partial\xi\partial\xi$ serves as the mass matrix. This is distinct from the intrinsic Fisher-Rao metric on the statistical manifold; it is a "Hessian mass matrix" whose curvature depends on the full free energy landscape including priors, sensory evidence, and social coupling.

B.5.1 Mean Sector: $\partial^2\mathcal{F}/\partial\mu\partial\mu^T$

Diagonal blocks $(i = k)$**:** From prior:

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i\|p_i)}{\partial\mu_i\partial\mu_i^T} = \bar{\Lambda}_{p_i} \tag{129}$$

From consensus (as receiver):

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i\|\tilde{q}_k)}{\partial\mu_i\partial\mu_i^T} = \tilde{\Lambda}_{q_k} = \Omega_{ik}\Lambda_{q_k}\Omega_{ik}^T \tag{130}$$

From consensus (as sender to agent j):

$$\frac{\partial^2 D_{\mathrm{KL}}(q_j\|\tilde{q}_i)}{\partial\mu_i\partial\mu_i^T} = \Omega_{ji}^T\tilde{\Lambda}_{qi}^{(j)}\Omega_{ji} = \Lambda_{q_i} \tag{131}$$

From sensory evidence:

$$\frac{\partial^2}{\partial\mu_i\partial\mu_i^T}\left[-\mathbb{E}_{q_i}[\log p(o_i \mid \theta)]\right] = \Lambda_{o_i} \tag{132}$$

Total diagonal mass:

$$\boxed{[\mathbf{M}^\mu]_{ii} = \underbrace{\bar{\Lambda}_{p_i}}_{\text{prior}} + \underbrace{\sum_k \beta_{ik}\tilde{\Lambda}_{q_k}}_{\text{incoming social}} + \underbrace{\sum_j \beta_{ji}\Lambda_{q_i}}_{\text{outgoing recoil}} + \underbrace{\Lambda_{o_i}}_{\text{sensory}}} \tag{133}$$

Off-diagonal blocks $(i \neq k)$**:** From $D_{\mathrm{KL}}(q_i\|\tilde{q}_k)$:

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i\|\tilde{q}_k)}{\partial\mu_i\partial\mu_k^T} = -\tilde{\Lambda}_{q_k}\Omega_{ik} = -\Omega_{ik}\Lambda_{q_k} \tag{134}$$

From $D_{\mathrm{KL}}(q_k\|\tilde{q}_i)$ (if k also listens to i):

$$\frac{\partial^2 D_{\mathrm{KL}}(q_k\|\tilde{q}_i)}{\partial\mu_i\partial\mu_k^T} = -\Lambda_{q_i}\Omega_{ki}^T \tag{135}$$

The sensory term does not couple different agents. Therefore:

$$\boxed{[\mathbf{M}^\mu]_{ik} = -\beta_{ik}\Omega_{ik}\Lambda_{q_k} - \beta_{ki}\Lambda_{q_i}\Omega_{ki}^T \quad (i \neq k)} \tag{136}$$

By Schwarz's theorem, this Hessian is symmetric: one can verify that $[\mathbf{M}^\mu]_{ik} = ([\mathbf{M}^\mu]_{ki})^T$ holds for any β_{ik} and Ω_{ik}.

B.5.2 Covariance Sector: $\partial^2\mathcal{F}/\partial\Sigma\partial\Sigma$

For matrix-valued variables, we use the directional derivative convention:

$$\frac{\partial^2 f}{\partial\Sigma\partial\Sigma}[V,W] = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\left.\frac{\partial f}{\partial\Sigma}\right|_{\Sigma+\epsilon W} - \left.\frac{\partial f}{\partial\Sigma}\right|_{\Sigma}\right)[V] \tag{137}$$

Key identity:

$$\frac{\partial}{\partial\Sigma}(\Sigma^{-1}) = -\Sigma^{-1}\otimes\Sigma^{-1} \tag{138}$$

Diagonal blocks $(i = k)$**:** From prior:

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i\|p_i)}{\partial\Sigma_i\partial\Sigma_i}[V,W] = \frac{1}{2}\mathrm{tr}\left[\Lambda_{q_i}V\Lambda_{q_i}W\right] \tag{139}$$

In tensor notation:

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i\|p_i)}{\partial\Sigma_i\partial\Sigma_i} = \frac{1}{2}(\Lambda_{q_i}\otimes\Lambda_{q_i}) \tag{140}$$

From consensus (as receiver and sender), identical contributions arise.

Critical observation: The sensory term $\frac{1}{2}\mathrm{tr}(\Lambda_{o_i}\Sigma_i)$ is *linear* in Σ_i, so its second derivative **vanishes**:

$$\frac{\partial^2}{\partial\Sigma_i\partial\Sigma_i}\left[\mathrm{tr}(\Lambda_{o_i}\Sigma_i)\right] = 0 \tag{141}$$

Therefore:

$$\boxed{[\mathbf{M}^\Sigma]_{ii} = \frac{1}{2}(\Lambda_{q_i}\otimes\Lambda_{q_i})\cdot\left(1+\sum_k\beta_{ik}+\sum_j\beta_{ji}\right)} \tag{142}$$

The sensory precision Λ_{o_i} does **not** contribute to the covariance-sector mass.

B.5.3 Mean-Covariance Cross Blocks

Prior term:

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i\|p_i)}{\partial\mu_i\partial\Sigma_i} = 0 \tag{143}$$

Sensory term: The sensory free energy decomposes into a component quadratic in μ_i, namely $(o_i-\mu_i)^T\Lambda_{o_i}(o_i-\mu_i)$, and a component linear in Σ_i, namely $\mathrm{tr}(\Lambda_{o_i}\Sigma_i)$. These are independent, so:

$$[\mathbf{C}^{\mu\Sigma}]_{ii}^{\text{sensory}} = 0 \tag{144}$$

Consensus (cross-agent): From $\partial D_{\mathrm{KL}}(q_i\|\tilde{q}_k)/\partial\mu_i = \tilde{\Lambda}_{q_k}(\mu_i-\tilde{\mu}_k)$, varying Σ_k:

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i\|\tilde{q}_k)}{\partial\mu_i\partial\Sigma_k}[V] = -\Omega_{ik}\Lambda_{q_k}V\Lambda_{q_k}\Omega_{ik}^T(\mu_i-\tilde{\mu}_k) \tag{145}$$

This vanishes at consensus $(\mu_i = \tilde{\mu}_k)$:

$$[\mathbf{C}^{\mu\Sigma}]_{ik} = 0 \quad \text{when } \mu_i = \tilde{\mu}_k \tag{146}$$

B.6 Complete Mass Matrix Assembly

The full state vector is $\xi = (\mu_1, \ldots, \mu_N, \Sigma_1, \ldots, \Sigma_N)$.

B.6.1 Block Structure

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}^\mu & \mathbf{C}^{\mu\Sigma} \\ (\mathbf{C}^{\mu\Sigma})^T & \mathbf{M}^\Sigma \end{pmatrix} \tag{147}$$

where each block is an $N \times N$ matrix of sub-blocks.

B.6.2 Explicit Formulae

Mean sector diagonal:

$$[\mathbf{M}^\mu]_{ii} = \underbrace{\bar{\Lambda}_{p_i}}_{\text{prior anchoring}} + \underbrace{\sum_k \beta_{ik}\Omega_{ik}\Lambda_{q_k}\Omega_{ik}^T}_{\text{incoming consensus}} + \underbrace{\sum_j \beta_{ji}\Lambda_{q_i}}_{\text{outgoing recoil}} + \underbrace{\Lambda_{o_i}}_{\text{sensory grounding}} \tag{148}$$

Mean sector off-diagonal:

$$[\mathbf{M}^\mu]_{ik} = -\beta_{ik}\Omega_{ik}\Lambda_{q_k} - \beta_{ki}\Lambda_{q_i}\Omega_{ki}^T \quad (i \neq k) \tag{149}$$

Covariance sector diagonal:

$$[\mathbf{M}^\Sigma]_{ii} = \frac{1}{2}(\Lambda_{q_i} \otimes \Lambda_{q_i}) \cdot \left(1 + \sum_k \beta_{ik} + \sum_j \beta_{ji}\right) \tag{150}$$

Cross mean-covariance (at consensus):

$$[\mathbf{C}^{\mu\Sigma}]_{ik} = 0 \quad \text{when } \mu_i = \tilde{\mu}_k \tag{151}$$

B.7 Interpretation

B.7.1 Mass as Precision

The mean-sector effective mass for agent i is:

$$\boxed{M_i = \bar{\Lambda}_{p_i} + \sum_k \beta_{ik}\tilde{\Lambda}_{q_k} + \sum_j \beta_{ji}\Lambda_{q_i} + \Lambda_{o_i}} \tag{152}$$

The term $\bar{\Lambda}_{p_i}$ represents **bare mass**, providing inertia against deviation from the prior. The term $\sum_k \beta_{ik}\tilde{\Lambda}_{q_k}$ represents **incoming relational mass**, the inertia from being "pulled" by neighbors. The term $\sum_j \beta_{ji}\Lambda_{q_i}$ represents **outgoing relational mass**, the inertia from "pulling" neighbors (recoil). Finally, Λ_{o_i} represents **sensory mass**, the inertia from grounding in observations.

B.7.2 Asymmetry of Sensory Contribution

The sensory precision Λ_{o_i} contributes to the mean-sector mass (Eq. 148) and the mean-sector force (Eq. 127), but not to the covariance-sector mass (Eq. 142). This asymmetry arises because the sensory term is quadratic in μ but only linear in Σ.

B.7.3 Kinetic Energy

$$T = \frac{1}{2}\dot{\mu}^T \mathbf{M}^\mu \dot{\mu} + \frac{1}{2}\mathrm{tr}\left[\mathbf{M}^\Sigma[\dot{\Sigma}, \dot{\Sigma}]\right] \tag{153}$$

The first term gives standard kinetic energy with precision-mass. The second gives "shape" kinetic energy on the SPD manifold; the metric $(\Lambda \otimes \Lambda)$ corresponds to the affine-invariant metric on the cone of symmetric positive-definite matrices.

B.8 The Hamiltonian

With conjugate momenta $\pi = (\pi^\mu, \Pi^\Sigma)$ and Hamiltonian:

$$\boxed{H = \frac{1}{2}\langle \pi, \mathbf{M}^{-1}\pi \rangle + \mathcal{F}[\xi]} \tag{154}$$

B.9 Hamilton's Equations

B.9.1 Equations of Motion

$$\dot{\mu}_i = \sum_k [\mathbf{M}^{-1}]^{\mu\mu}_{ik} \pi^\mu_k + \sum_k [\mathbf{M}^{-1}]^{\mu\Sigma}_{ik} \Pi^\Sigma_k \tag{155}$$

$$\dot{\Sigma}_i = \sum_k [\mathbf{M}^{-1}]^{\Sigma\mu}_{ik} \pi^\mu_k + \sum_k [\mathbf{M}^{-1}]^{\Sigma\Sigma}_{ik} \Pi^\Sigma_k \tag{156}$$

$$\dot{\pi}^\mu_i = -\frac{\partial \mathcal{F}}{\partial \mu_i} - \frac{1}{2}\pi^T \frac{\partial \mathbf{M}^{-1}}{\partial \mu_i} \pi \tag{157}$$

$$\dot{\Pi}^\Sigma_i = -\frac{\partial \mathcal{F}}{\partial \Sigma_i} - \frac{1}{2}\pi^T \frac{\partial \mathbf{M}^{-1}}{\partial \Sigma_i} \pi \tag{158}$$

B.9.2 Force Decomposition

The potential forces decompose into four distinct contributions:

$$\boxed{-\frac{\partial \mathcal{F}}{\partial \mu_i} = \underbrace{-\bar{\Lambda}_{p_i}(\mu_i - \bar{\mu}_i)}_{\text{prior restoring}} \underbrace{- \sum_k \beta_{ik} \tilde{\Lambda}_{q_k}(\mu_i - \tilde{\mu}_k)}_{\text{consensus}} \underbrace{- \sum_j \beta_{ji} \Lambda_{q_i} \Omega^T_{ji}(\tilde{\mu}^{(j)}_i - \mu_j)}_{\text{reciprocal}} \underbrace{-\Lambda_{o_i}(\mu_i - o_i)}_{\text{sensory evidence}}} \tag{159}$$

The geodesic force $f^{\text{geo}}_i = -\frac{1}{2}\sum_{jkl}(\pi^\mu_j)^T \frac{\partial [\mathbf{M}^{-1}]^{\mu\mu}_{jk}}{\partial \mu_i} \pi^\mu_k$ encodes manifold curvature.

B.9.3 Compact Form

$$\boxed{\begin{aligned} \dot{\xi} &= \mathbf{M}^{-1}\pi \\ \dot{\pi} &= -\nabla\mathcal{F} - \frac{1}{2}\nabla_\xi\langle\pi, \mathbf{M}^{-1}\pi\rangle \end{aligned}} \tag{160}$$

Since $\mathbf{M}$ depends on Σ, the Hamiltonian is non-separable, requiring implicit or splitting methods for symplectic integration.

B.10 Damped Dynamics

Including dissipation yields:

$$M_i\ddot{\mu}_i + \gamma_i\dot{\mu}_i + \nabla_{\mu_i}\mathcal{F} = 0 \tag{161}$$

For small displacements from equilibrium with stiffness $K_i = \nabla^2\mathcal{F}|_{\mu^*}$:

$$M_i\ddot{\delta\mu} + \gamma_i\dot{\delta\mu} + K_i\delta\mu = 0 \tag{162}$$

The discriminant $\Delta = \gamma_i^2 - 4K_iM_i$ determines three regimes: the **overdamped** regime ($\Delta > 0$) exhibits monotonic decay corresponding to standard Bayesian updating; the **critically damped** regime ($\Delta = 0$) achieves fastest equilibration; and the **underdamped** regime ($\Delta < 0$) exhibits oscillatory approach with overshooting.

B.11 Momentum Current with Sensory Coupling

Between agents, the momentum current is:

$$J_{k\to i} = \beta_{ik}\tilde{\Lambda}_{q_k}(\tilde{\mu}_k - \mu_i) \tag{163}$$

The continuity equation becomes:

$$\dot{\pi}_i + \gamma_i\dot{\mu}_i + \bar{\Lambda}_{p_i}(\mu_i - \bar{\mu}_i) + \Lambda_{o_i}(\mu_i - o_i) = \sum_k J_{k\to i} \tag{164}$$

The sensory term $\Lambda_{o_i}(\mu_i - o_i)$ acts as an additional "anchoring force" that grounds the agent in observations, distinct from the social momentum currents.

B.12 Summary

The Complete Theory with Sensory Evidence

State: Each agent i has belief $q_i = \mathcal{N}(\mu_i, \Sigma_i)$ with fixed prior $p_i = \mathcal{N}(\bar{\mu}_i, \bar{\Sigma}_i)$ and observations o_i with precision Λ_{o_i}.
Free Energy:

$$\mathcal{F} = \sum_i D_{\mathrm{KL}}(q_i \| p_i) + \sum_{i,k} \beta_{ik} D_{\mathrm{KL}}(q_i \| \Omega_{ik} \cdot q_k) - \sum_i \mathbb{E}_{q_i}[\log p(o_i \mid \theta)] \tag{165}$$

Mass Matrix:

$$\mathbf{M} = \frac{\partial^2 \mathcal{F}}{\partial \xi \partial \xi} = \text{Hessian of free energy} = \text{Precision} \tag{166}$$

Effective Mass:

$$M_i = \bar{\Lambda}_{p_i} + \sum_k \beta_{ik} \tilde{\Lambda}_{q_k} + \sum_j \beta_{ji} \Lambda_{q_i} + \Lambda_{o_i} \tag{167}$$

Dynamics:

$$\dot{\xi} = \mathbf{M}^{-1}\pi, \qquad \dot{\pi} = -\nabla \mathcal{F} - \frac{1}{2} \nabla_{\xi} \langle \pi, \mathbf{M}^{-1}\pi \rangle \tag{168}$$

Interpretation: Position μ_i represents what agent i believes; momentum π_i represents the rate of belief change times precision; mass equals precision (tight beliefs are heavy); and force represents the pull toward prior, consensus, and observations.

C Derivation of Softmax Attention from Maximum Entropy

The softmax attention weights β_{ij} are not arbitrary but emerge from a principled maximum entropy argument. We seek attention weights that satisfy a constraint on expected dissimilarity while being maximally uncertain otherwise.

C.1 The Maximum Entropy Problem

For agent i, let β_{ij} denote the attention weight assigned to neighbor j, subject to the constraint $\sum_j \beta_{ij} = 1$. We seek the distribution over attention that maximizes entropy subject to a constraint on expected belief dissimilarity:

$$\max_{\{\beta_{ij}\}} \left[-\sum_j \beta_{ij} \log \beta_{ij} \right] \quad \text{subject to} \quad \sum_j \beta_{ij}\, d_{ij}^2 = C_i, \quad \sum_j \beta_{ij} = 1 \tag{169}$$

where $d_{ij}^2 = \|\mu_i - \mu_j\|_{\Lambda}^2$ is the precision-weighted squared distance between beliefs.

C.2 Lagrangian Solution

Introducing Lagrange multipliers λ and ν, the Lagrangian is:

$$\mathcal{L} = -\sum_j \beta_{ij} \log \beta_{ij} - \lambda \left(\sum_j \beta_{ij} d_{ij}^2 - C_i \right) - \nu \left(\sum_j \beta_{ij} - 1 \right) \tag{170}$$

Taking the derivative with respect to β_{ik} and setting it to zero:

$$\frac{\partial \mathcal{L}}{\partial \beta_{ik}} = -\log \beta_{ik} - 1 - \lambda d_{ik}^2 - \nu = 0 \tag{171}$$

Solving for β_{ik}:

$$\beta_{ik} = \exp\left(-1 - \nu - \lambda d_{ik}^2\right) = \frac{\exp(-\lambda d_{ik}^2)}{Z_i} \tag{172}$$

where $Z_i = \sum_j \exp(-\lambda d_{ij}^2)$ is the normalizing partition function determined by the constraint $\sum_j \beta_{ij} = 1$.

C.3 Interpretation

The result is exactly the softmax (or Gibbs/Boltzmann) distribution over attention:

$$\boxed{\beta_{ij} = \frac{\exp(-d_{ij}^2/\kappa)}{\sum_k \exp(-d_{ik}^2/\kappa)}} \tag{173}$$

where $\kappa = 1/\lambda$ is an ”attention temperature” controlling how sharply attention concentrates on similar neighbors. In the limit $\kappa \to 0$, attention becomes deterministic (winner-take-all), while $\kappa \to \infty$ yields uniform attention regardless of similarity. This derivation shows that softmax attention is not merely a convenient mathematical choice but the unique solution that maximizes entropy subject to a constraint on expected dissimilarity. The same principle underlies the Boltzmann distribution in statistical mechanics and the exponential family in statistics.

D Why Forward KL Divergence?

A natural question arises: why use the forward KL divergence $D_{\mathrm{KL}}(q_i \| \Omega_{ik} \cdot q_k)$ rather than the reverse $D_{\mathrm{KL}}(\Omega_{ik} \cdot q_k \| q_i)$ in the social coupling term? This appendix provides both intuitive and formal justifications.

D.1 Intuitive Argument: Zero-Forcing vs. Mean-Seeking

The forward KL $D_{\mathrm{KL}}(q\|p)$ is called ”mean-seeking” or ”moment-matching” in that it penalizes q for placing probability mass where p has little probability mass, encouraging q to cover the support of p. The reverse KL $D_{\mathrm{KL}}(p\|q)$ is called ”mode-seeking” or ”zero-forcing” in that it penalizes q for failing to place probability mass where p has probability mass, encouraging q to concentrate on modes of p.

For social influence, the forward direction is more natural. Agent i should be penalized for holding beliefs that disagree with transported neighbor beliefs, regardless of how uncertain those neighbors are. The forward KL achieves this: if agent i is certain about something that neighbor k (in i's frame) considers unlikely, i pays a high cost.

D.2 Formal Uniqueness Argument

Consider the general family of f-divergences $D_f(q\|p) = \int p(x)f(q(x)/p(x))dx$ for convex f with $f(1) = 0$. We seek divergences satisfying the following desiderata for social coupling:

Desideratum 1: Convexity in q. The coupling term should be convex in agent i's belief q_i, ensuring a well-defined optimization landscape. All f-divergences satisfy this when the first argument varies.

Desideratum 2: Additivity under transport. Under the gauge transport Ω_{ik}, the divergence should decompose sensibly. For Gaussians, this requires the divergence to depend on beliefs only through sufficient statistics (μ, Σ).

Desideratum 3: Fisher information as Hessian. The Hessian of the coupling term at $q_i = \Omega_{ik} \cdot q_k$ should recover the Fisher information metric, ensuring consistency with the mass interpretation. Only the forward KL satisfies this: for $D_{\mathrm{KL}}(q\|p)$, we have $\nabla_q^2 D_{\mathrm{KL}}(q\|p)\big|_{q=p} = \mathcal{I}_p$, the Fisher information at p.

The reverse KL has Hessian $\nabla_q^2 D_{\mathrm{KL}}(p\|q)\big|_{q=p} = \mathcal{I}_p$ as well, but crucially, it is not convex in q globally—only locally near $q = p$. This makes the reverse KL unsuitable for the free energy minimization framework, as it can create spurious local minima.

D.3 Connection to Variational Inference

In variational inference, the forward KL $D_{\mathrm{KL}}(q\|p)$ is the natural objective when q is the variational approximation and p is the target. Minimizing this drives q toward p in an information-theoretic sense. Our social coupling term $D_{\mathrm{KL}}(q_i\|\Omega_{ik} \cdot q_k)$ follows this convention: agent i's belief q_i is driven toward the transported neighbor belief $\Omega_{ik} \cdot q_k$. The reverse direction would have neighbors' beliefs driving toward agent i, inverting the causal structure of social influence.

E Derivation of Social Coupling from Normalized Generative Model

While it is possible to construct a fully general variational free energy involving hyper-priors, prior-prior agreement, and general bundle morphisms (e.g. q and p could, in principle live in entirely different statistical fibers) we focus only on the quasi-static limit whereby priors evolve much slower than beliefs.

This construction leads to a generalized variational free energy given by

$$\begin{aligned}\mathcal{F}[\{q_i\},\{s_i\}] = &\underbrace{\sum_i D_{\mathrm{KL}}(q_i\|p_i)}_{\text{(1) Belief prior}} + \underbrace{\sum_i D_{\mathrm{KL}}(s_i\|r_i)}_{\text{(2) Model prior}} \\ &+ \underbrace{\sum_{i,j} \beta_{ij} D_{\mathrm{KL}}(q_i\|\Omega_{ij} q_j)}_{\text{(3) Belief alignment}} \\ &+ \underbrace{\sum_{i,j} \gamma_{ij} D_{\mathrm{KL}}(s_i\|\tilde{\Omega}_{ij} s_j)}_{\text{(4) Model alignment}} \\ &- \underbrace{\mathbb{E}_q[\log p(o \mid \{k_i\},\{m_i\})]}_{\text{(5) Observation likelihood}}\end{aligned} \tag{174}$$

where r_i is an agents hyper-prior, $\tilde{\Omega}_{ij}$ is a transport operator distinct from Ω_{ij} on the "model fiber" $\mathcal{B}_q$, s_i is an agent's prior and γ_{ij} is the model-fiber analogue to β_{ij} (k, m are latents).

In the limit where all agents share gauge-equivalent priors, which do not vary substantially over time, then we can focus on the fast variables q to consider the variational free energy given in the text. All that remains is to explicitly show how the social/attention coupling term arises.

The social coupling term $\sum_{i,k} \beta_{ik} D_{\mathrm{KL}}(q_i\|\Omega_{ik} \cdot q_k)$ in our free energy is not an ad hoc regularizer but emerges naturally from a properly normalized generative model with auxiliary agreement variables. This appendix provides the derivation, establishing the coupling as a consequence of gauge-transported Gaussian consistency constraints. The generalized variational free energy, as given above, then emerges analogously.

E.1 Auxiliary Agreement Variables

To enforce consistency between agents after gauge transport, we introduce an auxiliary agreement variable for each ordered pair (i, k):

$$z_{ik} \in \mathbb{R}^d \tag{175}$$

The variable z_{ik} represents what agent i believes agent k's belief looks like after transporting k's belief into i's gauge frame. These auxiliary variables serve as latent mediators that, when integrated out, yield the effective pairwise coupling between agents.

E.2 The Product-of-Gaussians Construction

Each agreement variable z_{ik} is drawn from a product of two Gaussians that simultaneously constrain it to match both agent i's belief and the gauge-transported belief of agent k:

$$p(z_{ik} \mid \mu_i, \mu_k) \propto \mathcal{N}(z_{ik}; \mu_i, \Lambda_{ik}^{-1}) \cdot \mathcal{N}(z_{ik}; \Omega_{ik}\mu_k, \Lambda_{ik}^{-1}) \tag{176}$$

where $\Lambda_{ik} \succ 0$ is the alignment precision matrix and $\Omega_{ik} \in \mathrm{SO}(d)$ is the gauge transport operator from agent k's frame to agent i's frame.

When we integrate out the agreement variable z_{ik}, the normalizing constant depends on (μ_i, μ_k). Using the standard result for products of Gaussians:

$$\int \mathcal{N}(z; a, \Sigma) \cdot \mathcal{N}(z; b, \Sigma)\, dz = \mathcal{N}(a; b, 2\Sigma) \propto \exp\left(-\frac{1}{4}(a-b)^\top \Sigma^{-1}(a-b)\right) \tag{177}$$

This induces an effective coupling between agents. The marginal prior over belief means, after integrating out all agreement variables, takes the form:

$$p(\{\mu_i\}) \propto \left[\prod_i p_i(\mu_i)\right] \cdot \exp\left(-\frac{1}{4}\sum_{i,k}(\mu_i - \Omega_{ik}\mu_k)^\top \Lambda_{ik}(\mu_i - \Omega_{ik}\mu_k)\right) \tag{178}$$

where $p_i(\mu_i) = \mathcal{N}(\mu_i; \bar{\mu}_i, \bar{\Sigma}_i)$ is agent i's local prior. The quadratic terms in the exponential encode pairwise alignment costs, with the factor of 1/4 arising directly from the Gaussian product formula. This construction yields tractable quadratic potentials, in contrast to general unnormalized Markov random fields where partition functions are intractable.

E.3 Variational Free Energy

Under a mean-field posterior approximation $q(\{\mu_i\}) = \prod_i q_i(\mu_i)$ with Gaussian factors $q_i = \mathcal{N}(\mu_i; \mu_{q,i}, \Sigma_{q,i})$, the variational free energy is:

$$\mathcal{F} = \mathbb{E}_q[\log q] - \mathbb{E}_q[\log p(\{\mu_i\})] - \mathbb{E}_q[\log p(o \mid \{\mu_i\})] \tag{179}$$

Using the marginal prior from the previous section, we have $-\log p(\{\mu_i\}) = \sum_i[-\log p_i(\mu_i)] + \frac{1}{4}\sum_{i,k}(\mu_i - \Omega_{ik}\mu_k)^\top \Lambda_{ik}(\mu_i - \Omega_{ik}\mu_k) + \text{const}$. Taking expectations:

$$\mathcal{F} = \sum_i D_{\mathrm{KL}}(q_i \| p_i) + \frac{1}{4}\sum_{i,k}\mathbb{E}_{q_i q_k}\left[(\mu_i - \Omega_{ik}\mu_k)^\top \Lambda_{ik}(\mu_i - \Omega_{ik}\mu_k)\right] - \mathbb{E}_q[\log p(o \mid \{\mu_i\})] \tag{180}$$

For independent Gaussians, the quadratic expectation evaluates to:

$$\mathbb{E}_{q_i q_k}[(\mu_i - \Omega_{ik}\mu_k)^\top \Lambda_{ik}(\mu_i - \Omega_{ik}\mu_k)] = \mathrm{tr}\left[\Lambda_{ik}(\Sigma_{q,i} + \Omega_{ik}\Sigma_{q,k}\Omega_{ik}^\top)\right] + (\mu_{q,i} - \Omega_{ik}\mu_{q,k})^\top \Lambda_{ik}(\mu_{q,i} - \Omega_{ik}\mu_{q,k}) \tag{181}$$

E.4 The Alignment Regime

We choose the coupling precision to be proportional to the transported neighbor's precision:

$$\Lambda_{ik} := \tau_{ik}(\Omega_{ik}\Sigma_{q,k}\Omega_{ik}^\top)^{-1} \tag{182}$$

where $\tau_{ik} > 0$ is a dimensionless coupling strength.

In the alignment regime where the coupling drives beliefs toward consistency, we have $\Sigma_{q,i} \approx \Omega_{ik}\Sigma_{q,k}\Omega_{ik}^\top$. Under this condition, the quadratic expectation becomes proportional to the KL divergence:

$$\frac{1}{4}\mathbb{E}_{q_i q_k}[(\mu_i - \Omega_{ik}\mu_k)^\top \Lambda_{ik}(\mu_i - \Omega_{ik}\mu_k)] \approx \frac{\tau_{ik}}{2} D_{\mathrm{KL}}(q_i \| \Omega_{ik} \cdot q_k) + \text{const} \tag{183}$$

where the constant absorbs dimension-dependent terms. Defining the normalized attention weight $\beta_{ik} := \tau_{ik}/2$, we obtain the final form of the social coupling term.

E.5 The Complete Free Energy

Assembling all terms, the variational free energy takes the form used throughout this paper:

$$\boxed{\mathcal{F}[\{q_i\}] = \sum_i D_{\mathrm{KL}}(q_i \| p_i) + \sum_{i,k} \beta_{ik} D_{\mathrm{KL}}(q_i \| \Omega_{ik} \cdot q_k) - \sum_i \mathbb{E}_{q_i}[\log p(o_i \mid \theta)]} \tag{184}$$

This derivation establishes several key points. First, the KL-based social coupling is not an ad hoc choice but emerges from integrating out agreement variables in a normalized generative model. Second, the attention weights β_{ik} have a clear interpretation as half the dimensionless coupling strength τ_{ik}. Third, the forward KL direction arises naturally from the construction where agent i's belief is compared against transported neighbor beliefs. Fourth, the gauge transport operators Ω_{ik} enter through the agreement variable construction, ensuring gauge covariance of the full theory.

The agreement variable construction can be understood intuitively: z_{ik} represents a negotiated belief state that both agents would accept as consistent. The product-of-Gaussians prior penalizes disagreement between agent i's belief and the transported belief of agent k. Integrating out this mediator leaves an effective attraction between the beliefs, weighted by how precisely they are required to agree.

F Generalization Beyond Gaussian Beliefs

The Gaussian assumption, while analytically convenient, restricts the theory to unimodal beliefs with elliptical uncertainty. This appendix outlines how the framework extends to exponential family distributions and discusses the challenges of fully general distributions.

F.1 Exponential Family Formulation

An exponential family distribution takes the form:

$$p(x \mid \theta) = h(x) \exp\left[\theta^T T(x) - A(\theta)\right] \tag{185}$$

where $\theta \in \mathbb{R}^k$ are natural parameters, $T(x)$ are sufficient statistics, and $A(\theta)$ is the log-partition function ensuring normalization.

F.1.1 Fisher Information as Mass

The Fisher information matrix for exponential families has the elegant form:

$$\mathcal{I}(\theta) = \nabla^2 A(\theta) = \mathrm{Cov}_{p(\cdot|\theta)}[T(x)] \tag{186}$$

This is the Hessian of the log-partition function, equivalently the covariance of sufficient statistics. Our identification of mass with Fisher information thus generalizes immediately:

$$\boxed{M_i = \nabla^2 A(\theta_i) = \mathcal{I}(\theta_i)} \tag{187}$$

For Gaussians with natural parameters $\theta = (\Sigma^{-1}\mu, -\frac{1}{2}\Sigma^{-1})$, this recovers $M = \Lambda$ (precision) in the mean sector.

F.1.2 KL Divergence Structure

For exponential families, the KL divergence takes the Bregman divergence form:

$$D_{\mathrm{KL}}(q\|p) = A(\theta_p) - A(\theta_q) - \nabla A(\theta_q)^T(\theta_p - \theta_q) \tag{188}$$

The Hessian of this with respect to θ_q yields:

$$\frac{\partial^2 D_{\mathrm{KL}}(q\|p)}{\partial\theta_q \partial\theta_q^T} = \nabla^2 A(\theta_q) = \mathcal{I}(\theta_q) \tag{189}$$

Thus the mass-as-Fisher-information identification holds for all exponential families.

F.2 Multi-Modal Beliefs: Mixture Models

Real beliefs are often multi-modal, representing hypothesis competition or attitude ambivalence. Gaussian mixtures provide a tractable extension:

$$q(x) = \sum_{m=1}^{M} w_m \mathcal{N}(x; \mu_m, \Sigma_m), \quad \sum_m w_m = 1 \tag{190}$$

F.2.1 Extended State Space

The state space becomes $\xi = (\{w_m\}, \{\mu_m\}, \{\Sigma_m\})$, comprising mixture weights on the $(M-1)$-simplex, component means in $\mathbb{R}^{d\times M}$, and component covariances in $\mathrm{SPD}(d)^M$.

F.2.2 Fisher Information for Mixtures

The Fisher information for mixture models does not decompose cleanly. The full Fisher matrix couples all parameters:

$$\mathcal{I} = \begin{pmatrix} \mathcal{I}^{ww} & \mathcal{I}^{w\mu} & \mathcal{I}^{w\Sigma} \\ (\mathcal{I}^{w\mu})^T & \mathcal{I}^{\mu\mu} & \mathcal{I}^{\mu\Sigma} \\ (\mathcal{I}^{w\Sigma})^T & (\mathcal{I}^{\mu\Sigma})^T & \mathcal{I}^{\Sigma\Sigma} \end{pmatrix} \tag{191}$$

The weight-weight block takes the form:

$$\mathcal{I}^{ww}_{mn} = \int \frac{1}{q(x)} \frac{\partial q}{\partial w_m} \frac{\partial q}{\partial w_n} dx = \int \frac{\mathcal{N}_m(x)\mathcal{N}_n(x)}{q(x)} dx \tag{192}$$

This integral generally requires numerical evaluation but captures the intuition that weight-inertia depends on component overlap: well-separated modes have nearly independent weight dynamics.

F.3 Fully General Distributions

For arbitrary distributions $q(x)$ without parametric form, the state space becomes infinite-dimensional (a functional space). The Fisher-Rao metric generalizes to:

$$\langle \delta q_1, \delta q_2 \rangle_q = \int \frac{\delta q_1(x)\, \delta q_2(x)}{q(x)} dx \tag{193}$$

This defines a Riemannian metric on the space of probability densities. Gradient flow with respect to this metric yields the natural gradient dynamics studied in information geometry (Amari, 2016).

F.4 Gauge Structure for Non-Gaussian Beliefs

The gauge transport operators Ω_{ij} generalize beyond rotations when beliefs are non-Gaussian.

For exponential families transport acts on natural parameters: $\theta \mapsto \Omega \cdot \theta$ where Ω preserves the parameter space structure.

In the case of mixtures transport can permute components, rotate within-component means, and transform covariances.

For general distributions transport becomes a diffeomorphism of the sample space such that $q(x) \mapsto q(\Omega^{-1}(x))|\det D\Omega^{-1}|$.

The gauge group expands from SO(d) to the group of measure-preserving diffeomorphisms, with corresponding enlargement of the connection structure.

F.5 Summary

The ansatz that mass equals Fisher information generalizes immediately to exponential families and, in principle, to arbitrary distributions. The Gaussian case developed in the main text is the simplest instance of a broader geometric framework. Extension to mixtures captures multi-modal beliefs at the cost of increased computational complexity. Fully general distributions require infinite-dimensional analysis or computational approximation, but the underlying geometric principle remains unchanged.

References

Adams, H. (1918). *The Education of Henry Adams.* Houghton Mifflin, Boston.

Amari, S. (2016). *Information Geometry and Its Applications.* Springer.

Anderson, C. A., Lepper, M. R., and Ross, L. (1980a). Perseverance of social theories. *Journal of Personality and Social Psychology*, 39(6):1037.

Anderson, C. A., Lepper, M. R., and Ross, L. (1980b). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, 39(6):1037–1049.

Arnold, V. I. (1989). *Mathematical methods of classical mechanics.* Springer.

Bernevig, B. A. and Hughes, T. L. (2013). *Topological insulators and topological superconductors*. Princeton University Press.

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, 76:198–211.

Burge, T. (2010). *Origins of Objectivity*. Oxford University Press.

Busemeyer, J. R. and Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge University Press.

Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204.

Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–2678.

Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(01n04):87–98.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121.

Eagly, A. H. and Chaiken, S. (1993). *The Psychology of Attitudes*. Harcourt Brace Jovanovich.

Fink, E. L., Kaplowitz, S. A., and Hubbard, S. M. (2002). Oscillation in beliefs and decisions. *The Persuasion Handbook: Developments in Theory and Practice*, pages 17–37.

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., and Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4).

Friedkin, N. E. and Johnsen, E. C. (1990). Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206.

Friedkin, N. E. and Johnsen, E. C. (2011). *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*. Cambridge University Press.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, 4(11):e1000211.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11:127–138.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., and Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, 68:862–879.

Galam, S. (2012). *Sociophysics: A physicist's modeling of psycho-political phenomena.* Springer.

Goldenfeld, N. (1992). *Lectures on phase transitions and the renormalization group.* Addison-Wesley.

Hegselmann, R. and Krause, U. (2002a). Opinion dynamics and bounded confidence models, analysis, and simulation. In *Journal of Artificial Societies and Social Simulation*, volume 5.

Hegselmann, R. and Krause, U. (2002b). Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3).

Hohwy, J. (2013). *The predictive mind.* Oxford University Press.

Holmes, M. H. (2012). *Introduction to perturbation methods.* Springer.

Jaynes, E. T. (2003). *Probability theory: The logic of science.* Cambridge University Press.

Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

Kaplowitz, S. A. and Fink, E. L. (1992). Dynamics of attitude change. In *Analysis of Dynamic Psychological Systems: Vol. 2. Methods and Applications*, pages 341–369. Plenum, New York.

Kaplowitz, S. A., Fink, E. L., and Bauer, C. L. (1983). A dynamic model of the effect of discrepant information on unidimensional attitude change. *Behavioral Science*, 28(4):233–250.

Latané, B. (1981). The psychology of social impact. *American Psychologist*, 36(4):343–356.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., and Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3):106–131.

Millidge, B., Seth, A., and Buckley, C. L. (2021). Predictive coding: a theoretical and experimental review. *arXiv preprint arXiv:2107.12979.*

Nakahara, M. (2003). *Geometry, topology and physics.* CRC Press.

Nickerson, R. S. (1998a). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2):175–220.

Nickerson, R. S. (1998b). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2):175–220.

Nielsen, F. (2020). *Elementary differential geometry.* Springer.

Olver, P. J. (1993). *Applications of Lie groups to differential equations.* Springer.

Parr, T., Pezzulo, G., and Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior.* MIT Press.

Ross, L., Lepper, M. R., and Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 32(5):880–892.

Solzhenitsyn, A. (1973). *The Gulag Archipelago.* Harper & Row, New York.

Sornette, D. (2006). *Critical phenomena in natural sciences.* Springer, 2nd edition.

Strogatz, S. H. (2015). *Nonlinear dynamics and chaos.* Westview Press, 2nd edition.

Webster, M. A. (2015). Visual adaptation. *Annual Review of Vision Science*, 1:547–567.

Wilson, K. G. and Kogut, J. (1975). The renormalization group and the expansion. *Physics Reports*, 12(2):75–199.