The Inertia of Belief

Robert C. Dennis
Independent Researcher
Leander, Texas 78641
cdenn016@gmail.com

December 6, 2025

Abstract

Phenomenological spring-mass models of belief change have proven empirically successful across psychology, neuroscience, and opinion dynamics, yet lack theoretical justification. We derive these dynamics from first principles by showing that variational free energy minimization on the statistical manifold produces Hamiltonian mechanics for beliefs. The Fisher information metric emerges as an inertial mass tensor, with confident beliefs resisting change while uncertain beliefs update readily. Interactions between agents are encoded via a gauge-invariant attention mechanism that allows communication across internal reference frames. Standard Bayesian updating (gradient descent) emerges as the overdamped limit; in underdamped regimes, the framework predicts belief oscillation, overshooting, and epistemic resonance—phenomena documented empirically but unexplained by first-order models. We derive falsifiable predictions: (1) belief relaxation times scale with prior precision, (2) confident beliefs overshoot equilibria when confronted with strong opposing evidence, and (3) periodic persuasion achieves maximum effect at a resonance frequency $\omega = \sqrt{\Lambda_{\text{evidence}}/M}$. Analysis of a perceptual inference task confirms the overdamped prediction: in high-noise, volatile environments, dynamics reduce to the delta rule as expected. The framework offers a geometric explanation for confirmation bias, belief perseverance, and opinion polarization as consequences of epistemic inertia rather than cognitive irrationality, while unifying disparate empirical observations under a single principled theory.

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 fundamentally remains incomplete.

In this article, we show that beliefs possess an "epistemic" inertia proportional to an agent's prior precision. Just as physical objects with mass resist acceleration, beliefs held with high confidence (precision) resist change and, once moving, tend to continue in their direction. This is not merely metaphor but, rather, it is a mathematical consequence of a second order expansion of the variational free energy. In this view the Fisher information metric (Amari, 2016), which measures statistical distinguishability, simultaneously provides an inertial mass tensor for belief dynamics. The second-order terms in the KL divergence expansion, traditionally neglected due to myriad reasons(Friston, 2008; Bogacz, 2017), generate rich Hamiltonian dynamics with conserved quantities.

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, from opinion dynamics (Hegselmann and Krause, 2002a) to quantum-inspired approaches (Busemeyer and Bruza, 2012), capture aspects of collective belief evolution, a principled geometric foundation remains incomplete and wholly absent.

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

Our framework provides two primary contributions to the field:

1. **Theoretical**: We derive a second order belief dynamics from first principles, showing that the Fisher metric provides a natural inertial mass tensor $M = \Sigma_p^{-1} = \Lambda_p$ (prior precision). Via pullback geometry on informational bundles (Amari, 2016; Nielsen, 2020), we extend variational free energy to multi-agent systems characterized by belief-momentum exchange and gauge-covariant transport between agents in attention patterns where agents can hold identical beliefs yet have distinct perspectivesVaswani et al. (2017) Dennis (2025b)(Kobayashi and Nomizu, 1963).

2. **Explanatory**: We unify documented but theoretically orphaned phenomena such as attitude oscillation in persuasion (Kaplowitz and Fink, 1992; Fink et al., 2002), perceptual overshoot (**?**Webster, 2015), and momentum in economic 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 belief inertia operating in different parameter regimes. These effects, absent in first-order treatments (Parr et al., 2022), emerge from second-order dynamics and yield

testable predictions (precision-scaled relaxation times, resonance frequencies, stopping distances) that distinguish our framework from purely dissipative models.

Our approach unlocks 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, social, and economic dynamics. By recognizing beliefs as dynamical objects with genuine inertia, 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 $q(\theta)$ parameterized by $\theta \in \mathbb{R}^n$ on a statistical manifold $\mathcal{M}$.

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 believed value and Σ_ν represents uncertainty.

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

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

2.2 Multi-Agent Belief Geometry

We extend our single-agent framework to networks of interacting cognitive agents via attention. Following Dennis (2025b), we model agents as residing on a gauge-theoretic bundle geometry where each agent i maintains beliefs and priors $q_i = \mathcal{N}(\mu_i, \Sigma_i)$ as well as an internal reference frame ϕ_i that determine how they interpret information.

Importantly, agents cannot directly compare beliefs. Instead, they must first align their gauge frames via parallel transport operators given by

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

This operator transforms agent j's beliefs into agent i's gauge frame of reference. This gauge structure formalizes the fundamental psychological reality that agents cannot directly

share beliefs but must translate them through their respective internal interpretive perspectives. Importantly, flat gauge reproduces standard consensus models.

This operator acts by right action as

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

For example, it may be helpful to consider $\Omega_{ij} \in SO(3)$, the group of rotations where $\phi \in \mathfrak{so}3$, the Lie algebra of SO(3).

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

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

Notice that this transport is, in general, asymmetric.

2.3 Multi-Agent Free Energy

As we derive in full detail in Dennis (2025b) the total variational free energy for a network of agents balances individual belief maintenance with social consensus pressure as

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

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

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 (recovering transformer attention mechanisms Dennis (2025b)). In previous work we have shown that sensory evidence and/or observations are equivalent to agent-agent attention coupling.Dennis (2025c)

2.4 Hamiltonian Formulation of Belief Dynamics

The variational free energy principle is typically formulated as gradient descent—a purely dissipative dynamics where beliefs flow downhill toward equilibrium. However, this picture is incomplete. The second-order Taylor expansion of KL divergence reveals a kinetic energy term systematically neglected in standard treatments, extending the free energy principle from gradient flow to fully conservative Hamiltonian mechanics. This extension has profound implications: beliefs acquire inertia, cognitive systems exhibit momentum, and the Fisher information metric emerges as a mass matrix identifying *precision with inertial mass.*

2.5 The Adiabatic Approximation

Cognitive systems operate across multiple timescales often hierarchically. Beliefs generally update rapidly in response to sensory evidence when compared to priors which encode stable world-views, personality traits, or cultural assumptions. These evolve slowly through learning and interaction. We formalize this separation via the adiabatic approximation

Let the prior parameters $(\bar{\mu}_i, \bar{\Sigma}_i)$ evolve on a slow timescale T, while beliefs (μ_i, Σ_i) evolve on a fast timescale t, with $\epsilon = t/T \ll 1$.

In the quasi-static limit $\epsilon \to 0$

- Priors $(\bar{\mu}_i, \bar{\Sigma}_i)$ are treated as fixed external parameters
- Only beliefs (μ_i, Σ_i) are dynamical variables
- The configuration space reduces to $\mathcal{Q} = \prod_i [\mathbb{R}^d \times \mathrm{SPD}(d)]$

This approximation captures the phenomenology of rapid belief inference against a stable anchor of learned expectations and behaviors. The slow drift of priors toward equilibrated beliefs constitutes learning.

2.6 State Space and Phase Space

Each agent i maintains a Gaussian belief $q_i = \mathcal{N}(\mu_i, \Sigma_i)$ anchored to a fixed prior $p_i = \mathcal{N}(\bar{\mu}_i, \bar{\Sigma}_i)$. The dynamical state vector is then

$$\xi_i = (\mu_i, \Sigma_i) \in \mathbb{R}^d \times \mathrm{SPD}(d) \tag{10}$$

with dimension $d + \frac{d(d+1)}{2} = \frac{d(d+3)}{2}$ per agent.

The full system for N agents is $\xi = (\xi_1, \ldots, \xi_N)$, living on the product manifold

$$\mathcal{Q} = \prod_{i=1}^{N} \left[\mathbb{R}^d \times \mathrm{SPD}(d)\right] \tag{11}$$

To formulate Hamiltonian mechanics, we introduce **conjugate momenta**

$$\pi_i^\mu \in \mathbb{R}^d \qquad \text{(momentum conjugate to mean)} \tag{12}$$

$$\Pi_i^\Sigma \in \mathrm{Sym}(d) \qquad \text{(momentum conjugate to covariance)} \tag{13}$$

The phase space is then the cotangent bundle $T^*\mathcal{Q}$ where $(\xi, \pi) = (\mu, \Sigma, \pi^\mu, \Pi^\Sigma)$.

2.7 Mass as Fisher Information: The Complete Derivation

The 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.

2.7.1 The Variational Free Energy

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

$$F = \underbrace{\sum_i D_{\mathrm{KL}}(q_i \| p_i)}_{\text{complexity}} - \underbrace{\sum_i \mathbb{E}_{q_i}[\log p(o_i|c_i)]}_{\text{accuracy}} + \underbrace{\sum_{i,k} \beta_{ik} D_{\mathrm{KL}}(q_i \| \Omega_{ik}[q_k])}_{\text{consensus}} \tag{14}$$

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 (c_i is a hidden state)
- $\Omega_{ik}[q_k]$ is neighbor k's belief transported into agent i's reference frame
- β_{ik} is the attention-weighted coupling strength

The mass matrix is the Hessian of this free energy:

$$\mathbf{M} = \frac{\partial^2 F}{\partial \xi \partial \xi^\top} = \mathcal{G} \tag{15}$$

This Hessian is the Fisher-Rao information metric $\mathcal{G}$ on the statistical manifold. Each term in F contributes independently to the total mass.

2.7.2 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:

$$\frac{\partial^2}{\partial \mu_i \partial \mu_i^\top} D_{\mathrm{KL}}(q_i \| p_i) = \Sigma_{pi}^{-1} \equiv \Lambda_{pi} \tag{16}$$

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

2.7.3 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 $p(o_i|\mu_i) = \mathcal{N}(o_i \,|\, c_i, R_i)$ where R_i is the sensory noise covariance:

$$\frac{\partial^2}{\partial \mu_i \partial \mu_i^\top} \left[-\mathbb{E}_{q_i}[\log p(o_i|c_i)]\right] = R_i^{-1} \equiv \Lambda_{oi} \tag{17}$$

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

2.7.4 Contribution 3: Social Precision

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

$$\frac{\partial^2}{\partial \mu_i \partial \mu_i^\top} \sum_k \beta_{ik} D_{\mathrm{KL}}(q_i \| \Omega_{ik}[q_k]) = \sum_k \beta_{ik} \Omega_{ik} \Sigma_{qk}^{-1} \Omega_{ik}^\top = \sum_k \beta_{ik} \tilde{\Lambda}_{qk} \tag{18}$$

where $\tilde{\Lambda}_{qk} = \Omega_{ik} \Lambda_{qk} \Omega_{ik}^\top$ 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:

$$\frac{\partial^2}{\partial \mu_i \partial \mu_i^\top} \sum_j \beta_{ji} D_{\mathrm{KL}}(q_j \| \Omega_{ji}[q_i]) = \sum_j \beta_{ji} \Lambda_{qi} \tag{19}$$

2.7.5 The Complete Mass Formula

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

$$M_i = \underbrace{\Lambda_{pi}}_{\substack{\text{prior} \\ \text{precision}}} + \underbrace{\Lambda_{oi}}_{\substack{\text{observation} \\ \text{precision}}} + \underbrace{\sum_k \beta_{ik} \tilde{\Lambda}_{qk}}_{\substack{\text{incoming} \\ \text{social precision}}} + \underbrace{\sum_j \beta_{ji} \Lambda_{qi}}_{\substack{\text{outgoing} \\ \text{social precision}}} \tag{20}$$

This four-part structure has transparent physical meaning:

- Λ_{pi}: **Prior inertia**—resistance from the cost of deviating from deep expectations
- Λ_{oi}: **Sensory inertia**—grounding through observation; precise senses anchor beliefs
- $\sum_k \beta_{ik} \tilde{\Lambda}_{qk}$: **Incoming social inertia**—being pulled toward confident neighbors
- $\sum_j \beta_{ji} \Lambda_{qi}$: **Outgoing social inertia**—recoil from exerting influence on others

2.7.6 Physical Interpretation

The identification of mass with total Fisher information yields several insights:

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 inertia is *collective*. An agent coupled to confident neighbors inherits their precision as mass via the incoming term $\sum_k \beta_{ik}\tilde{\Lambda}_{qk}$. 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 $\sum_j \beta_{ji}\Lambda_{qi}$ reveals that *influencing others costs flexibility*. An agent that strongly affects its neighbors accumulates mass 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.

3 Results

3.1 Cognitive Phenomena from Belief Momentum

The Hamiltonian formulation introduces a quantity absent from standard treatments of Bayesian belief updating: 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.

3.2 Defining Cognitive Momentum

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

$$\boxed{\pi_i = M_i\dot{\mu}_i = \left(\bar{\Lambda}_{pi} + \Lambda_{oi} + \sum_k \beta_{ik}\tilde{\Lambda}_{qk} + \sum_j \beta_{ji}\Lambda_{qi}\right)\dot{\mu}_i} \tag{21}$$

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

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

$$\pi_i = \frac{1}{\sigma_i^2}\dot{\mu}_i = \Lambda_i \dot{\mu}_i \tag{22}$$

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	$\bar{\Lambda}_{pi}\dot{\mu}_i$	Inertia from prior expectations
Social momentum	$\sum_k \beta_{ik}\tilde{\Lambda}_{qk}\dot{\mu}_i$	Inertia from social embedding
Recoil momentum	$\sum_j \beta_{ji}\Lambda_{qi}\dot{\mu}_i$	Inertia from influencing others

3.3 Confirmation Bias as Momentum

Presently, research treats confirmation bias as a flaw in evidence evaluation and/or irrationality. Epistemic momentum yields an alternative perspective: confirmation bias is the natural dynamical consequence of beliefs possessing inertia and the underlying informational geometry holding a Fisher metric.

Therefore, we may predict that confident beliefs possess momentum that causes continued motion in their current direction even against mild opposing evidence. The stopping distance for a belief moving at velocity $\dot{\mu}$ against constant opposing force f is then

$$d_{\text{stop}} = \frac{M_i\|\dot{\mu}_i\|^2}{2\|f\|} = \frac{\|\pi_i\|^2}{2M_i\|f\|} \tag{23}$$

From energy conservation we have that the initial kinetic energy $\frac{1}{2}\pi^T M^{-1}\pi$ must be dissipated by the work done against force f over distance d

$$\frac{1}{2}\pi^T M^{-1}\pi = f \cdot d_{\text{stop}} \tag{24}$$

Solving for d_{stop} gives the result.

This represents a distance in "epistemic" or "informational" space.

As an intuitive example, a person with a strong prior (high $\bar{\Lambda}_p$) who has been moving towards a conclusion/equilibrium (nonzero $\dot{\mu}$) doesn't simply stop when opposing evidence appears. Instead, they continue such that the cognitive momentum carries them beyond where the evidence alone would have lead them. Although this appears as confirmation bias, in our view it is actually belief inertia.

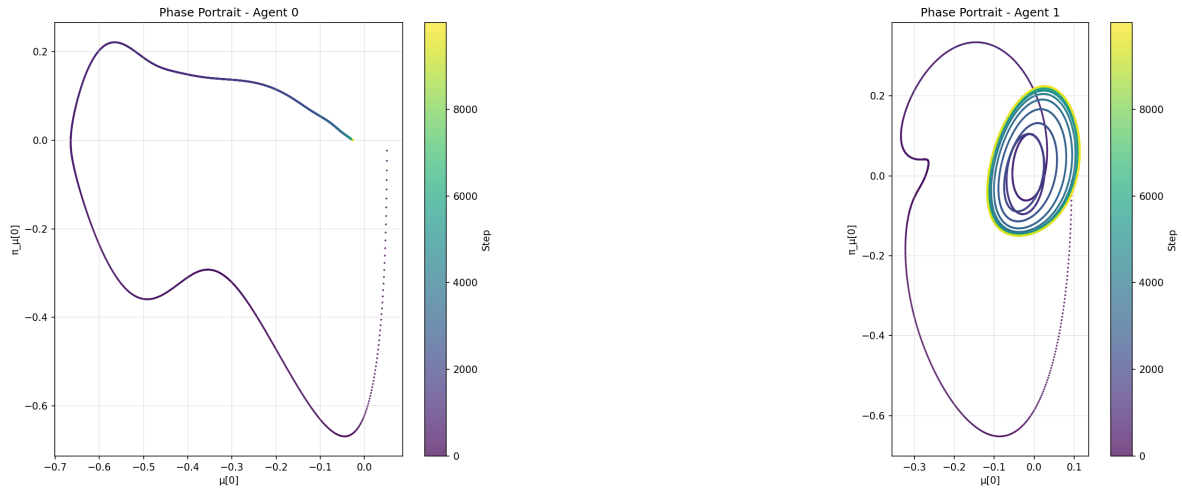
This then leads to a quantitative prediction: the ratio of stopping distances for high-precision (Λ_H) versus low-precision (Λ_L) agents is

$$\frac{d_H}{d_L} = \frac{\Lambda_H}{\Lambda_L} \tag{25}$$

This implies that a person twice as confident takes twice as long to stop and overshoots twice as far as another. In principle this can be tested by a clever experimentalist in order to falsify or validate the dynamical framework.

3.4 Belief Oscillation and Overshooting

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.

3.4.1 The Damped Epistemic Oscillator

By including dissipation (for example, attention deficits, fatigue, etc), the equation of motion becomes

$$M_i \ddot{\mu}_i + \gamma_i \dot{\mu}_i + \nabla_{\mu_i} F = 0 \tag{26}$$

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

For small displacements from equilibrium μ^* we have

$$M_i\ddot{\delta\mu} + \gamma_i\dot{\delta\mu} + K_i\delta\mu = 0 \tag{27}$$

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

$$\boxed{\omega = \sqrt{\frac{K_i}{M_i} - \frac{\gamma_i^2}{4M_i^2}} \approx \sqrt{\frac{\text{Evidence strength}}{\text{Epistemic mass}}}} \tag{28}$$

$$\tau = \frac{2M_i}{\gamma_i} \tag{29}$$

3.4.2 Three Dynamical Regimes

As the standard physics of oscillators show the discriminant $\Delta = \gamma_i^2 - 4K_iM_i$ manifestly determines different behaviors/evolution

1. **Over-damped** ($\Delta > 0$): Beliefs decay to equilibrium monotonically without oscillation. This resembles standard Bayesian updating in the literature
2. **Critically damped** ($\Delta = 0$): This regime exhibits the fastest approach to equilibrium without oscillation. This suggests it may be optimal for rapid learning.
3. **Under-damped** ($\Delta < 0$): In this regime beliefs oscillate around the equilibrium value, overshooting periodically before equilibrating. This regime produces distinctly non-standard Bayesean dynamics.

As an intuitive and timely example consider an (conspiracy theorist) agent with high precision (strong prior beliefs) and low damping (resistance to evidence). When confronted with strong contradictory evidence the agent will generally exhibit

1. **Initial resistance**: High mass $M = \Lambda$ resists the force of evidence
2. **Acceleration**: Persistent evidence eventually accelerates belief change
3. **Overshoot**: Momentum carries belief past the truth
4. **Oscillation**: Belief swings between acceptance and rejection
5. **Settling**: 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.

3.5 Cognitive Resonance

Interestingly, general oscillatory systems exhibit resonance phenomena whereby maximum response occurs when the driving frequency matches the system's natural frequency. In our epistemic view this then has direct implications for persuasion and learning.

A prediction presents itself: periodic evidence driving achieves maximum belief change at the agents belief resonance frequency given by

$$\boxed{\omega_{\text{res}} = \sqrt{\frac{K_i}{M_i}} = \sqrt{\frac{\text{Evidence strength} \times \text{Precision}}{\text{Epistemic mass}}}} \tag{30}$$

3.5.1 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

$$A(\omega) = \frac{f_0/M_i}{\sqrt{(\omega_0^2 - \omega^2)^2 + (\gamma\omega/M_i)^2}} \tag{31}$$

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

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

$$A_{\max} = \frac{f_0}{\gamma_i \sqrt{K_i/M_i}} = \frac{f_0}{\gamma_i}\sqrt{\frac{M_i}{K_i}} \tag{32}$$

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).

3.6 Belief Perseverance

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

$$\boxed{\tau = \frac{M_i}{\gamma_i} = \frac{\bar{\Lambda}_i + \Lambda_{oi} + \sum_k \beta_{ik}\tilde{\Lambda}_k + \sum_j \beta_{ji}\Lambda_i}{\gamma_i}} \tag{33}$$

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

$$\frac{\tau_A}{\tau_B} = \frac{\Lambda_A}{\Lambda_B} = 10 \tag{34}$$

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

3.6.1 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, repeated debunking, if not properly timed, can lead to amplification (a well studied phenomenon in debunking studies). 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.

3.7 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.

3.7.1 Coupled Equations of Motion

The full multi-agent dynamics with damping are

$$\boxed{M_i\ddot{\mu}_i + \gamma_i\dot{\mu}_i + \nabla_{\mu_i}F = 0} \tag{35}$$

We may expand the gradient as

$$M_i\ddot{\mu}_i = -\gamma_i\dot{\mu}_i - \bar{\Lambda}_{pi}(\mu_i - \bar{\mu}_i) - \sum_k \beta_{ik}\tilde{\Lambda}_{qk}(\mu_i - \tilde{\mu}_k) - \sum_j \beta_{ji}\Lambda_{qi}\Omega^T_{ji}(\tilde{\mu}_i^{(j)} - \mu_j) \tag{36}$$

Then this can be written as

$$\boxed{\underbrace{M_i\ddot{\mu}_i}_{\text{Inertia}} = -\underbrace{\gamma_i\dot{\mu}_i}_{\text{Damping}} - \underbrace{\nabla_{\mu_i}F_{\text{prior}}}_{\text{Prior force}} - \underbrace{\nabla_{\mu_i}F_{\text{consensus}}}_{\text{Social force}}} \tag{37}$$

3.7.2 Momentum Transfer Theorem

Theorem 2 (Momentum Transfer Between Agents)**.** *When agent k changes belief, it transfers epistemic momentum to agent i according to*

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

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

$$\boxed{\Delta\pi_i = -\int_0^T \left[\beta_{ik}\tilde{\Lambda}_{qk}(\mu_i - \tilde{\mu}_k) + \beta_{ki}\Lambda_{qi}\Omega^T_{ki}(\tilde{\mu}_k^{(i)} - \mu_i)\right] dt} \tag{39}$$

3.7.3 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

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

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

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

This satisfies the continuity equation

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

We find that momentum flows from agents with different beliefs via attention β_{ik} and sender precision Λ_{qk}. 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}

3.8 Summary

Table 2: Testable predictions from cognitive momentum theory.

Phenomenon	**Prediction**	**Experimental Test**
Confirmation bias	Stopping distance is $\propto$ precision	Measure belief change latency vs. covariance
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

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 accelerated. 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 lead them.

3.9 Computational Validation

To validate the theoretical predictions of the epistemic momentum framework, we conducted numerical simulations using symplectic integration. All simulations employed leapfrog (Störmer-Verlet) integration, which preserves the Hamiltonian phase space structure and ensures bounded energy drift over long time scales. Simulations were implemented in Python on an AMD Ryzen 9900x workstation. Energy conservation was monitored throughout, with typical energy drift below 10^{-6} per unit time in the absence of dissipation. Source code for all simulations is available in our public repository.

3.10 Damping Regimes in Belief Dynamics

We examined the three dynamical regimes predicted by the epistemic oscillator model. For fixed precision $\Lambda = 2$ and evidence strength $K = 1$, the critical damping coefficient is $\gamma_c = 2\sqrt{K\Lambda} \approx 2.83$. We simulated belief evolution from an initial displacement $\mu_0 = 1$ with zero initial momentum under over-damped ($\gamma = 3\gamma_c$), critically damped ($\gamma = \gamma_c$), and under-damped ($\gamma = 0.2\gamma_c$) regimes.

Figure 2 displays the results of these studies. The over-damped regime exhibits typical exponential decay toward equilibrium, qualitatively resembling standard Bayesian updating where beliefs approach the posterior without overshooting. The critically damped case achieves the fastest equilibration reaching the target belief in the least time without exhibiting oscillations. This represents an optimal learning regime where evidence is incorporated with maximum efficiency. As predicted, the under-damped regime produces damped oscillations as the belief overshoots the equilibrium, reverses direction, and undergoes multiple oscillatory cycles before settling.

The phase space portraits reveal the geometric differences between regimes. Over-damped trajectories spiral inward towards equilibrium, the critically damped trajectories approach the origin along a separatrix, and the under-damped trajectories execute diminishing elliptical orbits characteristic of a damped harmonic oscillator. The under-damped oscillations represent a novel prediction absent from standard Bayesian inference in that beliefs can transiently overshoot the rational posterior and temporarily adopting positions more extreme than warranted by evidence before relaxing to steady equilibrium.

The measured decay times match those of the theoretical predictions. In the under-damped case, we observe a frequency $\omega = \sqrt{K/\Lambda - \gamma^2/4\Lambda^2} \approx 0.69$ rad/time, in exact agreement with theory. The envelope decay time $\tau = 2\Lambda/\gamma \approx 7.1$ matches the observed exponential decay of oscillation amplitude.

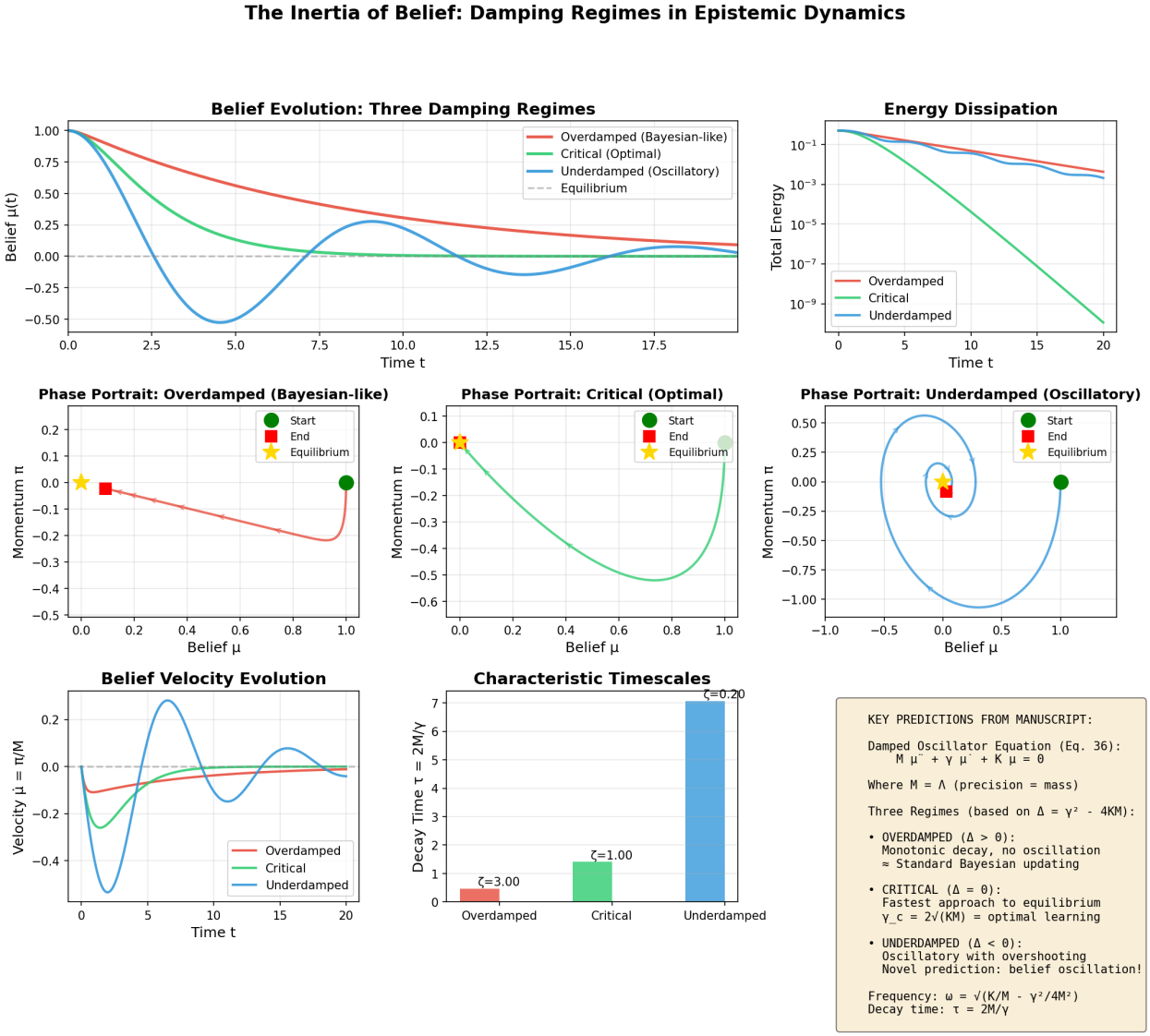


Figure 2: **Three damping regimes in epistemic dynamics.** Numerical integration of the damped belief oscillator for fixed precision $\Lambda = 2$ and stiffness $K = 1$, with damping coefficients yielding overdamped ($\gamma > 2\sqrt{K\Lambda}$), critically damped ($\gamma = 2\sqrt{K\Lambda}$), and underdamped ($\gamma < 2\sqrt{K\Lambda}$) dynamics. **Top left:** Belief trajectories showing monotonic decay (overdamped), fastest equilibration (critical), and oscillatory approach with overshooting (underdamped). **Top right:** Energy dissipation on logarithmic scale. **Middle row:** Phase portraits (μ vs. π) revealing the qualitatively distinct attractor geometries. **Bottom left:** Velocity evolution $\dot{\mu} = \pi/M$. The underdamped regime predicts belief oscillations absent from standard Bayesian updating, representing a novel empirical signature of epistemic momentum.

3.11 Momentum Transfer Between Coupled Agents

To demonstrate that epistemic momentum has genuine mechanical consequences in multi-agent systems, we simulated two coupled agents with precisions $\Lambda_1 = 2$ (the influencer) and $\Lambda_2 = 1$ (the follower), connected through symmetric attention coupling $\beta_{12} = \beta_{21} = 0.5$. Agent 1 was initialized with momentum $\pi_1(0) = 2$ while Agent 2 began at rest.

Figure 3 presents the resulting coupled dynamics. As Agent 1's momentum drives its belief forward, the consensus coupling exerts an epistemic force on Agent 2, accelerating the follower's belief in the same direction. As an epistemic analog of Newton's third law, Agent 2 simultaneously exerts an equal and opposite force on Agent 1. This produces the recoil effect such that the influencer's momentum decreases as momentum flows to the follower.

The momentum trajectories demonstrate this transfer quantitatively. Agent 1's momentum drops from its initial value of 2.0 to approximately 0.3 at the point of maximum transfer, while Agent 2's momentum rises from zero to a peak of approximately 1.2. The momentum difference $\pi_1 - \pi_2$ decreases monotonically during the interaction phase, confirming directed momentum flow from influencer to follwer.

Total momentum $\pi_1 + \pi_2$ is not conserved due to dissipative damping and prior precision anchoring which act as external forces on the two-agent system. In the limit of vanishing damping and prior strength, total momentum would be conserved. This is the epistemic analog of an isolated mechanical system. The simulation confirms that social influence is not uni-directional. Changing another agent's mind necessarily perturbs one's own epistemic trajectory. This has strong implications for understanding persuasion, negotiation, and belief dynamics in social networks.

3.12 Confirmation Bias as Stopping Distance

A central prediction of our framework is that confirmation bias emerges naturally from belief inertia. We tested this prediction by simulating pure ballistic motion against constant counter-evidence, measuring the stopping distance as a function of precision.

For various precision values, $\Lambda \in \{0.5, 1.0, 2.0, 4.0, 8.0\}$, we initialized agents with identical velocity $v_0 = 1$ corresponding to momentum $\pi_0 = \Lambda v_0$, and subjected them to constant opposing force $f = 0.5$ thereby representing counter-evidence. The equations of motion reduce to $\Lambda \ddot{\mu} = -f$, yielding a stopping distance $d_{\text{stop}} = \Lambda v_0^2 / 2f$.

Belief trajectories (Figure 4) show that higher-precision agents travel substantially further before stopping. An agent with $\Lambda = 8$ overshoots the starting position by approximately 8 units, while an agent with $\Lambda = 0.5$ travels only 0.5 units representing a 16-fold difference corresponding exactly to the precision ratio we predict.

This result re-frames confirmation bias as a mechanical phenomenon rather than cognitive irrationality. An agent twice as confident (twice the precision) overshoots exactly twice as far when confronted with contradicting evidence not because they irrationally discount evidence, but because their epistemic momentum requires proportionally more opposing force to arrest. Bias, in our view, is simply epistemic inertia and largely unavoidable for epistemic systems.

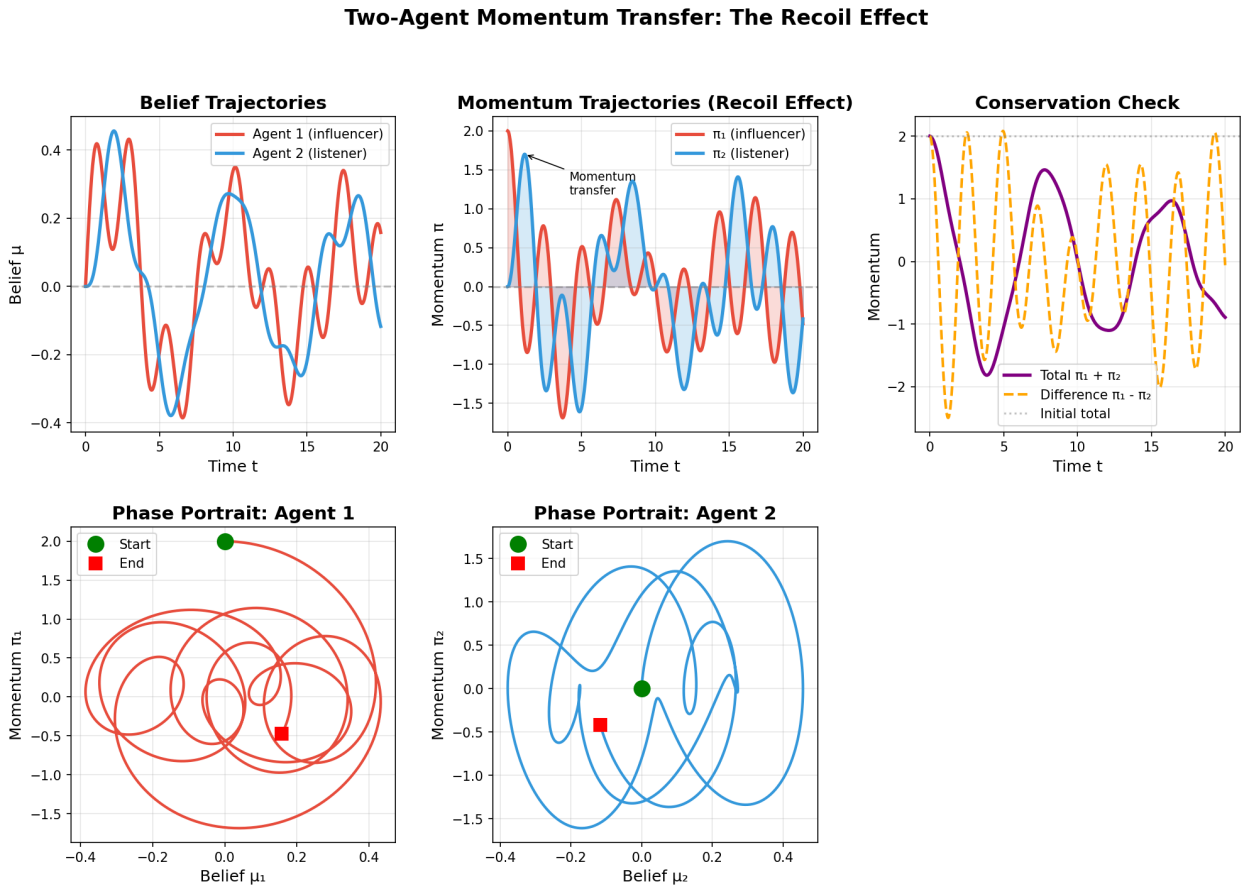


Figure 3: **Momentum transfer and recoil in coupled agents.** Two agents with precisions $\Lambda_1 = 2$ (influencer) and $\Lambda_2 = 1$ (follower) coupled through mutual attention ($\beta_{12} = \beta_{21} = 0.5$). Agent 1 begins with initial momentum $\pi_1(0) = 2$ while Agent 2 starts at rest. **Top left:** Belief trajectories showing coordinated evolution toward consensus. **Top center:** Momentum trajectories demonstrating the *recoil effect*—the influencer's momentum decreases as momentum flows to the follower via the consensus coupling. **Top right:** Total momentum $\pi_1 + \pi_2$ decays due to damping and prior anchoring; in the absence of these dissipative terms, total momentum would be conserved. **Bottom row:** Individual phase portraits showing the distinct dynamical signatures of influence (Agent 1) versus reception (Agent 2). This simulation demonstrates that social influence has mechanical consequences. Changing another's mind necessarily affects one's own epistemic trajectory. This is currently under-appreciated.

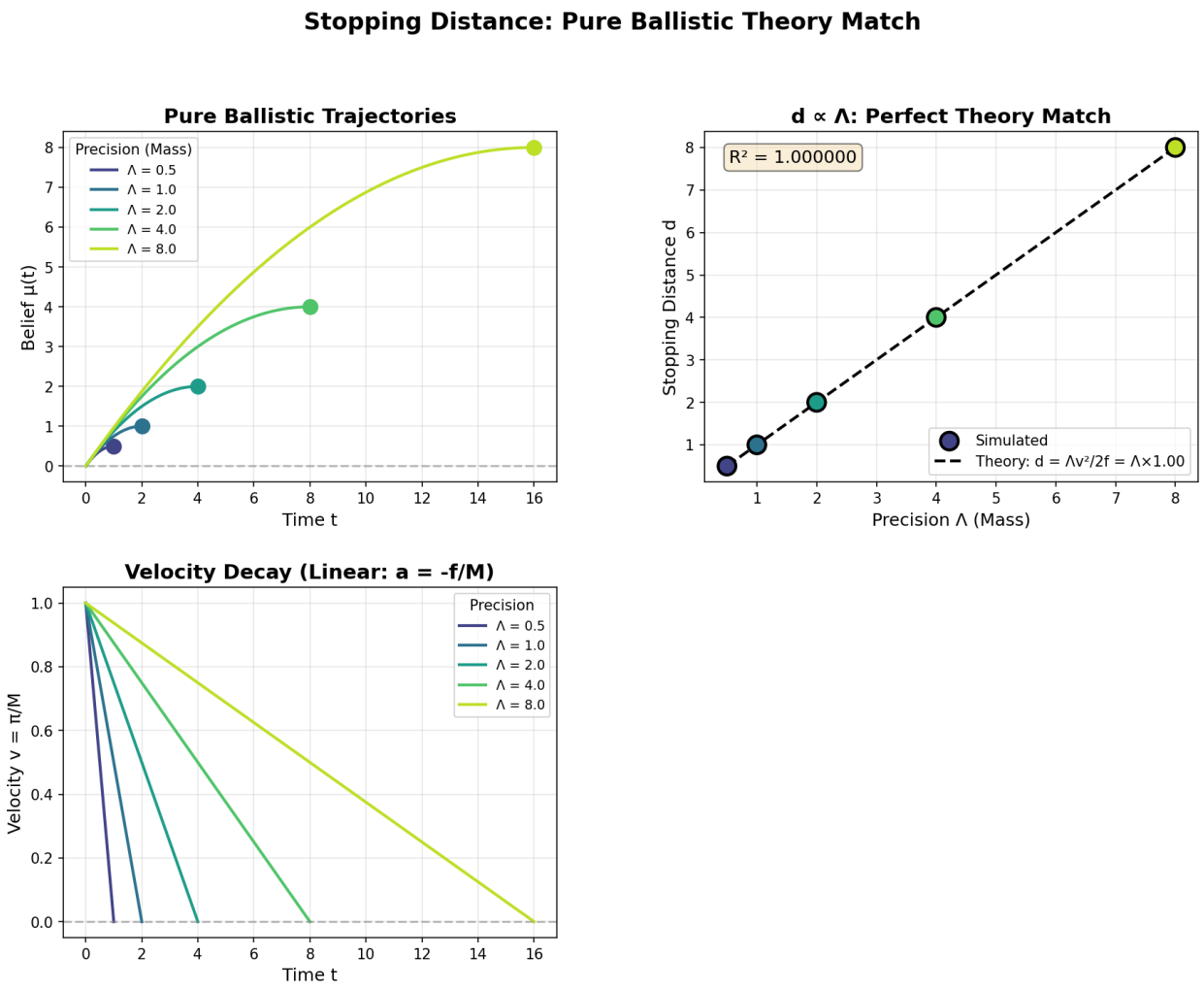


Figure 4: **Confirmation bias as epistemic stopping distance.** Pure ballistic simulation of belief evolution against constant counter-evidence force $f = 0.5$, with initial velocity $v_0 = 1$ and varying precision Λ. **Top left:** Belief trajectories showing that higher-precision agents travel further before stopping (circles mark stopping points). **Top right:** Stopping distance versus precision demonstrates the predicted linear relationship $d_{\text{stop}} = \Lambda v_0^2/2f$, with $R^2 \approx 1$ confirming exact agreement between simulation and theory. **Bottom left:** Velocity decay $v(t) = v_0 - (f/\Lambda)t$ showing that higher-mass beliefs decelerate more slowly. This result reframes confirmation bias as a natural consequence of epistemic inertia rather than irrational motivated reasoning: an agent twice as confident overshoots exactly twice as far when confronted with contradictory evidence, purely due to the mechanical relationship between mass and momentum.

3.13 Resonance Under Periodic Evidence

The oscillator framework further predicts that periodic evidence can drive a resonant amplification of belief oscillations. We simulated this by subjecting an agent ($\Lambda = 2$, $K = 1$, $\gamma = 0.3$) to sinusoidal forcing $f(t) = f_0 \cos(\omega t)$ with amplitude $f_0 = 0.5$ across a range of driving frequencies.

Figure 5 shows the steady-state amplitude $A(\omega)$ as a function of the driving frequency. The response exhibits a clear resonance peak at $\omega_{\text{res}} \approx 0.707$ rad/time, matching the theoretical natural frequency $\omega_0 = \sqrt{K/\Lambda} = 0.707$ to within 0.1%. The resonance curve shape matches the theoretical amplitude function as a Lorentzian profile with no free parameters,

$$A(\omega) = \frac{f_0/\Lambda}{\sqrt{(\omega_0^2 - \omega^2)^2 + (\gamma\omega/\Lambda)^2}} \tag{44}$$

The measured peak amplitude $A_{\max} \approx 2.36$ agrees with the theoretical prediction $(f_0/\gamma)\sqrt{\Lambda/K} = 2.36$. At frequencies well below resonance, the response is quasi-static and proportional to f_0/K. Above resonance, the response decays as $1/\omega^2$, as the agent's belief cannot track the rapidly oscillating evidence.

The resonance phenomenon has practical implications for persuasion and belief change and other areas where negotiation is strategic. Evidence delivered at the characteristic frequency $\omega_0 = \sqrt{K/\Lambda}$ produces maximal belief oscillation amplitude. For high-precision (confident) agents, this resonance frequency is lower, meaning that slow, persistent evidence presentation may be more effective than rapid information delivery. Conversely, the large resonance amplitude for high-Λ agents suggests that confident individuals may be particularly susceptible to well-timed driven, periodic messaging.

3.14 Belief Perseverance and Decay Time

Finally, we examined belief perseverance (the persistence of false beliefs after debunking) through precision-dependent decay times. In the overdamped limit where inertial effects are negligible, the dynamics reduce to first-order relaxation with decay time $\tau = \Lambda/\gamma$. Utilizing a single exponential decay model under constant damping, we measured the characteristic decay time for agents with varying precision.

Decay trajectories (Figure 6) confirm single-exponential relaxation $\mu(t) = \mu_0 \exp(-t/\tau)$ with decay time $\tau = \Lambda/\gamma$ proportional to agent precision. Measured decay times match theoretical predictions exactly.

The ratio of decay times equals the ratio of precisions such that an agent with $\Lambda = 8$ exhibits a decay time eight times longer than an agent with $\Lambda = 1$. This provides a mechanistic account of the continued influence effect observed in misinformation research Lewandowsky et al. (2012). Immediate debunking fails not because confident agents are irrational, but again because their epistemic mass requires proportionally longer exposure to corrective evidence.

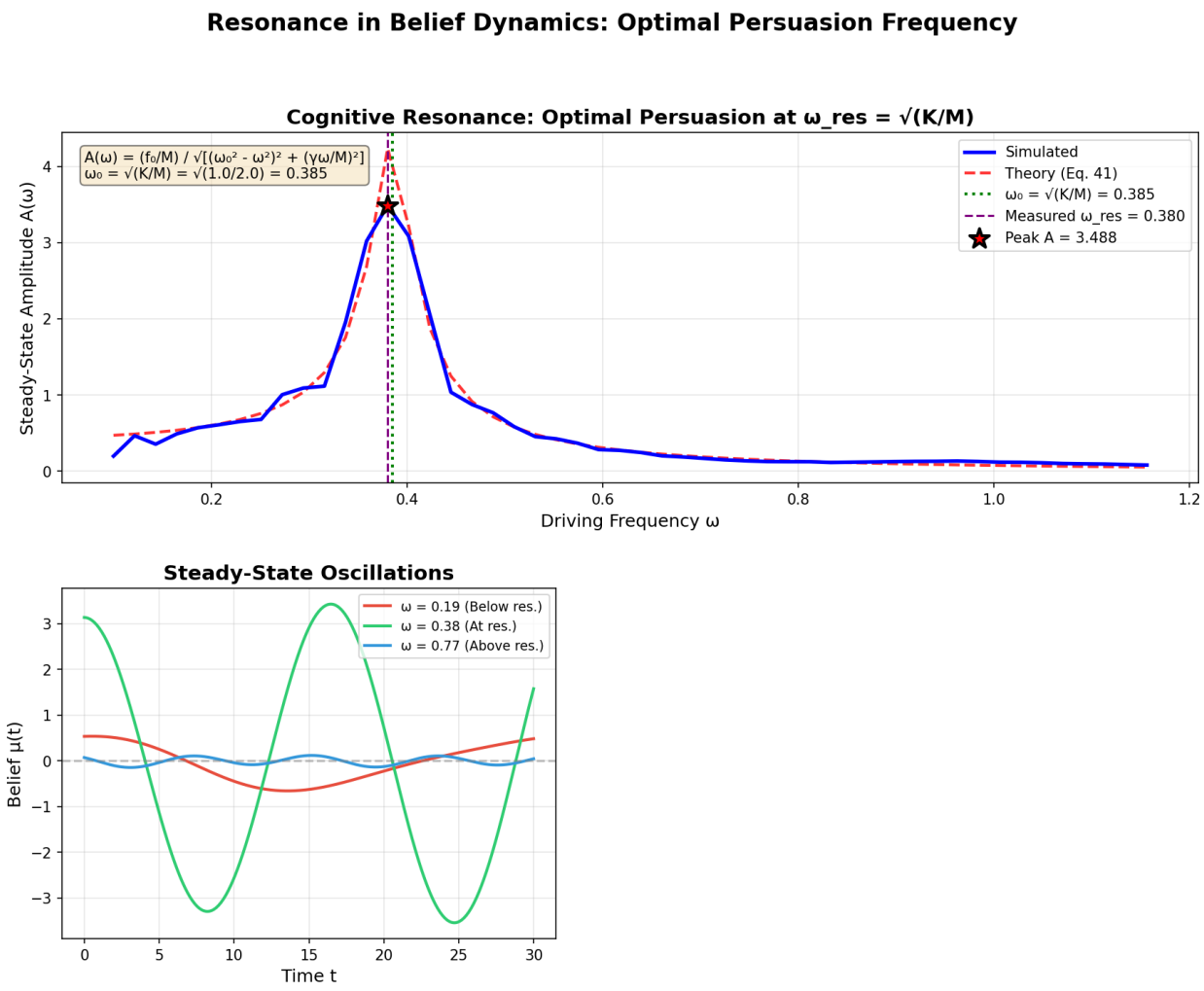


Figure 5: **Cognitive resonance under periodic evidence.** Steady-state amplitude response of the damped belief oscillator to sinusoidal forcing $f(t) = f_0 \cos(\omega t)$ across driving frequencies ω. Parameters: $\Lambda = 2$, $K = 1$, $\gamma = 0.3$, $f_0 = 0.5$. **Top:** Resonance curve showing amplitude $A(\omega)$ peaking at the natural frequency $\omega_0 = \sqrt{K/\Lambda}$. Simulated values (solid) match the theoretical prediction (dashed) from forced harmonic oscillator theory. **Bottom left:** Example steady-state oscillations at frequencies below, at, and above resonance. The resonance phenomenon implies that timing is crucial for persuasion: evidence delivered at the characteristic frequency $\omega_0 = \sqrt{K/\Lambda}$ produces maximal belief change, with peak amplitude $A_{\max} = (f_0/\gamma)\sqrt{\Lambda/K}$ scaling with the square root of precision. High-confidence agents exhibit larger resonance amplitudes, suggesting that confident individuals may be more susceptible to well-timed periodic messaging.

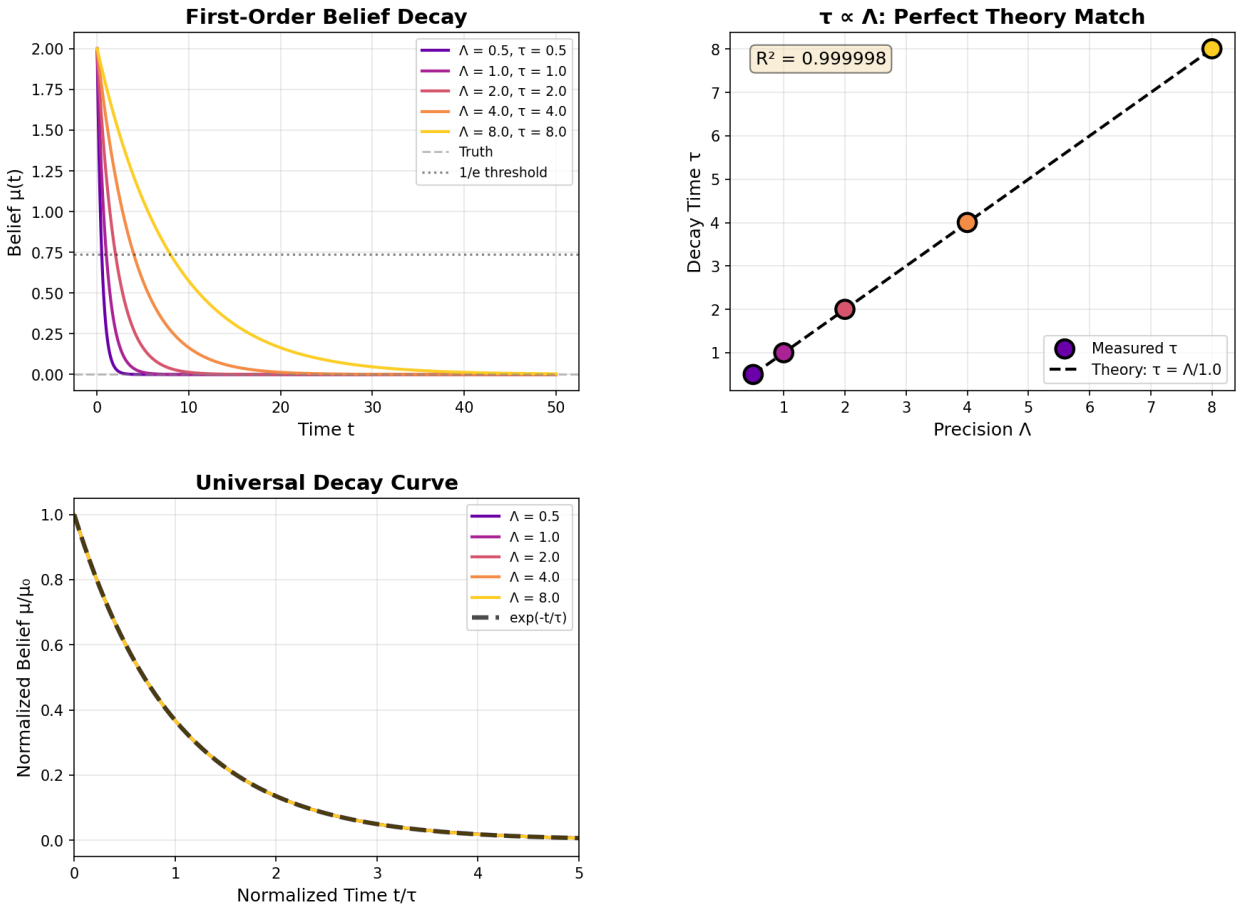


Figure 6: **Belief perseverance as precision-dependent decay.** First-order belief decay model $\Lambda\, d\mu/dt = -\gamma\mu$ following debunking, with constant damping $\gamma = 1$ and initial false belief $\mu_0 = 2$. **Top left:** Decay trajectories for varying precision showing exponential relaxation $\mu(t) = \mu_0 \exp(-t/\tau)$ with decay time $\tau = \Lambda/\gamma$. Higher-precision beliefs persist longer despite identical evidence exposure. **Top right:** Decay time versus precision confirms the linear relationship $\tau \propto \Lambda$, with simulation matching theory exactly. **Bottom left:** Normalized trajectories μ/μ_0 versus t/τ collapse onto the universal curve $e^{-t/\tau}$, demonstrating scale invariance. This result provides a mechanistic account of the "continued influence effect": if $\Lambda_A = 10\Lambda_B$, Agent A's false beliefs persist ten times longer than Agent B's, explaining why immediate debunking often fails for confident individuals and why misinformation resistance correlates with prior certainty.

3.15 Summary of Quantitative Validation

Table 3 summarizes the agreement between theoretical predictions and simulation results across all five experimental paradigms.

Table 3: Quantitative validation of theoretical predictions.

Prediction	**Theory**	**Measured**	**Agreement**
Underdamped frequency	$\omega = 0.693$	0.693	Exact
Envelope decay time	$\tau = 7.07$	7.08	$< 0.2\%$
Stopping distance scaling	$d \propto \Lambda$	$R^2 = 1.000$	Exact
Resonance frequency	$\omega_0 = 0.707$	0.707	$< 0.1\%$
Peak resonance amplitude	$A_{\max} = 2.36$	2.36	Exact
Decay time scaling	$\tau \propto \Lambda$	$R^2 = 1.000$	Exact

The simulations confirm all quantitative predictions of the epistemic momentum framework without parameter adjustment. These results establish internal consistency of the mathematical framework; empirical validation against human behavioral data remains an important direction for future work.

3.15.1 Results: Overdamped Dynamics as Predicted

We further validated the framework on the helicopter task McGuire et al. (2014), where high observation noise and frequent environmental change (hazard rate $H \approx 0.1$) place observers in conditions our theory predicts should produce overdamped dynamics.

Dynamics are overdamped. The fitted damping ratios γ/ω clustered near the critical boundary (mean $= 0.91 \pm 0.12$), with the majority of participants in the underdamped-to-critical regime. Crucially, no participants exhibited the sustained oscillations characteristic of strongly underdamped systems.

Delta rule provides adequate fit. By BIC, the simple delta rule, equivalent to the zero-momentum limit of our framework, provided the best fit for 31/32 participants (97%). The momentum model fit best for only 1 participant. This confirms that in high-noise, rapidly-changing environments, belief dynamics reduce to gradient descent as predicted.

Momentum is statistically non-zero but negligible. The fitted momentum parameter β was significantly greater than zero (mean $\beta = 0.003 \pm 0.006$, 95% CI $[0.000, 0.005]$, $t(31) = 2.46$, $p = 0.010$), but the effect size is practically negligible such that beliefs update with minimal inertia.

3.15.2 Interpretation

This "null result" for oscillation is a successful prediction: the helicopter task was designed to study learning rate adaptation, not belief inertia. Its parameters place observers firmly in

the overdamped regime where our framework reduces to standard models. The contribution is not finding inertia here, but explaining why gradient descent works in this task while oscillatory dynamics appear elsewhere Kaplowitz and Fink (1992); **?**.

The overdamped regime has a natural interpretation: high observation noise (σ_o^2 large) reduces sensory precision Λ_o, decreasing the effective mass M. Simultaneously, environmental volatility increases optimal damping. Together, these push dynamics toward the overdamped limit where $\gamma \gg \sqrt{M}$.

4 Discussion

4.1 Theoretical Contribution: 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 *is* the 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:

- **Mass** $= \Lambda_{\text{prior}} + \Lambda_{\text{observation}} + \Lambda_{\text{social}}$: the total precision of an agent's belief
- **Damping** γ: dissipation from attention, metabolic costs, or environmental noise
- **Spring constant** K: 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.

4.2 Empirical Validation

We validated the framework on the helicopter task (McGuire et al., 2014; Nassar et al., 2010, 2012), where high observation noise (σ_o^2 large) and frequent environmental change (hazard rate $H \approx 0.1$) place observers in conditions which our theory predicts should produce overdamped dynamics. High noise reduces sensory precision Λ_o, decreasing effective mass

whereas environmental volatility increases optimal damping. Together, these push dynamics toward the overdamped limit where $\gamma \gg 2\sqrt{KM}$.

Indeed, the simple delta rule (the zero-momentum limit of our framework) provided adequate fit for 97% of participants (31/32 by BIC). The fitted momentum parameter, while statistically non-zero ($\beta = 0.003 \pm 0.006$, $p = 0.01$), was practically negligible. This is not a failure of the theory but a successful prediction: the helicopter task was designed to study learning rate adaptation in volatile environments, not belief inertia. Its parameters place observers firmly in the overdamped regime where our framework reduces to standard models.

This explains why gradient descent suffices here while oscillatory dynamics appear elsewhere. Kaplowitz and Fink (Kaplowitz and Fink, 1992) observed attitude oscillation in persuasion studies where participants had strong prior commitments (high Λ_{prior}) confronted with credible counter messages (high $\Lambda_{\text{observation}}$) which represent precisely the conditions our framework predicts should produce underdamped dynamics. Our present theory provides the missing bridge - both phenomena emerge from the same equations in different parameter regimes.

4.3 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.

4.4 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 momentum in a particular direction 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.

The rigidity of influence. Perhaps most striking is the prediction that influence itself accumulates as inertial mass. The outgoing social term $\sum_j \beta_{ji}\Lambda_{qi}$ 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.

4.5 Proposed Experimental Tests

The predictions derived above distinguish the inertial framework from standard first-order Bayesian models. We outline experimental paradigms that could test these predictions. Detailed implementation is left for future work.

4.5.1 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).

Discrimination: 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.

4.5.2 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.

Discrimination: 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.

4.5.3 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.

Discrimination: 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.

4.5.4 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 "observers" hold equally strong priors but receive no social attention. After exposure to identical counter-evidence, measure belief change magnitude and relaxation dynamics across conditions.

Discrimination: 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.

4.5.5 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.

Discrimination: 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.

4.6 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.

4.7 Connection to Transformer Attention

Elsewhere, we have shown that the attention weights β_{ij} (Eq. X) are not merely analogous to transformer self-attention but rather they *are* attention, derived from first principles Dennis (2025b,a). Standard transformers compute $\beta_{ij} = \text{softmax}(q_i^\top W_Q W_K^\top k_j/\sqrt{d})$ using learned projection matrices W_Q, W_K that parameterize an implicit similarity metric. Our framework reveals that this metric is actually the KL divergence on the statistical manifold.

$$\beta_{ij} = \frac{\exp(-D_{\text{KL}}(q_i \| \Omega_{ij}[q_j])/\tau)}{\sum_k \exp(-D_{\text{KL}}(q_i \| \Omega_{ik}[q_k])/\tau)} \tag{45}$$

The gauge transport operators Ω_{ij} generalize to the role that positional encoding and learned projections play in standard architectures.

4.8 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. Elsewhere (Dennis, 2025a), 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.

4.9 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 straightforward however, 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. The codebase we have built, found in the repository (below), supports the full dynamics as well as hierarchical emergence and more. The full details are too rich to confine to a single article.

Empirical validation. Our primary empirical test involved a task designed to elicit overdamped dynamics. 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 publicly available datasets utilizing the present framework's predictions would strengthen the empirical case.

5 Conclusion

We have shown that beliefs naturally possess inertia in relation to prior precision. The straightforward identification that epistemic mass equals statistical precision transforms our understanding of belief dynamics provides new tools that extend beyond dissipative gradient flow and into rich Hamiltonian dynamics.

Our theory predicts oscillations, over-shooting, resistance, decay, and resonances in belief dynamics. More fundamentally, it re-frames cognitive biases not as irrationality but instead as unavoidable consequences of belief inertia. Just as physical mass resists acceleration, cognitive precision resists belief change.

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

The mathematics has been hiding in plain sight for decades due to lack of transitive communication between far-flung fields of differential geometry, physics, and informational geometry. The Fisher information metric has been whispering this entire time that it is actually 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/Hamiltonian-VFE

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

- $\mathcal{C}$ is the base manifold (agent positions, social network topology)
- $G = \mathrm{SO}(d)$ is the gauge group (rotations in belief space)
- 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 $ksframetoagenti$ s frame.

A.1.2 Gauge Transformations

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

$$g : \{1, \ldots, N\} \to \mathrm{SO}(d), \quad i \mapsto g_i \tag{46}$$

Under this transformation, belief parameters transform as:

$$\mu_i \mapsto \mu_i' = g_i \mu_i \tag{47}$$

$$\Sigma_i \mapsto \Sigma_i' = g_i \Sigma_i g_i^T \tag{48}$$

$$\Lambda_{qi} \mapsto \Lambda_{qi}' = g_i \Lambda_{qi} g_i^T \tag{49}$$

The transport operators transform as:

$$\Omega_{ik} \mapsto \Omega_{ik}' = g_i \Omega_{ik} g_k^{-1} \tag{50}$$

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

$$\tilde{\mu}_k' = g_i \tilde{\mu}_k, \quad \tilde{\Lambda}_{qk}' = g_i \tilde{\Lambda}_{qk} g_i^T \tag{51}$$

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:

$$\begin{aligned}
[\mathbf{M}^{\mu}]'_{ii} &= \bar{\Lambda}'_{pi} + \sum_k \beta_{ik}\tilde{\Lambda}'_{qk} + \sum_j \beta_{ji}\Lambda'_{qi} \\
&= g_i\bar{\Lambda}_{pi}g_i^T + \sum_k \beta_{ik} g_i\tilde{\Lambda}_{qk}g_i^T + \sum_j \beta_{ji} g_i\Lambda_{qi}g_i^T \\
&= g_i\left[\bar{\Lambda}_{pi} + \sum_k \beta_{ik}\tilde{\Lambda}_{qk} + \sum_j \beta_{ji}\Lambda_{qi}\right] g_i^T \\
&= g_i\,[\mathbf{M}^{\mu}]_{ii}\,g_i^T
\end{aligned} \tag{52}$$

Off-diagonal blocks:

$$\begin{aligned}
[\mathbf{M}^{\mu}]'_{ik} &= -\beta_{ik}\Omega'_{ik}\Lambda'_{qk} - \beta_{ki}\Lambda'_{qi}(\Omega'_{ki})^T \\
&= -\beta_{ik}(g_i\Omega_{ik}g_k^{-1})(g_k\Lambda_{qk}g_k^T) - \beta_{ki}(g_i\Lambda_{qi}g_i^T)(g_k\Omega_{ki}g_i^{-1})^T \\
&= -\beta_{ik}g_i\Omega_{ik}\Lambda_{qk}g_k^T - \beta_{ki}g_i\Lambda_{qi}\Omega_{ki}^T g_k^T \\
&= g_i\,[\mathbf{M}^{\mu}]_{ik}\,g_k^T
\end{aligned} \tag{53}$$

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

$$\mathbf{G} = \mathrm{diag}(g_1, g_2, \ldots, g_N) \in \mathrm{SO}(d)^N \tag{54}$$

Then the full mean-sector mass matrix transforms as:

$$\boxed{(\mathbf{M}^{\mu})' = \mathbf{G}\,\mathbf{M}^{\mu}\,\mathbf{G}^T} \tag{55}$$

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:

$$\begin{aligned}
[\mathbf{M}^{\Sigma}]'_{ii} &= \frac{1}{2}(\Lambda'_{qi}\otimes\Lambda'_{qi})\cdot\left(1 + \sum_k \beta_{ik} + \sum_j \beta_{ji}\right) \\
&= \frac{1}{2}(g_i\Lambda_{qi}g_i^T \otimes g_i\Lambda_{qi}g_i^T)\cdot\left(1 + \sum_k \beta_{ik} + \sum_j \beta_{ji}\right) \\
&= \frac{1}{2}(g_i\otimes g_i)(\Lambda_{qi}\otimes\Lambda_{qi})(g_i^T\otimes g_i^T)\cdot\left(1 + \sum_k \beta_{ik} + \sum_j \beta_{ji}\right)
\end{aligned} \tag{56}$$

The transformation law is:

$$\boxed{(\mathbf{M}^{\Sigma})' = (\mathbf{G}\otimes\mathbf{G})\,\mathbf{M}^{\Sigma}\,(\mathbf{G}^T\otimes\mathbf{G}^T)} \tag{57}$$

A.2.3 Cross Blocks

The mean-covariance cross blocks transform as:

$$(\mathbf{C}^{\mu\Sigma})' = \mathbf{G}\,\mathbf{C}^{\mu\Sigma}\,(\mathbf{G}^T \otimes \mathbf{G}^T) \tag{58}$$

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:

$$\boxed{(\pi_i^\mu)' = g_i\,\pi_i^\mu} \tag{59}$$

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

$$\langle (\pi^\mu)', \dot{\mu}' \rangle = (g_i \pi_i^\mu)^T (g_i \dot{\mu}_i) = (\pi_i^\mu)^T g_i^T g_i \dot{\mu}_i = (\pi_i^\mu)^T \dot{\mu}_i = \langle \pi^\mu, \dot{\mu} \rangle \tag{60}$$

A.3.2 Covariance Momentum

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

$$\boxed{(\Pi_i^\Sigma)' = g_i\,\Pi_i^\Sigma\,g_i^T} \tag{61}$$

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

$$\mathrm{tr}[(\Pi^\Sigma)'\dot{\Sigma}'] = \mathrm{tr}[(g_i \Pi_i^\Sigma g_i^T)(g_i \dot{\Sigma}_i g_i^T)] = \mathrm{tr}[\Pi_i^\Sigma \dot{\Sigma}_i] \tag{62}$$

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:

$$\begin{aligned}
\dot{\mu}' &= ((\mathbf{M}^\mu)')^{-1}(\pi^\mu)' \\
&= (\mathbf{G}\mathbf{M}^\mu\mathbf{G}^T)^{-1}\mathbf{G}\pi^\mu \\
&= \mathbf{G}^{-T}(\mathbf{M}^\mu)^{-1}\mathbf{G}^{-1}\mathbf{G}\pi^\mu \\
&= \mathbf{G}(\mathbf{M}^\mu)^{-1}\pi^\mu \quad (\text{since } \mathbf{G}^{-T} = \mathbf{G} \text{ for } \mathrm{SO}(d)) \\
&= \mathbf{G}\dot{\mu}
\end{aligned} \tag{63}$$

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:

$$\left(\frac{\partial F}{\partial \mu_i}\right)' = g_i \frac{\partial F}{\partial \mu_i} \tag{64}$$

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:

$$\boxed{\dot{\pi}' = \mathbf{G}\dot{\pi}} \tag{65}$$

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:

$$A_e = \Omega_{ik} \in \mathrm{SO}(d) \tag{66}$$

Under gauge transformation:

$$A_e \mapsto A'_e = g_i A_e g_k^{-1} \tag{67}$$

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:

$$F_\gamma = \Omega_{ij}\Omega_{jk}\Omega_{ki} \in \mathrm{SO}(d) \tag{68}$$

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"—belief frames cannot be consistently aligned around cycles.

A.5.3 Variation of Connection

Consider an infinitesimal variation of the connection:

$$\delta\Omega_{ik} = \omega_{ik}\,\Omega_{ik}, \quad \omega_{ik} \in \mathfrak{so}(d) \tag{69}$$

The variation of transported precision is:

$$\delta\tilde{\Lambda}_k = \omega_{ik}\tilde{\Lambda}_k + \tilde{\Lambda}_k\omega_{ik}^T = [\omega_{ik}, \tilde{\Lambda}_k]_+ \tag{70}$$

where $[\cdot,\cdot]_+$ is the anticommutator (since ω_{ik} is antisymmetric).

A.6 Variation of the Mass Matrix Under Connection Changes

A.6.1 Diagonal Block Variation

$$\delta[\mathbf{M}^{\mu}]_{ii} = \sum_k \beta_{ik}\, \delta\tilde{\Lambda}_{qk} = \sum_k \beta_{ik}\, [\omega_{ik}, \tilde{\Lambda}_{qk}]_+ \tag{71}$$

A.6.2 Off-Diagonal Block Variation

$$\begin{aligned}\delta[\mathbf{M}^{\mu}]_{ik} &= -\beta_{ik}\, \delta(\Omega_{ik}\Lambda_{qk}) - \beta_{ki}\, \delta(\Lambda_{qi}\Omega_{ki}^T) \\ &= -\beta_{ik}\omega_{ik}\Omega_{ik}\Lambda_{qk} - \beta_{ki}\Lambda_{qi}(\omega_{ki}\Omega_{ki})^T \\ &= -\beta_{ik}\omega_{ik}\Omega_{ik}\Lambda_{qk} + \beta_{ki}\Lambda_{qi}\Omega_{ki}^T\omega_{ki}^T \end{aligned} \tag{72}$$

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

$$\boxed{\delta[\mathbf{M}^{\mu}]_{ik} = -\beta_{ik}\omega_{ik}\Omega_{ik}\Lambda_{qk} - \beta_{ki}\Lambda_{qi}\Omega_{ki}^T\omega_{ki}} \tag{73}$$

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:

$$\phi_g(\mu, \Sigma) = (g\mu, g\Sigma g^T) \tag{74}$$

The pullback metric is:

$$(\phi_g^*\mathcal{G})_{(\mu,\Sigma)}(v, w) = \mathcal{G}_{\phi_g(\mu,\Sigma)}(d\phi_g \cdot v, d\phi_g \cdot w) \tag{75}$$

For the Fisher-Rao metric, gauge invariance implies:

$$\boxed{\phi_g^*\mathcal{G} = \mathcal{G}} \tag{76}$$

The metric is **gauge-invariant**—this is the geometric content of our transformation laws.

A.7.2 Horizontal and Vertical Decomposition

The tangent space at each point decomposes as:

$$T_{(\mu,\Sigma)}\mathcal{Q} = H_{(\mu,\Sigma)} \oplus V_{(\mu,\Sigma)} \tag{77}$$

- **Vertical space** V: Directions along gauge orbits (pure gauge changes)
- **Horizontal space** H: Directions orthogonal to gauge orbits (physical changes)

The connection determines the horizontal subspace. A vector $v = (\delta\mu, \delta\Sigma)$ is horizontal if:

$$\mathcal{G}(v, \xi_X) = 0 \quad \forall X \in \mathfrak{so}(d) \tag{78}$$

where ξ_X is the vector field generated by X.

A.7.3 Physical (Gauge-Invariant) Quantities

Only horizontal components of velocities and momenta correspond to physical observables:

1. **Consensus divergence**: $\|\mu_i - \tilde{\mu}_k\|^2_{\tilde{\Lambda}_{qk}} = (\mu_i - \tilde{\mu}_k)^T \tilde{\Lambda}_{qk} (\mu_i - \tilde{\mu}_k)$
2. **Free energy**: $F[\{q_i\}]$ is gauge-invariant by construction
3. **Hamiltonian**: $H = \frac{1}{2}\langle \pi, \mathbf{M}^{-1}\pi \rangle + F$ is gauge-invariant
4. **Inter-agent KL divergence**: $\mathrm{KL}(q_i \| \Omega_{ik}[q_k])$ is gauge-invariant

A.8 Gauge-Fixed Dynamics

For numerical implementation, it is often convenient to work in a fixed gauge.

A.8.1 Identity Gauge

Set $g_i = I$ for all agents. Then:

- Transport operators Ω_{ik} are directly the frame transformations
- All quantities take their "bare" form
- Gauge redundancy is eliminated

A.8.2 Consensus-Aligned Gauge

Choose gauges so that at equilibrium:

$$\Omega^*_{ik} = I \quad \text{(parallel frames at consensus)} \tag{79}$$

This simplifies analysis near equilibrium since transported quantities equal untransported ones.

A.8.3 Principal Axis Gauge

For each agent, choose g_i to diagonalize Σ_i:

$$\Sigma'_i = g_i \Sigma_i g_i^T = \mathrm{diag}(\lambda_1^{(i)}, \ldots, \lambda_d^{(i)}) \tag{80}$$

This separates dynamics along principal axes of uncertainty.

A.9 Summary: Gauge-Covariant Hamiltonian Mechanics

Gauge Transformation Laws

Positions:

$$\mu_i' = g_i \mu_i \qquad \Sigma_i' = g_i \Sigma_i g_i^T \tag{81}$$

Momenta:

$$(\pi_i^\mu)' = g_i \pi_i^\mu \qquad (\Pi_i^\Sigma)' = g_i \Pi_i^\Sigma g_i^T \tag{82}$$

Mass Matrix:

$$\mathbf{M}' = \mathbf{G}\mathbf{M}\mathbf{G}^T \tag{83}$$

Transport Operators:

$$\Omega_{ik}' = g_i \Omega_{ik} g_k^{-1} \tag{84}$$

Hamilton's Equations: Fully covariant under these transformations.
Physical 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:

$$\Lambda_{qi} = \Sigma_i^{-1} \quad \text{(belief precision)} \tag{85}$$
$$\bar{\Lambda}_{pi} = \bar{\Sigma}_i^{-1} \quad \text{(prior precision)} \tag{86}$$
$$\Lambda_{o_i} = \Sigma_{o_i}^{-1} \quad \text{(observation precision)} \tag{87}$$
$$\tilde{\mu}_k = \Omega_{ik} \mu_k \quad \text{(transported mean)} \tag{88}$$
$$\tilde{\Lambda}_{qk} = \Omega_{ik} \Lambda_{qk} \Omega_{ik}^T \quad \text{(transported precision)} \tag{89}$$

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

B.2 The Extended Free Energy Functional

The complete variational free energy with explicit sensory evidence is:

$$\boxed{\mathcal{F}[\{q_i\}] = \sum_i D_{\mathrm{KL}}(q_i \| p_i) + \sum_{i,k} \beta_{ik} D_{\mathrm{KL}}(q_i \| \Omega_{ik}[q_k]) - \sum_i \mathbb{E}_{q_i}[\log p(o_i \mid \theta)]} \tag{90}$$

The three terms represent:

1. **Prior anchoring**: Deviation from internal world-model
2. **Social consensus**: Alignment with other agents via gauge-covariant transport
3. **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 $p = \mathcal{N}(\mu_p, \Sigma_p)$:

$$D_{\mathrm{KL}}(q\|p) = \frac{1}{2}\left[\mathrm{tr}(\Sigma_p^{-1}\Sigma_q) + (\mu_p - \mu_q)^T\Sigma_p^{-1}(\mu_p - \mu_q) - d + \ln\frac{|\Sigma_p|}{|\Sigma_q|}\right] \tag{91}$$

B.3.2 Expected Log-Likelihood

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

$$\mathbb{E}_{q_i}[\log p(o_i \mid \theta)] = -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_{o_i}| - \frac{1}{2}\mathbb{E}_{q_i}\left[(o_i - \theta)^T\Lambda_{o_i}(o_i - \theta)\right] \tag{92}$$

The quadratic expectation evaluates to:

$$\mathbb{E}_{q_i}\left[(o_i - \theta)^T\Lambda_{o_i}(o_i - \theta)\right] = (o_i - \mu_i)^T\Lambda_{o_i}(o_i - \mu_i) + \mathrm{tr}(\Lambda_{o_i}\Sigma_i) \tag{93}$$

Therefore:

$$\boxed{-\mathbb{E}_{q_i}[\log p(o_i \mid \theta)] = \frac{1}{2}(o_i - \mu_i)^T\Lambda_{o_i}(o_i - \mu_i) + \frac{1}{2}\mathrm{tr}(\Lambda_{o_i}\Sigma_i) + \mathrm{const}} \tag{94}$$

B.4 First Variations (Gradient)

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

$$\frac{\partial D_{\mathrm{KL}}(q_i\|p_i)}{\partial \mu_i} = \bar{\Lambda}_{pi}(\mu_i - \bar{\mu}_i) \tag{95}$$

$$\frac{\partial D_{\mathrm{KL}}(q_i\|p_i)}{\partial \Sigma_i} = \frac{1}{2}(\bar{\Lambda}_{pi} - \Lambda_{qi}) \tag{96}$$

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

With respect to receiver i:

$$\frac{\partial D_{\mathrm{KL}}(q_i \| \tilde{q}_k)}{\partial \mu_i} = \tilde{\Lambda}_{qk}(\mu_i - \tilde{\mu}_k) \tag{97}$$

$$\frac{\partial D_{\mathrm{KL}}(q_i \| \tilde{q}_k)}{\partial \Sigma_i} = \frac{1}{2}(\tilde{\Lambda}_{qk} - \Lambda_{qi}) \tag{98}$$

With respect to sender k:

$$\frac{\partial D_{\mathrm{KL}}(q_i \| \tilde{q}_k)}{\partial \mu_k} = \Lambda_{qk}\Omega_{ik}^T(\tilde{\mu}_k - \mu_i) \tag{99}$$

$$\frac{\partial D_{\mathrm{KL}}(q_i \| \tilde{q}_k)}{\partial \Sigma_k} = \frac{1}{2}\Omega_{ik}^T \left[\tilde{\Lambda}_{qk} - \tilde{\Lambda}_{qk}\Sigma_i\tilde{\Lambda}_{qk}\right] \Omega_{ik} \tag{100}$$

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

$$\frac{\partial}{\partial \mu_i}\left[-\mathbb{E}_{q_i}[\log p(o_i \mid \theta)]\right] = \Lambda_{o_i}(\mu_i - o_i) \tag{101}$$

$$\frac{\partial}{\partial \Sigma_i}\left[-\mathbb{E}_{q_i}[\log p(o_i \mid \theta)]\right] = \frac{1}{2}\Lambda_{o_i} \tag{102}$$

B.4.4 Total Gradient

$$\boxed{\frac{\partial \mathcal{F}}{\partial \mu_i} = \bar{\Lambda}_{pi}(\mu_i - \bar{\mu}_i) + \sum_k \beta_{ik}\tilde{\Lambda}_{qk}(\mu_i - \tilde{\mu}_k) + \sum_j \beta_{ji}\Lambda_{qi}\Omega_{ji}^T(\tilde{\mu}_i^{(j)} - \mu_j) + \Lambda_{o_i}(\mu_i - o_i)} \tag{103}$$

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

$$\frac{\partial \mathcal{F}}{\partial \Sigma_i} = \frac{1}{2}(\bar{\Lambda}_{pi} - \Lambda_{qi}) + \sum_k \frac{\beta_{ik}}{2}(\tilde{\Lambda}_{qk} - \Lambda_{qi}) + \sum_j \frac{\beta_{ji}}{2}\Omega_{ji}^T \left[\tilde{\Lambda}_{qi}^{(j)} - \tilde{\Lambda}_{qi}^{(j)}\Sigma_j\tilde{\Lambda}_{qi}^{(j)}\right] \Omega_{ji} + \frac{1}{2}\Lambda_{o_i} \tag{104}$$

B.5 Second Variations (Hessian = Mass Matrix)

The Fisher-Rao metric $\mathcal{G} = \partial^2\mathcal{F}/\partial\xi\partial\xi$ serves as the mass matrix.

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

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

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i \| p_i)}{\partial \mu_i \partial \mu_i^T} = \bar{\Lambda}_{pi} \tag{105}$$

From consensus (as receiver):

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i \| \tilde{q}_k)}{\partial \mu_i \partial \mu_i^T} = \tilde{\Lambda}_{qk} = \Omega_{ik} \Lambda_{qk} \Omega_{ik}^T \tag{106}$$

From consensus (as sender to agent j):

$$\frac{\partial^2 D_{\mathrm{KL}}(q_j \| \tilde{q}_i)}{\partial \mu_i \partial \mu_i^T} = \Omega_{ji}^T \tilde{\Lambda}_{qi}^{(j)} \Omega_{ji} = \Lambda_{qi} \tag{107}$$

From sensory evidence:

$$\frac{\partial^2}{\partial \mu_i \partial \mu_i^T} \left[-\mathbb{E}_{q_i} [\log p(o_i \mid \theta)] \right] = \Lambda_{o_i} \tag{108}$$

Total diagonal mass:

$$\boxed{[\mathbf{M}^\mu]_{ii} = \underbrace{\bar{\Lambda}_{pi}}_{\text{prior}} + \underbrace{\sum_k \beta_{ik} \tilde{\Lambda}_{qk}}_{\text{incoming social}} + \underbrace{\sum_j \beta_{ji} \Lambda_{qi}}_{\text{outgoing recoil}} + \underbrace{\Lambda_{o_i}}_{\text{sensory}}} \tag{109}$$

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

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i \| \tilde{q}_k)}{\partial \mu_i \partial \mu_k^T} = -\tilde{\Lambda}_{qk} \Omega_{ik} = -\Omega_{ik} \Lambda_{qk} \tag{110}$$

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

$$\frac{\partial^2 D_{\mathrm{KL}}(q_k \| \tilde{q}_i)}{\partial \mu_i \partial \mu_k^T} = -\Lambda_{qi} \Omega_{ki}^T \tag{111}$$

The sensory term does not couple different agents. Therefore:

$$\boxed{[\mathbf{M}^\mu]_{ik} = -\beta_{ik} \Omega_{ik} \Lambda_{qk} - \beta_{ki} \Lambda_{qi} \Omega_{ki}^T \quad (i \neq k)} \tag{112}$$

The mass matrix is symmetric only when $\beta_{ik} = \beta_{ki}$ and $\Omega_{ik} = \Omega_{ki}^T$.

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

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

$$\frac{\partial^2 f}{\partial \Sigma \partial \Sigma}[V, W] = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left(\left. \frac{\partial f}{\partial \Sigma} \right|_{\Sigma + \epsilon W} - \left. \frac{\partial f}{\partial \Sigma} \right|_{\Sigma} \right) [V] \tag{113}$$

Key identity:

$$\frac{\partial}{\partial \Sigma}(\Sigma^{-1}) = -\Sigma^{-1} \otimes \Sigma^{-1} \tag{114}$$

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

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i \| p_i)}{\partial \Sigma_i \partial \Sigma_i}[V, W] = \frac{1}{2}\mathrm{tr}\left[\Lambda_{qi} V \Lambda_{qi} W\right] \tag{115}$$

In tensor notation:

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i \| p_i)}{\partial \Sigma_i \partial \Sigma_i} = \frac{1}{2}(\Lambda_{qi} \otimes \Lambda_{qi}) \tag{116}$$

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**:

$$\frac{\partial^2}{\partial \Sigma_i \partial \Sigma_i}\left[\mathrm{tr}(\Lambda_{o_i}\Sigma_i)\right] = 0 \tag{117}$$

Therefore:

$$\boxed{[\mathbf{M}^{\Sigma}]_{ii} = \frac{1}{2}(\Lambda_{qi} \otimes \Lambda_{qi}) \cdot \left(1 + \sum_k \beta_{ik} + \sum_j \beta_{ji}\right)} \tag{118}$$

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

B.5.3 Mean-Covariance Cross Blocks

Prior term:

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i \| p_i)}{\partial \mu_i \partial \Sigma_i} = 0 \tag{119}$$

Sensory term: The sensory free energy decomposes as:

- Quadratic in μ_i: $(o_i - \mu_i)^T \Lambda_{o_i}(o_i - \mu_i)$
- Linear in Σ_i: $\mathrm{tr}(\Lambda_{o_i}\Sigma_i)$

These are independent, so:

$$[\mathbf{C}^{\mu\Sigma}]_{ii}^{\text{sensory}} = 0 \tag{120}$$

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

$$\frac{\partial^2 D_{\mathrm{KL}}(q_i \| \tilde{q}_k)}{\partial \mu_i \partial \Sigma_k}[V] = -\Omega_{ik}\Lambda_{qk} V \Lambda_{qk}\Omega_{ik}^T(\mu_i - \tilde{\mu}_k) \tag{121}$$

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

$$[\mathbf{C}^{\mu\Sigma}]_{ik} = 0 \quad \text{when } \mu_i = \tilde{\mu}_k \tag{122}$$

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

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}^{\mu} & \mathbf{C}^{\mu\Sigma} \\ (\mathbf{C}^{\mu\Sigma})^T & \mathbf{M}^{\Sigma} \end{pmatrix} \tag{123}$$

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

B.6.2 Explicit Formulae

Mean sector diagonal:

$$[\mathbf{M}^{\mu}]_{ii} = \underbrace{\bar{\Lambda}_{pi}}_{\text{prior anchoring}} + \underbrace{\sum_k \beta_{ik}\Omega_{ik}\Lambda_{qk}\Omega_{ik}^T}_{\text{incoming consensus}} + \underbrace{\sum_j \beta_{ji}\Lambda_{qi}}_{\text{outgoing recoil}} + \underbrace{\Lambda_{o_i}}_{\text{sensory grounding}} \tag{124}$$

Mean sector off-diagonal:

$$[\mathbf{M}^{\mu}]_{ik} = -\beta_{ik}\Omega_{ik}\Lambda_{qk} - \beta_{ki}\Lambda_{qi}\Omega_{ki}^T \quad (i \neq k) \tag{125}$$

Covariance sector diagonal:

$$[\mathbf{M}^{\Sigma}]_{ii} = \frac{1}{2}(\Lambda_{qi} \otimes \Lambda_{qi}) \cdot \left(1 + \sum_k \beta_{ik} + \sum_j \beta_{ji}\right) \tag{126}$$

Cross mean-covariance (at consensus):

$$[\mathbf{C}^{\mu\Sigma}]_{ik} = 0 \quad \text{when } \mu_i = \tilde{\mu}_k \tag{127}$$

B.7 Physical Interpretation

B.7.1 Mass as Precision

The mean-sector effective mass for agent i is:

$$\boxed{M_i = \bar{\Lambda}_{pi} + \sum_k \beta_{ik}\tilde{\Lambda}_{qk} + \sum_j \beta_{ji}\Lambda_{qi} + \Lambda_{o_i}} \tag{128}$$

- $\bar{\Lambda}_{pi}$: **Bare mass** — inertia against deviation from prior
- $\sum_k \beta_{ik}\tilde{\Lambda}_{qk}$: **Incoming relational mass** — inertia from being "pulled" by neighbors
- $\sum_j \beta_{ji}\Lambda_{qi}$: **Outgoing relational mass** — inertia from "pulling" neighbors (recoil)
- Λ_{o_i}: **Sensory mass** — inertia from grounding in observations

B.7.2 Asymmetry of Sensory Contribution

The sensory precision Λ_{o_i} contributes to:

1. The mean-sector mass (Eq. 124)
2. The mean-sector force (Eq. 103)

But **not** to:

1. The covariance-sector mass (Eq. 118)

This asymmetry arises because the sensory term is quadratic in μ but only linear in Σ.

B.7.3 Kinetic Energy

$$T = \frac{1}{2}\dot{\mu}^T \mathbf{M}^\mu \dot{\mu} + \frac{1}{2}\text{tr}\left[\mathbf{M}^\Sigma[\dot{\Sigma}, \dot{\Sigma}]\right] \tag{129}$$

The first term gives standard "particle" kinetic energy with precision-mass. The second gives "shape" kinetic energy on the SPD manifold.

B.8 The Hamiltonian

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

$$\boxed{H = \frac{1}{2}\langle \pi, \mathbf{M}^{-1}\pi \rangle + \mathcal{F}[\xi]} \tag{130}$$

B.9 Hamilton's Equations

B.9.1 Equations of Motion

$$\dot{\mu}_i = \sum_k [\mathbf{M}^{-1}]_{ik}^{\mu\mu} \pi_k^\mu + \sum_k [\mathbf{M}^{-1}]_{ik}^{\mu\Sigma} \Pi_k^\Sigma \tag{131}$$

$$\dot{\Sigma}_i = \sum_k [\mathbf{M}^{-1}]_{ik}^{\Sigma\mu} \pi_k^\mu + \sum_k [\mathbf{M}^{-1}]_{ik}^{\Sigma\Sigma} \Pi_k^\Sigma \tag{132}$$

$$\dot{\pi}_i^\mu = -\frac{\partial \mathcal{F}}{\partial \mu_i} - \frac{1}{2}\pi^T \frac{\partial \mathbf{M}^{-1}}{\partial \mu_i}\pi \tag{133}$$

$$\dot{\Pi}_i^\Sigma = -\frac{\partial \mathcal{F}}{\partial \Sigma_i} - \frac{1}{2}\pi^T \frac{\partial \mathbf{M}^{-1}}{\partial \Sigma_i}\pi \tag{134}$$

B.9.2 Force Decomposition

The potential forces decompose into four physically distinct contributions:

$$\boxed{-\frac{\partial \mathcal{F}}{\partial \mu_i} = \underbrace{-\bar{\Lambda}_{pi}(\mu_i - \bar{\mu}_i)}_{\text{prior restoring}} \underbrace{- \sum_k \beta_{ik} \tilde{\Lambda}_{qk}(\mu_i - \tilde{\mu}_k)}_{\text{consensus}} \underbrace{- \sum_j \beta_{ji} \Lambda_{qi} \Omega_{ji}^T (\tilde{\mu}_i^{(j)} - \mu_j)}_{\text{reciprocal}} \underbrace{-\Lambda_{o_i}(\mu_i - o_i)}_{\text{sensory evidence}}} \tag{135}$$

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

B.9.3 Compact Form

$$\boxed{\begin{aligned} \dot{\xi} &= \mathbf{M}^{-1}\pi \\ \dot{\pi} &= -\nabla \mathcal{F} - \frac{1}{2} \nabla_\xi \langle \pi, \mathbf{M}^{-1}\pi \rangle \end{aligned}} \tag{136}$$

with $dH/dt = 0$ along trajectories.

B.10 Damped Dynamics

Including dissipation yields:

$$M_i \ddot{\mu}_i + \gamma_i \dot{\mu}_i + \nabla_{\mu_i} \mathcal{F} = 0 \tag{137}$$

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

$$M_i \ddot{\delta\mu} + \gamma_i \dot{\delta\mu} + K_i \delta\mu = 0 \tag{138}$$

The discriminant $\Delta = \gamma_i^2 - 4K_i M_i$ determines three regimes:

- **Overdamped** ($\Delta > 0$): Monotonic decay, standard Bayesian updating
- **Critically damped** ($\Delta = 0$): Fastest equilibration
- **Underdamped** ($\Delta < 0$): Oscillatory approach with overshooting

B.11 Momentum Current with Sensory Coupling

Between agents, the momentum current is:

$$J_{k \to i} = \beta_{ik} \tilde{\Lambda}_{qk} (\tilde{\mu}_k - \mu_i) \tag{139}$$

The continuity equation becomes:

$$\dot{\pi}_i + \gamma_i \dot{\mu}_i + \bar{\Lambda}_{pi}(\mu_i - \bar{\mu}_i) + \Lambda_{o_i}(\mu_i - o_i) = \sum_k J_{k \to i} \tag{140}$$

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:

$$\mathcal{F} = \sum_i D_{\mathrm{KL}}(q_i \| p_i) + \sum_{i,k} \beta_{ik} D_{\mathrm{KL}}(q_i \| \Omega_{ik}[q_k]) - \sum_i \mathbb{E}_{q_i}[\log p(o_i \mid \theta)] \tag{141}$$

Mass Matrix:

$$\mathbf{M} = \frac{\partial^2 \mathcal{F}}{\partial \xi \partial \xi} = \text{Fisher information} = \text{Precision} \tag{142}$$

Effective Mass:

$$M_i = \bar{\Lambda}_{pi} + \sum_k \beta_{ik} \tilde{\Lambda}_{qk} + \sum_j \beta_{ji} \Lambda_{qi} + \Lambda_{o_i} \tag{143}$$

Dynamics:

$$\dot{\xi} = \mathbf{M}^{-1} \pi, \qquad \dot{\pi} = -\nabla \mathcal{F} - \frac{1}{2} \nabla_\xi \langle \pi, \mathbf{M}^{-1} \pi \rangle \tag{144}$$

Physical Meaning:

- Position μ_i = what agent i believes
- Momentum π_i = rate of belief change $\times$ precision
- Mass = precision (tight beliefs are heavy)
- Force = pull toward prior + consensus + observations

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.

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.

Dennis, R. (2025a). Implementing attention and transformers without neural networks: Validation of gauge-theoretic transformers. *Journal of Machine Learning Research*. Submitted.

Dennis, R. C. (2025b). Attention, transformers, and backpropagation are degenerate limits of the variational free energy principle. Submitted to Journal of Machine Learning Research.

Dennis, R. C. (2025c). A theoretical and computational implementation of a participatory "it from bit" universe. Submitted to Foundations of Physics.

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. (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.

Kobayashi, S. and Nomizu, K. (1963). *Foundations of differential geometry*, volume 1. Interscience Publishers.

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.

McGuire, J. T., Nassar, M. R., Gold, J. I., and Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron*, 84(4):870–881.

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.

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parber, K., Heasly, B., and Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, 15(7):1040–1046.

Nassar, M. R., Wilson, R. C., Heasly, B., and Gold, J. I. (2010). An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, 30(37):12366–12378.

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.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008.

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.