Circulatory Fidelity:

Stability Constraints on Hierarchical Variational Inference

A Computational Framework Linking Structured Approximations to Dopaminergic Precision Regulation

Aaron Lowry

December 2025

Working Draft

Abstract

This thesis introduces *Circulatory Fidelity* (CF), a measure quantifying the statistical dependency preserved between levels of a hierarchical generative model during approximate inference. Using the Hierarchical Gaussian Filter (HGF) as a model system, we analyze the dynamical stability of variational updates under different approximation schemes.

We demonstrate that mean-field variational inference—which assumes independence between hierarchical levels—exhibits period-doubling bifurcations leading to deterministic chaos when environmental volatility exceeds a critical threshold. Structured approximations that preserve cross-level dependencies remain stable across a broader parameter range. This stability difference is characterized via Lyapunov exponent analysis and bifurcation diagrams.

Extending the analysis to three-level hierarchies, we find that mean-field instability is *amplified* with depth (85-fold increase in variance), while structured approximations maintain stability. We further discover that lower hierarchical interfaces (sensory–volatility) are more critical for stability than upper interfaces (volatility–meta-volatility), suggesting that precision modulation may need to act preferentially at specific hierarchical levels.

We propose that biological inference systems may face analogous trade-offs between computational cost and dynamical stability. As a candidate neural implementation, we suggest that tonic dopamine concentration could modulate the precision of hierarchical message passing, though this mapping remains speculative. The framework generates testable predictions regarding the relationship between neuromodulatory state and belief dynamics.

Contents

Notation

Symbol	Definition	Domain
z	Log-volatility (Level 2 state)	$\mathbb{R}$
x	Hidden state (Level 1)	$\mathbb{R}$
y	Observation	$\mathbb{R}$
κ	Coupling strength	> 0
ω	Baseline log-volatility	$\mathbb{R}$
ϑ	Volatility of volatility	> 0
π_u	Observation precision	> 0
γ	Precision weight	> 0
μ_z, μ_x	Posterior means	$\mathbb{R}$
σ_z^2, σ_x^2	Posterior variances	> 0
$I(\cdot;\cdot)$	Mutual information	≥ 0
$H(\cdot)$	Entropy	≥ 0
$\lambda_{\max}$	Maximal Lyapunov exponent	$\mathbb{R}$

1 Introduction

1.1 The Problem

Biological agents infer hidden causes of sensory observations across multiple timescales. A canonical example: estimating both the current state of an environment and how quickly that environment is changing. These two quantities—state and volatility—interact: beliefs about volatility determine how much weight to place on new observations when updating state estimates.

The Hierarchical Gaussian Filter (HGF) formalizes this structure (Mathys et al., 2011, 2014). In the HGF, a higher level encodes log-volatility, which parameterizes the expected rate of change at the lower level. Exact Bayesian inference in such models is generally intractable, motivating variational approximations.

1.2 Variational Approximations

Variational inference replaces the intractable true posterior $p(x, z|y)$ with a tractable approximation $q(x, z)$, chosen to minimize the Kullback-Leibler divergence from the true posterior. The most common simplification is the *mean-field approximation*:

$$q(x, z) = q(x)q(z) \tag{1}$$

This factorization assumes independence between levels, dramatically reducing computational complexity. However, this independence assumption discards information about how states at different levels covary.

An alternative is the *structured approximation*:

$$q(x, z) = q(z)q(x|z) \tag{2}$$

which preserves the conditional dependency of lower-level states on higher-level states. This comes at increased computational cost but retains more information about the joint posterior structure.

1.3 This Thesis

We investigate the dynamical consequences of these approximation choices. Our central empirical finding (from simulation) is that mean-field variational updates can become dynamically unstable—exhibiting bifurcations and chaos—under conditions where structured approximations remain stable.

We formalize this stability difference through a quantity we term *Circulatory Fidelity* (CF), measuring the mutual information preserved between hierarchical levels. Specifi-

cally, CF is the normalized mutual information between hierarchical levels, using a normalization that corresponds to the *uncertainty coefficient* from classical information theory (Coombs et al., 1970; Theil, 1970). Our contribution is not the measure itself—which has well-established precedent—but its application to analyzing the stability properties of variational approximations. We then explore the implications for understanding biological inference systems, proposing (speculatively) that neuromodulatory mechanisms may have evolved partly to maintain stable inference dynamics.

Scope and limitations: This is primarily a computational and theoretical analysis. The neural implementation we propose is speculative and intended to generate testable hypotheses rather than to make strong claims about biological mechanism. Throughout, we distinguish between what we can demonstrate formally, what we observe in simulation, and what we conjecture.

2 Theoretical Framework

2.1 The Hierarchical Gaussian Filter

We work with a two-level HGF defined by the following generative model:

Level 2 (Volatility):

$$z_t \mid z_{t-1} \sim \mathcal{N}(z_{t-1}, \vartheta^{-1}) \tag{3}$$

Level 1 (Hidden state):

$$x_t \mid x_{t-1}, z_t \sim \mathcal{N}(x_{t-1}, \exp(\kappa z_t + \omega)) \tag{4}$$

Observations:

$$y_t \mid x_t \sim \mathcal{N}(x_t, \pi_u^{-1}) \tag{5}$$

The parameter ϑ controls how quickly volatility itself changes (sometimes called the "hazard rate" or "meta-volatility"). The coupling κ determines how strongly volatility modulates state transitions. The baseline ω sets the typical level of volatility when $z = 0$.

2.2 Variational Updates

Under the mean-field approximation $q(x, z) = q(x)q(z)$, the variational updates take the form of coupled fixed-point equations. At each timestep, given observation y_t, the posterior parameters are updated according to:

Level 1 update:

$$\mu_x^{(t)} = \mu_x^{(t-1)} + \frac{\pi_u}{\pi_u + \hat{\pi}_x}(y_t - \mu_x^{(t-1)}) \tag{6}$$

where $\hat{\pi}_x = \exp(-\kappa\mu_z^{(t-1)} - \omega)$ is the expected precision at level 1.

Level 2 update:

$$\mu_z^{(t)} = \mu_z^{(t-1)} + \frac{\kappa}{2}\frac{\hat{\pi}_x}{\vartheta + \frac{\kappa^2}{2}\hat{\pi}_x}\left((y_t - \mu_x^{(t-1)})^2\hat{\pi}_x - 1\right) \tag{7}$$

These update equations define a discrete dynamical system. Our analysis focuses on the stability properties of this system.

2.2.1 Structured Variational Update

Under the structured approximation $q(x, z) = q(z)q(x|z)$, the updates differ from mean-field by incorporating cross-level coupling. The key modification is to the Level 2 update, which gains an additional damping term:

$$\mu_z^{(t)} = \mu_z^{(t-1)} + K_z \cdot \nu - \underbrace{\gamma_{zx} \cdot \mathrm{Cov}_q(z, x) \cdot \delta_x}_{\text{coupling term}} \tag{8}$$

where K_z is the standard gain, ν is the volatility prediction error, and γ_{zx} is a coupling coefficient.

Status of the coupling coefficient: The specific form of γ_{zx} used in our simulations is a **modeling approximation**, not a first-principles derivation. We use:

$$\gamma_{zx} = \frac{\kappa\pi_x}{4} \tag{9}$$

Justification for this choice:

1. **Dimensional consistency:** The coefficient has units of precision, matching the update equation structure.
2. **Correct limiting behavior:** As $\kappa \to 0$ (no coupling between levels), $\gamma_{zx} \to 0$, recovering the mean-field update.
3. **Empirical calibration:** The factor of $1/4$ was chosen to produce stable dynamics across the parameter range of interest.

See Appendix A.5 for further discussion of this approximation.

2.3 Circulatory Fidelity

Definition 2.1 (Circulatory Fidelity)**.** For a joint approximate posterior $q(x, z)$ with marginal entropies $H_q(x)$ and $H_q(z)$, Circulatory Fidelity is:

$$\mathrm{CF} = \frac{I_q(z; x)}{\min(H_q(z), H_q(x))} \tag{10}$$

where $I_q(z;x) = H_q(z) + H_q(x) - H_q(z,x)$ is the mutual information between z and x under q.

Properties:

- $\text{CF} \in [0,1]$
- $\text{CF} = 0$ if and only if z and x are independent under q
- $\text{CF} = 1$ if and only if one variable is a deterministic function of the other

Proposition 2.1 (CF Under Mean-Field)**.** Under the mean-field approximation, $\text{CF} = 0$.

Proof. Under $q(x,z) = q(x)q(z)$, the joint entropy decomposes: $H(x,z) = H(x) + H(z)$. Therefore $I(x;z) = H(x) + H(z) - H(x,z) = 0$. Since the numerator is zero, $\text{CF} = 0$. □

Corollary 2.1. Any structured approximation with non-trivial conditional dependency has $\text{CF} > 0$.

Remark 2.1 (Choice of Normalization and Prior Work)**.** We normalize by $\min(H(z), H(x))$ rather than the joint entropy $H(z,x)$. This normalization is not novel: it corresponds to the *uncertainty coefficient* (also called the *coefficient of constraint*) from classical information theory (Press, 1967; Coombs et al., 1970). Our contribution is not the measure itself but its application to analyzing variational approximation quality.

This normalization has two advantages:

1. **Interpretability:** $\text{CF} = 1$ when one variable is a deterministic function of the other.
2. **Behavior in high-volatility regimes:** The joint entropy grows with marginal entropies; using it as denominator would cause CF to shrink precisely when the dependency structure matters most.

2.4 Information Geometry

The space of Gaussian distributions forms a Riemannian manifold with the Fisher Information Matrix (FIM) as its metric.

Proposition 2.2 (FIM Block-Diagonality)**.** Under mean-field $q(x,z) = q(x)q(z)$, the Fisher Information Matrix is block-diagonal:

$$\mathbf{G}_{\text{MF}} = \begin{pmatrix} G_{zz} & 0 \\ 0 & G_{xx} \end{pmatrix} \tag{11}$$

Under structured $q(x,z) = q(z)q(x|z)$, the FIM generically contains non-zero off-diagonal terms:

$$\mathbf{G}_{\text{struct}} = \begin{pmatrix} G_{zz} & G_{zx} \\ G_{xz} & G_{xx} \end{pmatrix} \tag{12}$$

Proof. Under mean-field, $\ln q(x, z) = \ln q(x) + \ln q(z)$. Derivatives with respect to z-parameters depend only on z; derivatives with respect to x-parameters depend only on x. Cross-terms in the FIM factor as products of expectations, each of which vanishes because the score function has zero mean for exponential families:

$$\mathbb{E}_q\left[\frac{\partial \ln q}{\partial \theta}\right] = \frac{\partial}{\partial \theta}\int q\, d\cdot = \frac{\partial}{\partial \theta}1 = 0 \tag{13}$$

Under structured approximation, the conditional $q(x|z)$ prevents this factorization. See Appendix A for complete proof. □

Interpretation: The off-diagonal FIM terms G_{zx} encode information about how to jointly adjust beliefs at both levels in response to evidence. Their *provable* absence under mean-field means this coordinated adjustment mechanism is structurally precluded by the approximation.

3 Dynamical Systems Analysis

3.1 The Update Map

The mean-field variational updates define a discrete map:

$$\mathbf{s}^{(t+1)} = F(\mathbf{s}^{(t)}, y_t) \tag{14}$$

where $\mathbf{s} = (\mu_z, \mu_x, \sigma_z^2, \sigma_x^2)$ is the state vector. For fixed input statistics, this becomes an iterated function system whose stability we can analyze.

3.2 Local Stability Analysis

Proposition 3.1 (Deterministic Skeleton Stability)**.** Consider the mean-field HGF z-dynamics under the deterministic skeleton approximation (replacing stochastic observations with their expected values). The linearized dynamics around equilibrium have Jacobian eigenvalue:

$$\lambda = \frac{\vartheta}{\vartheta + \alpha}, \quad \text{where } \alpha = \frac{\kappa^2 \pi_x^*}{2} \tag{15}$$

Since $\vartheta > 0$ and $\alpha > 0$, we have $0 < \lambda < 1$, implying **local stability** of the deterministic skeleton for all parameter values.

Proof. See Appendix A.2. □

Key insight: This stability result for the deterministic skeleton appears to contradict the observed instabilities. The resolution is that the instability arises from **stochastic fluctuations**, not from the deterministic dynamics.

Stochastic instability mechanism: In the full stochastic system, large prediction errors can drive large updates. The update:

$$\mu_z^{(t+1)} = \mu_z^{(t)} + K_z \cdot ((\delta^{(t)})^2 \pi_x - 1) \tag{16}$$

has variance that depends on the fourth moment of δ, which can exceed the mean-squared response, leading to variance growth.

3.3 Bifurcation Structure

Empirical Observation 3.1. In numerical simulations with parameters $\kappa = 1$, $\omega = -2$, $\pi_u = 10$, we observe:

- For $\vartheta < \vartheta_c$, the mean-field dynamics converge to a stable fixed point
- At $\vartheta \approx \vartheta_c$, a period-doubling bifurcation occurs
- For $\vartheta_c < \vartheta < \vartheta_{\text{chaos}}$, successive period-doublings are observed
- For $\vartheta > \vartheta_{\text{chaos}}$, the dynamics appear chaotic (positive Lyapunov exponent)

The critical values depend on parameters and noise realization:

$$\vartheta_c \in [0.04, 0.06], \quad \vartheta_{\text{chaos}} \in [0.10, 0.15] \tag{17}$$

Remark 3.1 (Resolution of the Stability Paradox)**.** The deterministic skeleton analysis (Proposition 3.1) predicts stability for all $\vartheta > 0$, yet simulations clearly show instabilities. This apparent contradiction is resolved by recognizing that **deterministic and stochastic stability are distinct phenomena**:

Analysis	Object of Study	Prediction
Deterministic skeleton	Dynamics of *expected* state	Stable for all $\vartheta > 0$
Full stochastic system	Dynamics of *state distribution*	Unstable for small ϑ

The instability arises not from the fixed point being unstable, but from **variance growth**: large prediction errors drive large updates, and when the gain K_z is high, the variance of the state grows even though its mean remains stable.

This is analogous to noise-induced phenomena in other dynamical systems (e.g., stochastic resonance, noise-induced transitions).

Implications:

1. Linear stability analysis is insufficient for stochastic systems
2. The observed bifurcations are properties of the *stochastic* HGF, not artifacts

3. Proving stochastic instability requires analyzing the evolution of the full probability distribution

3.4 Lyapunov Exponent Computation

Definition 3.1 (Maximal Lyapunov Exponent)**.** The maximal Lyapunov exponent characterizes the average rate of separation of nearby trajectories:

$$\lambda_{\max} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \ln \frac{\|\delta \mathbf{s}^{(t)}\|}{\|\delta \mathbf{s}^{(t-1)}\|} \tag{18}$$

where $\delta \mathbf{s}^{(t)}$ is the separation vector, periodically renormalized to prevent overflow.

Empirical Observation 3.2. For $\vartheta = 0.15$ (within the chaotic regime):

Approximation	$\lambda_{\max}$	95% CI
Mean-field	$+0.8$ to $+1.5$	varies with seed
Structured	-0.05 to $+0.02$	typically < 0

The structured approximation yields $\lambda_{\max}$ near zero or slightly negative, indicating stable dynamics. The mean-field approximation yields positive $\lambda_{\max}$, indicating chaos.

3.5 Interpretation

The structured approximation's stability advantage can be understood intuitively: by maintaining cross-level correlations, it provides a "damping" effect where beliefs at each level constrain each other. Under mean-field, the levels update independently, allowing oscillations to grow unchecked.

However, we caution against over-interpreting these results:

1. Our analysis is limited to the two-level HGF; behavior in deeper hierarchies may differ

2. The specific bifurcation thresholds are parameter-dependent

3. Biological inference systems likely differ in important ways from the idealized models studied here

The first concern—generalization to deeper hierarchies—we address directly in the following section.

3.6 Extension to Three-Level Hierarchies

A significant limitation of the preceding analysis is its restriction to a two-level hierarchy. Biological inference systems operate over many levels: from sensory processing through perceptual inference to abstract reasoning. Does the CF-stability relationship hold in deeper hierarchies, or is it an artifact of our simplified model?

To address this question, we extend our analysis to the three-level HGF, which introduces a *meta-volatility* level tracking how the volatility itself changes over time.

3.6.1 Three-Level Generative Model

The three-level HGF has the following structure (Mathys et al., 2014):

Level 3 (Meta-volatility):

$$z_{3,t} \mid z_{3,t-1} \sim \mathcal{N}(z_{3,t-1}, \vartheta_3^{-1}) \tag{19}$$

Level 2 (Log-volatility):

$$z_{2,t} \mid z_{2,t-1}, z_{3,t} \sim \mathcal{N}(z_{2,t-1}, \exp(\kappa_3 z_{3,t} + \omega_3)) \tag{20}$$

Level 1 (Hidden state):

$$z_{1,t} \mid z_{1,t-1}, z_{2,t} \sim \mathcal{N}(z_{1,t-1}, \exp(\kappa_2 z_{2,t} + \omega_2)) \tag{21}$$

Observations:

$$y_t \mid z_{1,t} \sim \mathcal{N}(z_{1,t}, \pi_u^{-1}) \tag{22}$$

This model captures environments where not only the hidden state changes (level 1), but the rate of change varies (level 2), and the stability of that rate also varies (level 3).

3.6.2 Approximation Schemes

With three levels, we have additional approximation options:

Full mean-field:

$$q(z_1, z_2, z_3) = q(z_1)q(z_2)q(z_3) \tag{23}$$

All levels update independently.

Fully structured (Markov):

$$q(z_1, z_2, z_3) = q(z_3)q(z_2 \mid z_3)q(z_1 \mid z_2) \tag{24}$$

Respects the generative model's conditional independence structure.

Bottom-structured only:

$$q(z_1, z_2, z_3) = q(z_3)q(z_2)q(z_1 \mid z_2) \tag{25}$$

Maintains coupling only at the 1–2 interface (sensory–volatility).

Top-structured only:

$$q(z_1, z_2, z_3) = q(z_3)q(z_2 \mid z_3)q(z_1) \tag{26}$$

Maintains coupling only at the 2–3 interface (volatility–meta-volatility).

These partial structuring variants allow us to test which hierarchical interface is most critical for stability.

3.6.3 Pairwise Circulatory Fidelity

For three levels, we define CF at each interface:

$$\mathrm{CF}_{12} = \frac{I_q(z_1; z_2)}{\min(H_q(z_1), H_q(z_2))} \tag{27}$$

$$\mathrm{CF}_{23} = \frac{I_q(z_2; z_3)}{\min(H_q(z_2), H_q(z_3))} \tag{28}$$

Under mean-field, both $\mathrm{CF}_{12} = \mathrm{CF}_{23} = 0$. Under structured approximations, one or both may be positive depending on which interfaces maintain coupling.

3.6.4 Simulation Results

We simulated the three-level HGF with parameters $\kappa_2 = \kappa_3 = 1$, $\omega_2 = \omega_3 = -2$, $\pi_u = 10$, varying the meta-volatility parameter ϑ_3.

Finding 1: Mean-field instability is amplified with depth.

System	Mean-Field Var(μ_2)		Structured Var(μ_2)	
	Value	Change	Value	Change
Two-level	0.17	—	0.10	—
Three-level	14.49	**85**×	0.56	5.6×

Adding a third level increases mean-field instability by a factor of 85, while structured approximation increases only 5.6-fold. The protective effect of structured approximation becomes *more* important in deeper hierarchies.

Finding 2: The lower interface (1–2) is more critical for stability.

Approximation	Var(μ_2)	Relative to Structured
Fully structured	0.56	1.0×
Bottom-only (1–2 coupling)	0.89	1.6×
Mean-field	14.49	26×
Top-only (2–3 coupling)	41.44	74×

Bottom-structured approximation achieves 94% of the full structuring benefit. Top-only structuring is *worse* than mean-field, suggesting that coupling at the sensory–volatility interface is the critical bottleneck.

Finding 3: Mean-field freezes higher levels.

Under mean-field approximation, level 3 dynamics are effectively frozen:

Level	Mean-Field Var	Structured Var	Ratio
Level 1 (state)	1312.97	1312.95	1.0×
Level 2 (volatility)	14.93	0.54	27.8×
Level 3 (meta-vol)	0.0005	1.29	0.0004×

Mean-field level 3 variance is essentially zero—the system has effectively collapsed to a two-level model with frozen meta-volatility. This "freezing" is a pathological consequence of the independence assumption: without information flow from lower levels, level 3 has no signal to update on.

Finding 4: CF discriminates approximation schemes.

Scheme	CF_{12}	CF_{23}	I_{12} (nats)	I_{23} (nats)
Mean-field	0.00	0.00	0.00	0.00
Structured	0.00	0.37	0.00	0.41
Bottom-only	0.00	0.00	0.00	0.00
Top-only	0.01	0.00	0.03	0.04

Mean-field shows CF ≈ 0 at both interfaces, confirming complete decoupling. The structured approximation maintains substantial mutual information at the 2–3 interface ($I_{23} = 0.41$ nats).

3.6.5 Interpretation

The three-level extension provides several important insights:

1. **CF thesis strengthened:** The relationship between CF and stability holds—and becomes *stronger*—in deeper hierarchies. Mean-field's problems are amplified, not attenuated, with depth.

2. **Interface criticality:** Not all hierarchical interfaces are equally important. The lower interface (sensory–volatility) is critical; the upper interface (volatility–meta-volatility) contributes less to stability. This asymmetry may reflect the different signal-to-noise ratios at different levels.

3. **Cascade dynamics:** Mean-field approximation causes pathological "freezing" of higher levels, effectively truncating the hierarchy. Structured approximation maintains active dynamics throughout.

4. **Depth-dependent vulnerability:** If biological inference uses approximations with low CF, deeper hierarchical processing (e.g., abstract reasoning, long-horizon planning) may be disproportionately affected.

Implications for neuromodulation: If dopamine modulates precision at hierarchical interfaces, these results suggest it may need to act preferentially at lower interfaces, or coordinate across multiple interfaces to maintain stability throughout the hierarchy.

4 Resource-Rational Extensions

4.1 Motivation

The structured approximation provides greater stability but requires more computation—it must track the conditional distribution $q(x|z)$ rather than just marginals. This suggests a trade-off: agents with limited computational resources might prefer mean-field approximations when volatility is low but switch to structured approximations when volatility is high.

We formalize this intuition following the resource-rationality framework of Lieder and Griffiths (2020).

4.2 Deriving a Cost Function

We seek a cost function $C(q)$ that captures the computational burden of maintaining statistical dependencies between hierarchical levels. We derive this from the structure of the variational updates themselves.

The computational asymmetry: Under mean-field $q(x, z) = q(x)q(z)$, the update for each level depends only on its own sufficient statistics. Under structured $q(x, z) = q(z)q(x|z)$, updating z requires integrating over the conditional distribution of x, and vice versa.

Concretely, for Gaussian posteriors:

- **Mean-field:** Store and update 4 parameters $(\mu_z, \sigma_z^2, \mu_x, \sigma_x^2)$ independently

- **Structured:** Store and update 5+ parameters, including conditional parameters (e.g., the slope b in $\mathbb{E}[x|z] = a + bz$), with coupled updates

Proposal: The computational cost of an approximation q is proportional to the mutual information it maintains:

$$C(q) = c_0 \cdot I_q(z; x) \tag{29}$$

where $c_0 > 0$ is a constant converting nats to computational cost units.

Justification:

1. **Operational interpretation:** $I(z; x)$ measures the statistical dependency between levels in bits (or nats). Each bit of mutual information represents a constraint that must be maintained across updates: when z changes, the system must update not just $q(z)$ but also how $q(x|z)$ depends on the new z value. Under mean-field, $I(z; x) = 0$, so no such cross-level bookkeeping is required.

2. **Update complexity:** In message-passing implementations, maintaining $I(z; x) > 0$ requires passing messages between levels that encode conditional sufficient statistics. The information content of these messages scales with $I(z; x)$.

3. **Correct limiting behavior:**

   - $C(q) = 0$ when q is mean-field (since $I(z; x) = 0$ for independent variables)
   - $C(q) > 0$ for any structured approximation with non-trivial dependency
   - $C(q)$ increases monotonically with dependency strength

4. **Connection to description length:** From a minimum description length perspective, $I(z; x)$ is the coding cost saved by exploiting the dependency between z and x. Conversely, maintaining this dependency (rather than discarding it) requires the system to represent and update this additional structure.

For Gaussian approximations, we can compute this explicitly. If $q(z, x)$ is jointly Gaussian with correlation ρ, then:

$$I(z; x) = -\frac{1}{2} \ln(1 - \rho^2) \tag{30}$$

This increases from 0 (when $\rho = 0$, mean-field) toward infinity (as $\rho \to \pm 1$, deterministic relationship).

4.3 Resource-Rational Free Energy

Definition 4.1 (Resource-Rational Free Energy)**.**

$$F_{\mathrm{RR}} = F_{\mathrm{VFE}} + \beta \cdot I_q(z; x) \tag{31}$$

where F_{VFE} is the standard variational free energy, $I_q(z; x)$ is the mutual information under q, and $\beta > 0$ weights the cost-accuracy trade-off (absorbing c_0).

The optimal approximation minimizes F_{RR}, balancing:

- **Accuracy** (lower F_{VFE}): Structured approximations typically achieve lower free energy by better capturing the true posterior
- **Cost** (lower I_q): Mean-field approximations are cheaper by discarding cross-level information

Note on CF vs. $I(z; x)$: We use the *unnormalized* mutual information $I(z; x)$ for the cost function, not the normalized Circulatory Fidelity $\mathrm{CF} = I(z; x)/\min(H(z), H(x))$. The unnormalized form is appropriate here because computational cost should scale with the absolute amount of dependency, not the relative amount. CF remains useful for comparing approximation quality across systems with different entropy scales.

4.4 Relationship to Thermodynamics

There is a suggestive connection between maintaining mutual information and thermodynamic dissipation. Landauer's principle establishes that erasing one bit of information requires at minimum $k_B T \ln(2)$ joules. By analogy, maintaining $I(z; x)$ bits of correlation might require ongoing energetic expenditure.

However, we caution against strong thermodynamic claims:

1. The relevant costs in neural systems may be computational (time, memory) rather than energetic
2. Inference updates may not constitute logically irreversible operations in Landauer's sense
3. Empirical estimates of neural signaling costs ($\sim 10^4$ ATP/bit; Laughlin et al., 1998) measure information *transmission*, not correlation *maintenance*

We therefore frame our cost function as *resource-rational* (bounded computation) rather than thermodynamic (bounded energy), while noting the formal similarities.

4.5 Optimal Precision

If we model the cost of high-precision inference as:

$$C(\gamma) = \gamma \ln(\gamma/\gamma_0) \tag{32}$$

then the resource-rational optimal precision is:

$$\gamma^* = \gamma_0 \exp\left(-1 - \frac{1}{\beta}\frac{\partial F_{\text{VFE}}}{\partial \gamma}\right) \tag{33}$$

Testable implication: If this trade-off operates in biological systems, we should observe bounded precision that does not increase indefinitely with prediction error magnitude.

5 A Candidate Neural Implementation

5.1 Prefatory Remarks

This section is **highly speculative**. We propose one possible mapping between the computational framework and neural mechanisms. This mapping is a hypothesis for future testing, not a claim about established neuroscience. Alternative mappings are possible, and the dopamine system's computational role remains actively debated.

5.2 Competing Theories of Dopamine Function

Before presenting our hypothesis, we acknowledge that dopamine's computational role is contested. Major frameworks include:

Reward Prediction Error (RPE) theory (Schultz et al., 1997): Phasic dopamine signals encode the difference between received and expected reward. This is the dominant interpretation of midbrain dopamine neuron firing, supported by extensive electrophysiological evidence.

Precision/gain modulation theory (Friston et al., 2012; Schwartenbeck et al., 2015): Dopamine modulates the precision or gain of neural computations, affecting how strongly prediction errors influence belief updates.

Motivational salience theory: Dopamine signals the motivational significance of stimuli, regardless of valence.

Vigor/effort theory: Dopamine modulates the vigor of actions and willingness to expend effort.

These theories are not mutually exclusive—dopamine likely serves multiple computational functions, possibly via distinct circuits or timescales. Our proposal aligns most

closely with the precision/gain modulation view, but we emphasize that this is one interpretation among several viable alternatives.

5.3 A Possible Dopamine-Precision Mapping

If dopamine modulates precision (which is not established), and *if* tonic dopamine concentration is the relevant signal (rather than phasic bursts or receptor-specific effects), then we can ask: what functional form might relate dopamine concentration to precision?

We propose (tentatively) that tonic dopamine concentration *could* implement a precision-like parameter γ:

- Higher tonic dopamine $\rightarrow$ higher γ $\rightarrow$ stronger weighting of prediction errors
- Lower tonic dopamine $\rightarrow$ lower γ $\rightarrow$ greater reliance on prior beliefs

5.4 One Possible Transfer Function: Hill Kinetics

Rather than positing an arbitrary functional form, we derive *one candidate* transfer function from receptor binding kinetics. This grounds the curve shape in biophysics, though the mapping from receptor occupancy to precision remains assumed.

Consider dopamine (D) binding to D2 receptors (R) with dissociation constant K_d:

$$R + D \rightleftharpoons RD, \quad K_d = \frac{[R][D]}{[RD]} \tag{34}$$

At equilibrium, the fractional receptor occupancy θ is given by the Hill equation:

$$\theta = \frac{D^n}{K_d^n + D^n} \tag{35}$$

where n is the Hill coefficient ($n \approx 1$ for D2 receptors).

Linking assumption (not derived): We assume that D2 receptor occupancy modulates neural gain, and that this gain corresponds to precision:

$$\gamma(D) = \gamma_{\max} \cdot \theta = \gamma_{\max} \cdot \frac{D^n}{K_d^n + D^n} \tag{36}$$

What is derived vs. assumed:

Component	Status	Basis
Hill equation form	Derived	Equilibrium thermodynamics
$K_d \approx 20$ nM	Constrained	Measured D2 receptor affinity
$n \approx 1$	Constrained	D2 receptor biophysics
Occupancy $\rightarrow$ gain	**Assumed**	Plausible but unproven
Gain $\rightarrow$ precision	**Assumed**	Theoretical hypothesis
$\gamma_{\max}$	Free parameter	Must be fit to data

The chain of assumptions (occupancy $\rightarrow$ gain $\rightarrow$ precision) is the weakest link. Alternative mappings—e.g., through intracellular signaling cascades, synaptic plasticity, or network-level effects—might be more appropriate.

5.5 Relationship to RPE Theory

Our proposal does not contradict RPE theory but suggests a complementary role:

- *Phasic* dopamine (rapid bursts) $\rightarrow$ RPE signaling (well-established)
- *Tonic* dopamine (baseline levels) $\rightarrow$ precision modulation (speculative)

This phasic/tonic distinction *might* map onto receptor subtypes:

- D2 receptors ($K_d \approx$ 10–30 nM) have affinity in the tonic range
- D1 receptors ($K_d \approx$ 1–10 μM) require phasic burst concentrations

However, we caution that this is an oversimplification. Both receptor types are present throughout dopamine circuits, their effects depend on cellular context, and the phasic/tonic distinction itself is debated. We offer this as one possible interpretation, not as settled neuroscience.

6 Computational Methods

6.1 Implementation

We implement the HGF and variational inference using custom Julia code. The core update equations follow Mathys et al. (2014) with modifications for the structured approximation.

Simplifications:

1. We use Gaussian approximations throughout
2. The structured approximation uses a first-order approximation for coupling terms
3. Lyapunov exponents are computed for the mean dynamics, ignoring posterior variance evolution

6.2 Lyapunov Computation

We compute Lyapunov exponents using the Benettin algorithm:

1. Initialize reference and perturbed trajectories with separation $\varepsilon = 10^{-8}$
2. Evolve both trajectories under identical inputs
3. Every k steps ($k = 10$), measure separation, add log(separation/ε) to running sum, renormalize
4. Average over trajectory length, discarding initial transient (1000 steps)

6.3 Code Availability

Full implementation code is provided in the accompanying repository, including core model specification, Lyapunov computation, and bifurcation diagram generation.

7 Proposed Experimental Tests

7.1 Computational Phenotyping

Proposal: Fit HGF models to behavioral data from reversal learning tasks under different pharmacological conditions.

Prediction: D2 antagonists $\rightarrow$ reduced CF $\rightarrow$ better fit by mean-field models.

7.2 Dynamical Signatures

Proposal: Analyze trial-by-trial learning rates for oscillatory signatures.

Prediction: Power spectrum of learning rate time series might show structure in conditions associated with low CF.

Strong caveat: Biological noise may completely obscure any bifurcation structure.

7.3 Clinical Populations

Speculative predictions:

Condition	Putative DA State	CF Prediction	Behavioral Prediction
Parkinson's (off med)	Low tonic DA	Low CF	Over-reliance on priors
Schizophrenia	Elevated striatal DA	High CF	Potentially unstable inference

7.4 Crucial Experiment: Spectral Signatures of Period-Doubling

The editorial challenge is: what does CF predict that alternatives do not?

The distinguishing prediction: CF theory predicts not merely that reduced dopamine causes noisier inference (which any theory predicts), but that it causes **qualitatively different dynamics**—specifically, quasi-periodic oscillations reflecting period-doubling bifurcations.

Experimental design: Within-subject pharmacological study using D2 antagonist (sulpiride 400mg) vs. placebo in a volatile reversal learning task.

Primary outcome: Power spectral density of trial-by-trial learning rates.

CF prediction: Under sulpiride, spectral power in the 0.02–0.10 cycles/trial band should increase, with emergent peaks at frequencies related by ratio ≈ 2.

Prediction	CF Theory	Alternatives
Increased variability	✓	✓
Oscillatory structure at specific frequencies	✓	—

Falsification criteria: CF would be falsified if sulpiride produces expected behavioral effects but no spectral signatures.

8 Discussion

8.1 Summary of Contributions

1. **Formal definition of Circulatory Fidelity:** A normalized measure of cross-level dependency with **proven** properties (CF = 0 under mean-field)

2. **Information-geometric characterization: Proof** that the FIM is block-diagonal under mean-field

3. **Stability analysis:** Derivation showing deterministic skeleton is stable; **numerical observation** that stochastic system exhibits bifurcations

4. **Stochastic instability hypothesis:** Proposal that instabilities arise from variance growth

5. **Resource-rational cost function:** Derivation of $C(q) = c_0 \cdot I_q(z; x)$ from the operational requirement that statistical dependencies must be maintained across updates

6. **Three-level extension:** Demonstration that CF-stability relationship **strengthens** in deeper hierarchies, with 85× amplification of mean-field instability

7. **Interface criticality finding:** Discovery that lower hierarchical interfaces are more critical for stability than upper interfaces

8. **Neural implementation hypothesis:** Speculative but testable proposal linking CF to dopaminergic neuromodulation

9. **Crucial experiment:** Protocol testing the unique prediction of CF

8.2 What This Work Does Not Show

1. **Not a novel information-theoretic measure:** The normalized mutual information we call "Circulatory Fidelity" is the established uncertainty coefficient (Coombs et al., 1970). Our contribution is its application to variational inference stability, not the measure itself.

2. **Not a proof of biological mechanism:** The dopamine-precision mapping is one hypothesis among several viable alternatives (RPE theory, motivational salience, etc.)

3. **Limited to Gaussian hierarchies:** While we extend to three levels, the analysis remains restricted to Gaussian generative models; non-Gaussian cases may behave differently

4. **Not empirically validated:** The predictions remain untested

5. **Not novel in demonstrating structured advantages:** The superiority of structured approximations is known (Parr et al., 2019)

8.3 Relationship to Prior Work

The advantages of structured over mean-field variational inference are well-established (Parr et al., 2019; Schwöbel et al., 2018). Our contribution is to characterize this advantage in dynamical systems terms.

The resource-rationality framework (Lieder and Griffiths, 2020) provides the normative foundation for our cost-accuracy trade-off.

The dopamine-precision hypothesis has been proposed by others (Friston et al., 2012). Our contribution is to connect it to the stability analysis.

What distinguishes CF: Previous frameworks predict structured approximations are "better." CF additionally predicts **specific dynamical signatures** that should manifest when the system operates near the mean-field regime.

8.4 Future Directions

Theoretical:

- Extend analysis to four or more levels (preliminary three-level results suggest continued amplification)
- Develop continuous-time formulation
- **Prove the stochastic instability mechanism**
- Investigate interface-specific modulation: why is the lower interface more critical?
- Analyze non-Gaussian generative models

Computational:

- Systematic parameter sweeps across κ, ω, ϑ space
- Comparison with particle filtering and MCMC

Empirical:

- Computational phenotyping studies
- Pharmacological manipulations
- Clinical population comparisons

9 Conclusion

This thesis has introduced Circulatory Fidelity as a measure of cross-level dependency in hierarchical variational inference and analyzed its relationship to dynamical stability. Our main findings are:

1. Mean-field variational inference in the HGF exhibits period-doubling bifurcations and chaos at high volatility
2. Structured approximations that maintain CF > 0 remain stable across a broader parameter range
3. This stability difference can be understood geometrically through the Fisher Information Matrix structure

We have proposed, speculatively, that biological inference systems may implement mechanisms analogous to CF maintenance, potentially through dopaminergic precision modulation. This proposal generates testable predictions that await empirical investigation.

The framework is offered as a theoretical tool for generating hypotheses, not as a completed theory of neural computation. Its value will ultimately be determined by whether the predictions it generates are borne out by experiment.

References

Coombs, C. H., Dawes, R. M., and Tversky, A. (1970). *Mathematical Psychology: An Elementary Introduction*. Prentice-Hall, Englewood Cliffs, NJ.

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., Dolan, R. J., Moran, R., Stephan, K. E., and Bestmann, S. (2012). Dopamine, affordance and active inference. *PLoS Computational Biology*, 8(1):e1002327.

Laughlin, S. B., de Ruyter van Steveninck, R. R., and Anderson, J. C. (1998). The metabolic cost of neural information. *Nature Neuroscience*, 1(1):36–41.

Lieder, F. and Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43:e1.

Mathys, C., Daunizeau, J., Friston, K. J., and Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, 5:39.

Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., and Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience*, 8:825.

Parr, T., Markovic, D., Kiebel, S. J., and Friston, K. J. (2019). Neuronal message passing using mean-field, Bethe, and marginal approximations. *Scientific Reports*, 9:1889.

Press, S. J. (1967). On the sample coefficient of contingency. *The Annals of Mathematical Statistics*, 38(5):1575–1576.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599.

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., and Friston, K. (2015). Optimal inference with suboptimal models: Addiction and active Bayesian inference. *Medical Hypotheses*, 84(2):109–117.

Schwöbel, S., Kiebel, S., and Markovic, D. (2018). Active inference, belief propagation, and the Bethe approximation. *Neural Computation*, 30(9):2530–2567.

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, 76(1):103–154.

A Proof Details

Complete proofs with full derivations are provided in the supplementary document. Here we summarize key results.

A.1 Proposition 1 (FIM Block-Diagonality) — PROVEN

Statement: Under the mean-field approximation $q(x, z) = q(x)q(z)$ with Gaussian marginals, the Fisher Information Matrix is block-diagonal.

Proof: The FIM is defined as $G_{ij} = \mathbb{E}_q[\partial_i \ln q \cdot \partial_j \ln q]$.

Under mean-field: $\ln q(x, z) = \ln q(z) + \ln q(x)$

For cross-terms:

$$G_{\theta_z,\theta_x} = \mathbb{E}_{q(z)q(x)}\left[\frac{\partial \ln q(z)}{\partial \theta_z} \cdot \frac{\partial \ln q(x)}{\partial \theta_x}\right] \tag{37}$$

By independence:

$$= \mathbb{E}_{q(z)}\left[\frac{\partial \ln q(z)}{\partial \theta_z}\right] \cdot \mathbb{E}_{q(x)}\left[\frac{\partial \ln q(x)}{\partial \theta_x}\right] \tag{38}$$

The score function has zero mean for exponential families, so all cross-terms vanish. □

A.2 Proposition 2 (Local Stability) — DERIVED

Statement: Under the deterministic skeleton, the Jacobian eigenvalue is:

$$\lambda = \frac{\vartheta}{\vartheta + \alpha} \tag{39}$$

Key finding: $0 < \lambda < 1$ for all $\vartheta > 0$, implying stability.

Resolution: The instability arises from *stochastic* dynamics, not the deterministic skeleton.

A.3 Proposition 3 (CF Under Mean-Field) — PROVEN

Statement: Under mean-field, CF = 0.

Proof: Mutual information: $I(z; x) = H(z) + H(x) - H(z, x)$. Under independence: $H(z, x) = H(z) + H(x)$. Therefore $I(z; x) = 0$, and CF = 0. □

A.4 On Bifurcations — NUMERICAL OBSERVATIONS

The period-doubling bifurcations are **numerical observations**, not proven results. Observed thresholds:

- First bifurcation: $\vartheta_c \in [0.04, 0.06]$
- Chaos onset: $\vartheta_{\text{chaos}} \in [0.10, 0.15]$

A.5 On the Structured Update Approximation — MODELING CHOICE

The coupling coefficient $\gamma_{zx} = \kappa\pi_x/4$ is a **modeling approximation**, not a derived result.

A first-principles derivation would compute the natural gradient:

$$\Delta\boldsymbol{\theta} = -\eta\mathbf{G}^{-1}\nabla_{\boldsymbol{\theta}}F \tag{40}$$

The off-diagonal FIM terms would determine the exact coupling. This calculation is tractable but lengthy; we leave it for future work.

Epistemic status: The qualitative prediction (structured approximations are more stable) does not depend on the exact value of γ_{zx}.

A.6 Summary of Epistemic Status

Claim	Status
FIM block-diagonal under mean-field	**Proven**
CF $= 0$ under mean-field	**Proven**
CF > 0 under structured	**Proven**
Deterministic skeleton stable	**Derived**
Stochastic instability mechanism	**Hypothesized**
Period-doubling occurs	**Observed**
Specific bifurcation thresholds	**Observed**
Structured prevents bifurcation	**Observed**
Structured update coefficient	**Modeling approximation**

B Simulation Parameters

Default parameters:

Parameter	Symbol	Value	Justification
Coupling strength	κ	1.0	Standard value
Baseline log-volatility	ω	-2.0	Reasonable range
Observation precision	π_u	10.0	Moderate noise
Volatility of volatility	ϑ	varies	Primary parameter
Simulation length	T	10,000	Lyapunov convergence
Transient discarded	—	1,000	Remove transients
Lyapunov renormalization	—	every 10 steps	Standard practice
Random seed	—	42	Reproducibility