# Verrell’s Law Ψμν / Collapse-Aware AI — Referee Response & Path B Revision Notes (v1.2)

Context: Response to referee feedback on Zenodo DOI 10.5281/zenodo.17416435 (v1.1b).

Objective: Reframe the framework as Computational Systems Theory and supply explicit, testable definitions for the core equation, metrics, and glossary.

### 1. Core Equation (Computational Reformulation)

We replace the former physics-laden “Core Equation Snapshot” with a stochastic differential form defined in logit space:

$$dz_t = b\Psi(z_t, M_t)\, dt + \Sigma\, dW_t$$

Here, $z_t$ is the logit vector, $p_t = \mathrm{softmax}(z_t/T)$, $b\Psi$ is the bias drift (logit/s), $\Sigma$ is the diffusion matrix (logit/s½), and $W_t$ is a Wiener process.

$$b\Psi = \alpha\, \nabla z \log \pi\_prior(z|M_t) + \beta\, \nabla z \log \pi\_anchor(z) - \gamma\, \nabla z\, H(p_t)$$

This reframes 'collapse' as probabilistic resolution under adaptive bias control rather than a physical field interaction.

### 2. Defined Metrics (for falsifiability)

| Symbol | Definition | Interpretation |
|---|---|---|
| $R_b$ | $\mathrm{corr}(\Delta z_t, b_t)$ | Response-to-bias correlation |
| $S_b$ | $\mathrm{Var}_t[R_b(t)]$ | Sessional stability |
| $\Delta\_KL$ | mean $KL(p\blacksquare \| p\blacksquare)$ | Distributional shift magnitude |
| $\Delta\_prior$ | mean $[KL(p\blacksquare \| \pi\_prior) – KL(p\blacksquare \| \pi\_prior)]$ | Alignment to memory prior |
| $\Delta\_anchor$ | mean $[KL(p\blacksquare \| \pi\_anchor) – KL(p\blacksquare \| \pi\_anchor)]$ | Stabilization against anchor |

Controls: fixed seeds, isotemperature runs, static-bias baseline, bootstrap CIs ($p < 0.01$).

KV-cache statistics: mean key-vector drift ($D_K$) and attention entropy ($H_{attn}$).

### 3. Terminology Update

Physics terminology has been replaced with computational systems theory language for clarity.

| Old term | Revised term |
|---|---|
| Collapse | Probabilistic Resolution |
| Bias Field | State-Space Bias Operator |
| Resonance | Feedback Weighting |
| Observer Effect | Contextual Conditioning |
| Memory = Information | Memory-Conditioned Prior |
| Governor | Adaptive Gain Regulator |

### 4. Probability Conservation (Sketch)

The induced density $p(x,t)$ obeys $\partial t\ p = -\nabla\cdot(b\Psi\ p) + \frac{1}{2}\nabla\cdot(D\nabla p)$ with $D = \Sigma\Sigma$■. Zero-flux boundaries ensure $d/dt \int p\ dx = 0$; stationary $p^* \propto e^{-U\Psi}$ exists for $D$ ■ $0$.

### 5. Author Response Summary

We acknowledge the referee’s requests and confirm the following revisions: 1. Dimensional consistency established. 2. Physical analogy removed. 3. Empirical reproducibility through explicit metrics. 4. Governance clarified as adaptive gain regulator. These revisions position the framework as speculative but internally coherent computational systems theory, satisfying the 'publishable with revisions' criterion.