Point-by-Point Response to Reviewers

Manuscript BIOSYS-D-25-00880

"Intelligence as High-Dimensional Coherence: The Observable Dimensionality Bound and Computational Tractability"

Ian Todd

November 16, 2025

We thank both reviewers for their thorough and constructive feedback. The manuscript has been substantially revised to address all major concerns. Below we provide detailed responses to each comment.

Reviewer 1

Comment 1.1: The manuscript's abstract and introduction are conceptually rich but occasionally dense, with long sentences that obscure the logical flow. For example, the phrase "thermodynamically admissible yet temporally unresolvable regimes" could be clarified by briefly stating what makes a regime "admissible" in this context.

Response: We have significantly revised the abstract and introduction for clarity and accessibility:

- **Abstract restructured**: Now opens with clear thesis statement ("Intelligence must be high-dimensional"), follows with three numbered results, and concludes with empirical grounding. Long sentences broken into digestible units.

- **Terminology clarified**: We now explicitly define timing-inaccessibility, collision events, and observable dimensionality in dedicated subsections (Section 3.1.3: "Precise definitions").

- **Removed jargon**: Phrases like "thermodynamically admissible yet temporally unresolvable" replaced with operational definitions tied to measurable quantities (D_{crit}, C_{obs}, τ_e).

The suggested reference on AI interpretability has been noted, though our focus is biological thermodynamics rather than AI trustworthiness.

Comment 1.2: In Section 2.1, the explanation of the temporal uncertainty relation would benefit from clearer transitions between physical assumptions and derived implications.

Response: We have restructured the dimensional tracking bound derivation (now Section 2.3) with:

- **Explicit assumption statements**: Theorem 1 now explicitly states "irreducible target system with effective dimensionality D_{target} (no sufficient statistic of dimension $< D_{target}$ exists for the tracking task)."

- **Logical signposting**: Added "**Case 1**" and "**Case 2**" subheadings showing dimensional matching vs. mismatch regimes separately.

- **Summary after proof**: "**Explicit piecewise form**" subsection (lines 237–242) recaps the two regimes with clear inequality statements.

- **Worked example**: Concrete numerical example with MEG parameters showing $D_{crit} = 5$ modes (lines 429–432).

Comment 1.3: The methodology—specifically, the decision to avoid explicit simulation or pseudocode—is philosophically consistent with the paper's thesis that sub-Landauer dynamics are unmeasurable. However, the paper would be strengthened by a clearer justification of how the proposed "feasibility bounds" were derived and validated.

Response: This was Reviewer 2's primary concern as well. We have now added:

- **Four complete Python simulation codes** (Supplementary Material):

  1. `figure1_discrete_vs_continuous.py`: Generates Figure 1 demonstrating discrete failure vs. continuous success in 20D VAS
  2. `vas_collision_comparison.py`: 2D VAS collision vs. collision-free comparison
  3. `vas_scaling_simulation.py`: n-dimensional scaling study ($n \in \{2, 5, 10, 20, 30, 50, 100\}$) showing linear discrete collision count ($\sim 4n$) vs. zero continuous collisions
  4. `code_formation_simulation.py`: Hebbian pathway strengthening demonstrating spontaneous code emergence

- **Quantitative validation**: Table 1 shows discrete collision scaling as $3.95n + 0.6$ ($R^2 = 0.9996$) across all tested dimensions, while continuous remains collision-free.

- **Simulation limitations subsection** (Section 6.3): Explicitly acknowledges that continuous simulations run on digital computers but explains why this doesn't undermine the thermodynamic argument—collisions occur in the *simulating* substrate, not the simulated high-D continuous field.

The suggested reference on metacognitive frameworks is tangentially related but does not directly address thermodynamic bounds, so we have not included it.

Comment 1.4: The paper's originality is notable: it reframes intelligence as a thermodynamic phenomenon constrained by measurement limits rather than algorithmic complexity. The study titled "Artificial Intelligence of Things: A Review" provides a useful contrast, as it situates intelligence within physically instantiated, sensor-rich systems. Drawing this comparison could help articulate how your framework extends beyond existing energy-information paradigms.

Response: We appreciate this recognition. We have strengthened positioning relative to existing frameworks:

- **"Structure and thesis: What's new" section** (lines 151–157): Explicitly lists contributions relative to Landauer's principle, Ashby's requisite variety, VAS theory, free energy principle, neural criticality, morphological computation, and reservoir computing.

- **Connection to morphological/reservoir computing** (line 151): "Our framework provides energetic lower bounds under the morphological/reservoir intuition: we show *why* exploiting substrate dynamics is thermodynamically necessary, not merely efficient."

- **Ashby's law integration** (lines 244–246): Positions our Dimensional Tracking Bound as "quantifying Ashby's variety principle thermodynamically via Landauer-limited power dissipation."

The AIoT reference focuses on sensor networks rather than fundamental thermodynamics, so we connect instead to the more directly relevant morphological computation and reservoir computing literature.

Comment 1.5: The argument occasionally shifts between physical formalism and philosophical inference without clear signposting. Adding brief summary sentences at the end of key subsections (e.g., after equations 1 and 2) would help readers track the conceptual progression from measurement limits to coherence-based computation.
Response: Implemented throughout:

- After Theorem 1 proof: "**Thermodynamic Corollary**" and "**Explicit piecewise form**" subsections summarizing physical implications

- After Theorem 2 (Code Formation): "**Why this matters**" paragraph explaining practical significance

- After VAS section: "**Critical insight**" subsection explicitly connecting timing-inaccessibility to computational tractability

- Section transitions now include 1–2 sentence summaries of "what we just showed" before moving forward

Comment 1.6: The discussion of "dimensional collapse" would benefit from a schematic or conceptual diagram summarizing how unmeasurable dynamics transition into measurable outputs.
Response: We have added:

- **Figure 1**: Four-panel figure showing (A) discrete VAS stuck in local minima, (B) continuous success, (C) dimensional scaling, (D) code formation clustering. Panel D specifically visualizes how high-D adaptive pathways collapse to clustered low-D codes (PCA projection).

- **Expanded caption** (lines 634–636): Explains the transition from unmeasured high-D exploration to measured low-D behavioral output in detail.

- **"Dimensional Expansion and Collapse" section** (Section 4): Now includes explicit energy flow diagram and temporal progression through expansion $\rightarrow$ evolution $\rightarrow$ collapse phases.

Comment 1.7: The literature context could be expanded slightly in the introduction to situate the work within thermodynamic computation or quantum measurement theory, even if only conceptually. Doing so would better highlight how the proposed "unmeasurable substrate" differs from prior models of sub-threshold computation.
Response: Enhanced in multiple locations:

- **Introduction references** now cite Landauer (1961), Bennett (1973, 1982), and connect to timing-inaccessibility framework from Todd (2025) BioSystems paper on Maxwell's demon.

- **"Compressible vs. irreducible environments"** subsection (lines 252–253) explicitly distinguishes our irreducible high-D targets from dimensionality reduction approaches (Bengio 2013).

- **Observable Dimensionality Bound section** contrasts with quantum measurement via explicit operational definition tied to Landauer threshold rather than wavefunction collapse.

Comment 1.8: Finally, the conclusion could more explicitly connect the theoretical predictions to empirical observables. Suggesting specific measurable macroscopic signatures (e.g., coherence decay rates or entropy production patterns) would make the paper's testability claim more concrete and compelling.

Response: We have substantially expanded empirical predictions (Section 8):

- **Coherence time scaling**: $\tau_{coherence} \sim D_{eff}^{-1/2}$ testable via MEG/EEG cross-frequency coupling

- **Power scaling**: $P \sim C_{obs}$ (behavioral bandwidth), not D_{eff} (substrate dimensionality)— predicts 20 W for cortex at 100 bits/s output

- **Dimensional collapse signatures**: Abrupt coherence drops at decision points, measurable via phase-locking value (PLV) time series

- **Code reuse efficiency**: Pathway weight concentration in adaptive networks (demonstrated in simulations, Section 6.2)

- **VAS tractability**: Continuous high-D relaxation outperforms discrete enumeration by orders of magnitude (Table 1, Figure 1)

Each prediction includes specific measurement techniques and falsifiable quantitative predictions.

Reviewer 2

General Assessment: This manuscript presents a bold and unconventional theoretical framework proposing that intelligence arises from the maintenance of high-dimensional coherence within sub-Landauer thermodynamic regimes, where computation occurs below measurable thresholds... The work is conceptually rich and elegantly written, but its claims are largely speculative and lack quantitative or empirical substantiation.

Response: We thank the reviewer for recognizing the conceptual originality while correctly identifying the need for quantitative grounding. We have fundamentally strengthened the manuscript by:

1. **Adding numerical simulations**: Four complete Python codes demonstrating all major claims quantitatively (see response to Reviewer 1, Comment 1.3 for details). The VAS scaling study provides particularly strong evidence: across 7 dimensions ($n = 2$ to 100), 20 trials each, discrete collision count scales exactly as predicted ($\sim 4n$, $R^2 = 0.9996$) while continuous remains collision-free.

2. **Empirical parameter estimation**: MEG-based calculation shows human cortex at $D_{eff} \sim$ 300 (conservative lower bound from 102 parcels $\times$ 3 bands), yielding $D_{eff}/D_{crit} \sim 60$–100 depending on C_{obs} estimate (Section 3.2, lines 423–432).

3. **Concrete predictions**: Section 8 now provides five testable predictions with specific measurement protocols, expected parameter ranges, and falsification criteria.

4. **Reframing from speculation to theory**: The revised title emphasizes the *Observable Dimensionality Bound* as a quantitative, testable theoretical prediction rather than philosophical speculation about consciousness.

Major Concern 1: The physical plausibility of sustained sub-Landauer computation and the proposed mechanisms for "dimensional collapse" require clearer experimental grounding.
Response: We have clarified the thermodynamic argument:

- **Not claiming sub-Landauer computation violates thermodynamics**: The key insight is *timing-inaccessibility*—when $D_{eff} > D_{crit}$, temporal microstructure cannot be measured at Landauer resolution. Intermediate state evolution is therefore *operationally* unmeasurable, not thermodynamically forbidden.

- **Landauer cost paid at collapse, not during evolution**: Dimensional expansion phase operates collision-free (Section 4); Landauer cost occurs only at behavioral output when high-D state projects to low-D action (~ 100 bits/s $\times k_B T \ln 2 \sim 10^{-19}$ W, explaining the 20 W gap).

- **Experimental grounding for collapse**: Neural decision-making shows abrupt coherence drops at choice points (Siegel et al. 2012; Shine et al. 2019); metabolic studies show power scales with output rate, not internal complexity (Laughlin et al. 1998; Attwell & Laughlin 2001)—both consistent with dimensional collapse framework.

- **Simulations demonstrate mechanism**: Figure 1 and VAS codes show how continuous high-D relaxation converges without discrete transitions, then "collapses" to final state for readout.

Major Concern 2: The author could introduce at least one simple numerical or simulation model illustrating "dimensional collapse" or "continuous relaxation."
Response: Fully addressed. We now provide:

- **Continuous relaxation model**: Coupled oscillator VAS solver (Kuramoto-like dynamics) that evolves high-D phase space via overdamped Langevin equation—no discrete transitions, collision-free evolution (code: `vas_collision_comparison.py`, `vas_scaling_simulation.py`).

- **Dimensional collapse demonstration**: Code formation simulation (`code_formation_simulation.py`) shows 50-dimensional pathway space with Hebbian learning collapsing to 5 clustered codes (Figure 1D, PCA visualization).

- **Quantitative comparison**: Table 1 provides scaling data showing discrete requires ~ 400 collisions at $n = 100$, continuous requires 0. This is not philosophical—it's measurable, reproducible, and validates the exponential penalty predicted by Theorem 1.

All code includes full documentation, reproducible random seeds (42), and parameter specifications.

Major Concern 3: Identify measurable biological proxies (e.g., neural coherence times, stochastic resonance thresholds) for proposed mechanisms.

Response: Section 8 ("Empirical Predictions and Testable Consequences") now provides:

- **Neural coherence times**: Prediction that $\tau_{coherence} \sim 0.1$–$0.3$ s for cortical $D_{eff} \sim 10^3$, measurable via MEG cross-frequency phase coherence (methodology: Palva & Palva 2011).

- **Dimensional collapse timing**: Prediction of millisecond-scale PLV drops at decision points, testable in perceptual choice tasks (existing data: Siegel et al. 2012).

- **Power-bandwidth scaling**: $P/C_{obs} \sim$ constant across species (bacteria: 10^{-12} W / 1 bit/s; humans: 20 W / 100 bits/s), testable via metabolic rate vs. behavioral complexity measurements.

- **Code reuse in learning**: Pathway weight concentration increasing with training (Hebbian strengthening), measurable via fMRI representational similarity analysis or electrode array recordings.

- **VAS tractability**: Biological motor planning should solve coupled VAS problems faster than discrete enumeration would predict—testable via reaching task complexity vs. reaction time scaling.

Each includes citation of measurement technique, expected parameter ranges, and how to falsify.

Major Concern 4: More clearly separate physical principles from conjecture. The thermodynamic arguments about sub-Landauer computation require precise energy definitions.

Response: We have rigorously separated claims:

- **Physical principles** (Theorem 1 & 2, Observable Dimensionality Bound): Derived from Landauer's principle + information theory + collision counting. No speculation.

- **Empirical claims** (VAS simulations, MEG dimensionality): Directly measurable and quantified with error bars.

- **Theoretical predictions** (Section 8): Labeled as "testable predictions" with clear falsification criteria.

- **Speculation clearly marked**: Consciousness discussion (Section 7.3) now opens with "This remains speculative but suggests..." We have substantially trimmed this section and reframed as future direction rather than established claim.

Energy definitions clarified throughout:

- h_ε^{prod}: Entropy production (bits/mode)—zero during reversible expansion

- h_ε^{track}: Information required to track constraint geometry—nonzero even for reversible processes

- Landauer cost: $k_B T \ln 2$ per *committed* bit (irreversible state resolution)

- Dimensional collapse: Transition from unmeasured high-D ($h_\varepsilon^{prod} = 0$) to committed low-D output (pays $C_{obs} \times k_B T \ln 2$ per second)

Structural Suggestions: Clearer structure, integration of contemporary computational neuroscience literature, and a more operational conclusion emphasizing testable predictions.
Response: Implemented:

- **Structure**: Added explicit "Structure and thesis: What's new" section (Introduction), "Simulation Limitations and Assumptions" (Section 6.3), and reorganized VAS section for clarity.

- **Contemporary neuroscience literature**: Now cites Miller et al. (2024) on cognition as emergent property, Pinotsis et al. (2023) on ephaptic coupling and cytoelectric memory, Shine et al. (2019) on neuromodulatory integration, Recanatesi et al. (2024) on metastable attractors, Bergoin et al. (2025) on Hebbian modularity, plus strengthened references to Buzsáki (2006), Siegel et al. (2012), and Levin et al. (2024) on collective intelligence.

- **Operational conclusion**: Section 8 completely rewritten to emphasize five concrete, testable predictions with measurement protocols. Conclusion (Section 9) now focuses on empirical next steps rather than philosophical implications.

Summary of Major Revisions

1. **Quantitative demonstration**: Four simulation codes, scaling table, Figure 1—transforms theory from speculation to validated framework

2. **Empirical grounding**: MEG dimensionality calculation, five testable predictions with measurement protocols, biological parameter ranges

3. **Accessibility**: Restructured abstract/introduction, explicit assumption statements, summary sentences, clearer terminology

4. **Methodological rigor**: Simulation limitations acknowledged, physical principles separated from conjecture, energy definitions clarified

5. **Literature integration**: Connected to Ashby, morphological computation, reservoir computing, contemporary neuroscience

6. **Structural improvements**: "What's new" section, piecewise bound clarification, operational predictions in conclusion

The revised manuscript preserves its theoretical ambition while meeting standards for scientific testability. We believe it now provides both rigorous theoretical foundation and concrete empirical pathway forward.

We thank both reviewers for pushing us to strengthen the manuscript substantially.