

The Topological Signature of Consciousness: A Gromov-Wasserstein Framework for Neural-Phenomenal Alignment

Éric Reis

Independent Researcher

eirikreisena@gmail.com

January 2026

Abstract

Contemporary theories of consciousness identify neural correlates but lack formal metrics to quantify the structural isomorphism between neural dynamics and phenomenal experience. We introduce the **Phenomenal Manifold Hypothesis (PMH)**, which posits that consciousness is a geometric structure isometric to the information-theoretic topology of the neural substrate. We propose a rigorous methodology using **Gromov-Wasserstein (GW) Optimal Transport** to align the metric space of neural states (\mathcal{N}) with the metric space of phenomenal distinctions (Ψ) without assuming *a priori* coordinate correspondence. To refute the “manifold illusion” critique, we employ **Topological Data Analysis (TDA)** via persistent homology to distinguish genuine topological invariants from high-dimensional noise. A computational validation using a Color Ring toy model demonstrates the recovery of circular topology ($\beta_1 = 1$) from noisy spike trains where linear methods fail (Normalized GW cost: 0.042 ± 0.005 vs. Control 0.350 ± 0.020 , $p < 0.001$). Anchored in **Ontic Structural Realism**, this framework dissolves the “hard problem” by re-framing the mind-body relation as a testable problem of metric space alignment. We propose three falsification protocols: (1) topological phase transitions during anesthesia, (2) cross-modal geometric isometry, and (3) adversarial topological validation.

Keywords: Consciousness, Gromov-Wasserstein Distance, Neural Manifolds, Topological Data Analysis, Persistent Homology, Structural Realism.

1 Introduction

The scientific study of consciousness is currently hindered by a fundamental *metric gap*. While neuroimaging provides high-dimensional data on neural activity (voltage traces, BOLD signals), and psychophysics provides low-dimensional reports of phenomenal experience (similarity judgments, difference thresholds), there exists no rigorous mathematical framework to quantify the structural identity between these disparate domains [1, 2].

Current dominant theories, such as Integrated Information Theory (IIT) and Global Neuronal Workspace Theory (GNWT), offer necessary conditions for consciousness but often rely on scalar measures (e.g., Φ) or functional descriptions [3–5]. They do not explicitly model the *intrinsic geometry* of phenomenal space—the specific topological structure that dictates why, for instance, the color space is perceived as a closed loop (hue wheel) while pitch is perceived as linear (or helical) [21].

We propose the **Phenomenal Manifold Hypothesis (PMH)**, which asserts that phenomenal structure is not epiphenomenal but is *isomorphic* to the information-geometric structure of the neural substrate. **Scope:** This paper presents a *theoretical framework* with computational validation; empirical testing on biological neural data is proposed as future work.

To operationalize this, we introduce a framework combining:

1. **Gromov-Wasserstein (GW) Optimal Transport:** A metric-measure space alignment technique that does not require pre-aligned coordinates [6].
2. **Topological Data Analysis (TDA):** A validation method using persistent homology to ensure detected manifolds are not artifacts of dimensionality reduction [9].
3. **Ontic Structural Realism (OSR):** A philosophical grounding that treats structural isomorphism as identity, dissolving the explanatory gap [18].

The framework is illustrated concretely in Section 2.4, where we show how the phenomenal structure of color perception emerges from the hybrid metric formalism.

2 Theoretical Framework

2.1 Ontic Structural Realism and Identity

The PMH adopts Ontic Structural Realism (OSR) [18–20], positing that “things” are derivative of relational structures. In this view, if the relational structure of a neural manifold \mathcal{N} is isometric to that of a phenomenal manifold Ψ , they are physically identical. The “qualities” of experience are defined purely by their geometric location and relations within Ψ .

2.2 Formal Definitions

Definition 2.1 (Neural Space \mathcal{N}). *Let $\mathcal{N} \subset \mathbb{R}^N$ be the manifold of accessible neural states [14, 15]. It is equipped with a metric $d_{\mathcal{N}}$, typically the Fisher Information Metric [26], which quantifies the distinguishability of probabilistic neural codes.*

Definition 2.2 (Phenomenal Space Ψ). *Let $\Psi \subset \mathbb{R}^P$ be the manifold of subjective experiences. It is equipped with a metric d_{Ψ} representing perceived dissimilarity (e.g., Weber-Fechner distances).*

2.3 The Hybrid Phenomenal Metric

We posit that the geometry of Ψ is induced by the neural substrate through a specific deformation. The metric tensor g_{Ψ} is defined as:

$$g_{\Psi} = \pi^* g_{\text{info}} + h(\mathcal{I}, \Gamma, \Delta) \quad (1)$$

where:

- $\pi^* g_{\text{info}}$ is the pullback of the neural information metric.
- h is a **structural deformation tensor** dependent on three global invariants:
 - **Integration (\mathcal{I}):** Ensures topological connectedness (prevents fragmentation).
 - **Coherence (Γ):** The Kuramoto order parameter, acting as a binding constant. If $\Gamma \rightarrow 0$, the metric degenerates (loss of structure).
 - **Differentiation (Δ):** Entropy of the repertoire, expanding the manifold’s volume form.

This formulation explains why consciousness fades during seizures (high synchronization/ Γ , but zero differentiation/ Δ) or deep sleep (low \mathcal{I}): the metric structure itself collapses.

2.4 Illustrative Example: The Geometry of Color

To ground these formal definitions, consider the phenomenological structure of color perception. Standard psychophysics models color space \mathcal{C} as a solid (e.g., HSV cone or cylinder) where:

- **Hue** corresponds to the angular coordinate $\theta \in [0, 2\pi)$ (topology S^1).
- **Saturation** corresponds to the radial coordinate $r \in [0, R]$.
- **Brightness** corresponds to the vertical coordinate z .

The metric on this manifold determines distinguishability. In our framework, the “Hybrid Metric” (Eq. 1) explicitly models how neural invariants shape this experience.

Consider the role of **Differentiation** (Δ). We can model the radius of the color solid as a function of differentiation: $R(\Delta) \propto \log(1 + \Delta)$.

- **High Δ (Wakefulness):** R is large. The manifold volume is expanded, allowing distinct separation between “Red” and “Pink” (large geodesic distance).
- **Low Δ (Drowsiness/Pathology):** As $\Delta \rightarrow 0$, the radius $R \rightarrow 0$. The manifold collapses toward its central axis ($r = 0$). Topological cycles persist ($\beta_1 = 1$) but metric volume vanishes. Experientially, this corresponds to the “fading” of vividness or collapse into greyscale (achromatopsia), where hue distinctions become impossible despite preserved brightness processing.

Similarly, consider the role of **Coherence** (Γ):

- **High Γ (Normal binding):** The angular coordinate θ is well-defined. Colors are experienced as unified qualia.
- **Low Γ (Fragmentation):** The metric along the θ direction degenerates. The “loop” of hue becomes fragmented—phenomenologically corresponding to the “dissolution” of color boundaries reported in certain psychedelic or dissociative states.

This example illustrates that “losing consciousness” is not merely a light switch turning off, but a specific **geometric deformation**: the contraction of the phenomenal manifold’s volume and the eventual collapse of its topological invariants (the hue loop closing into a point).

3 Methodology: Alignment and Validation

3.1 Gromov-Wasserstein Alignment

To compare \mathcal{N} and Ψ without shared coordinates, we use the Gromov-Wasserstein distance [6, 7], which compares the *internal distance matrices* of two metric spaces.

The optimal transport plan γ^* minimizes the distortion:

$$\mathcal{GW}_2^2(\mathcal{N}, \Psi) = \inf_{\gamma \in \Pi} \iint |d_{\mathcal{N}}(x, x') - d_{\Psi}(y, y')|^2 d\gamma(x, y) d\gamma(x', y') \quad (2)$$

Proposition 3.1. *If $\mathcal{GW}(\mathcal{N}, \Psi) \approx 0$, the spaces are isometric. The support of γ^* identifies the specific neural subspace (the Neural Correlate of Consciousness, NCC) that maps structurally to experience.*

3.2 Topological Validation via Persistent Homology

To address the “Manifold Illusion” (where algorithms like UMAP find structure in noise [17]), we use persistent homology, a technique that has proven powerful for detecting intrinsic structure in neural data [10, 11, 13]. We compute **Betti numbers** (β_k) across filtration scales.

Criterion 3.1 (Topological Authenticity). *A manifold structure is genuine iff:*

1. *It exhibits persistent topological features (long lifespans in the persistence diagram).*
2. *The integral of its Persistence Landscape (λ) is significantly greater than that of phase-shuffled surrogates: $\int \lambda_{real} \gg \int \lambda_{surr}$.*

4 Computational Validation: Color Ring Model

We validated the framework using a synthetic dataset representing the hue dimension of color perception (as described in Section 2.4).

4.1 Setup

- **Ground Truth (Ψ):** A unit circle S^1 representing Hue.
- **Neural Model (\mathcal{N}):** 50 neurons with Gaussian tuning curves and Poisson noise ($\lambda = 0.1$).
- **Control:** Rate-matched shuffled spike trains.

4.2 Results

We computed the \mathcal{GW} distance and Persistence Diagrams (PD).

Table 1: **Validation Metrics (Toy Model).** Comparisons between the structured neural model and shuffled control (n=100 trials).

Metric	Neural Model	Control (Shuffle)	Stat. Sig.
β_1 Count (Loops)	1.0	0.0	—
Persistence (Norm.)	0.82 ± 0.05	0.12 ± 0.08	$p < 0.001$
Norm. GW Cost	0.042 ± 0.005	0.350 ± 0.020	$p < 0.001$

The framework successfully recovered the circular topology ($\beta_1 = 1$) from noisy data and aligned it with the ground truth, whereas the control failed. This validates the core claim: the GW-TDA pipeline can detect phenomenal geometry from neural activity.

5 Falsifiable Predictions

The PMH is distinct from other theories because it predicts specific *geometric* transformations.

Table 2: Differential Predictions of PMH vs. Major Theories

Scenario	PMH Prediction	IIT Prediction	GNWT Prediction
Loss of Consciousness (Anesthesia)	Topological Phase Transition: Abrupt collapse of β_1, β_2 features. Metric space becomes contractible.	Reduction in scalar Φ .	Failure of global broadcast (P300 wave collapse).
Cross-Modal Perception	Geometric Isometry: Auditory Pitch and Visual Hue manifolds align via GW ($GW \approx 0$) in association cortex.	Similar conceptual structure size.	Broadcast of distinct content codes.
Psychedelic State	Manifold Expansion: Increase in volume form and curvature K ; creation of novel topological holes (β_k).	Increase in Φ (or ambiguous).	Hyper-associativity / Global ignition stability issues.

5.1 Experiment 1: Topological Collapse in Anesthesia

Protocol: High-density EEG during propofol induction.

Prediction: Loss of consciousness will correlate precisely with the disappearance of persistent homological cycles ($\beta_1 \rightarrow 0$), distinct from mere spectral power changes.

Falsification: If behavioral LOC occurs *without* topological collapse, PMH is falsified.

5.2 Experiment 2: Adversarial Manifold Validation

Protocol: Compare real neural manifolds against phase-shuffled surrogates.

Falsification: If the ‘‘Manifold Illusion’’ holds, surrogates (which lack genuine topology) should yield GW scores indistinguishable from real data. PMH requires $GW_{\text{real}} \ll GW_{\text{surr}}$.

6 Discussion

6.1 Scope and Limitations

This work proposes a *formal interpretative framework* with falsifiable predictions, not a complete empirical model of consciousness. Several limitations require acknowledgment:

1. **Computational Complexity:** GW computation is NP-hard for large systems; approximations introduce estimation errors.
2. **Temporal Dynamics:** The current formalism treats states as static. Extension to Lorentzian manifolds is needed to capture temporal flow.
3. **Phenomenal Space Construction:** Ψ is derived from behavioral reports, which may not faithfully reflect phenomenal structure.

6.2 Relation to Existing Theories

The PMH is not a replacement for IIT or GNWT but a *complementary geometric formalization*. IIT provides the scalar Φ ; PMH provides the *shape* of the integrated information [21, 25]. GNWT describes access mechanisms; PMH describes the *geometry* of what is accessed. Recent work on information decomposition [24] and categorical approaches [23] suggests that geometric and topological methods may provide the unifying language for consciousness science.

7 Conclusion

The Phenomenal Manifold Hypothesis provides a mathematically robust, falsifiable alternative to purely functionalist or scalar theories of consciousness. By combining **Gromov-Wasserstein alignment** with **Topological Data Analysis**, we bridge the gap between the high-dimensional objective brain and low-dimensional subjective experience.

Consciousness, in this view, is not a “ghost in the machine,” but the machine’s intrinsic topological signature. The “Hard Problem” is thus reformulated as a tractable problem of metric space alignment.

Acknowledgments

The author thanks anonymous reviewers for constructive feedback on earlier versions of this manuscript.

Data and Code Availability

Simulation code for the Color Ring toy model is available from the author upon reasonable request. No empirical data were used in this study.

Conflict of Interest

The author declares no competing interests.

Author Contributions

É.R. conceived the theoretical framework, developed the mathematical formalism, performed computational validation, and wrote the manuscript.

References

- [1] Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.
- [2] Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nat. Rev. Neurosci.*, 17(7), 450–461.
- [3] Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.
- [4] Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS Comp. Bio.*, 10(5), e1003588.
- [5] Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, 23(7), 439–452.

- [6] Mémoli, F. (2011). Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4), 417–487.
- [7] Peyré, G., & Cuturi, M. (2019). Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6), 355–607.
- [8] Alvarez-Melis, D., & Jaakkola, T. S. (2018). Gromov-Wasserstein alignment of word embedding spaces. *Proc. EMNLP*, 1881–1890.
- [9] Otter, N., et al. (2017). A roadmap for the computation of persistent homology. *EPJ Data Science*, 6, 17.
- [10] Giusti, C., Pastalkova, E., Curto, C., & Itskov, V. (2015). Clique topology reveals intrinsic geometric structure in neural correlations. *PNAS*, 112(44), 13455–13460.
- [11] Petri, G., et al. (2014). Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101), 20140873.
- [12] Sizemore, A. E., et al. (2018). Cliques and cavities in the human connectome. *Journal of Computational Neuroscience*, 44(1), 115–145.
- [13] Curto, C. (2017). What can topology tell us about the neural code? *Bulletin of the AMS*, 54(1), 63–78.
- [14] Gallego, J. A., et al. (2017). Neural manifolds for the control of movement. *Neuron*, 94(5), 978–984.
- [15] Chaudhuri, R., Gercek, B., Bhanu, K., Fiete, I., & Bhalla, U. S. (2019). The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature Neuroscience*, 22(9), 1512–1520.
- [16] Jazayeri, M., & Ostojic, S. (2021). Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Current Opinion in Neurobiology*, 70, 113–120.
- [17] Chari, T., & Pachter, L. (2023). The specious art of single-cell genomics. *PLoS Comp. Bio.*, 19(8), e1011288.
- [18] Ladyman, J., & Ross, D. (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford Univ. Press.
- [19] French, S. (2014). *The Structure of the World: Metaphysics and Representation*. Oxford Univ. Press.
- [20] Worrall, J. (1989). Structural realism: The best of both worlds? *Dialectica*, 43(1-2), 99–124.
- [21] Balduzzi, D., & Tononi, G. (2009). Qualia: The geometry of integrated information. *PLoS Comp. Bio.*, 5(8), e1000462.
- [22] Fekete, T., van Leeuwen, C., & Edelman, S. (2016). System, subsystem, hive: Boundary problems in computational theories of consciousness. *Frontiers in Psychology*, 7, 1041.
- [23] Tsuchiya, N., Taguchi, S., & Saigo, H. (2016). Using category theory to assess the relationship between consciousness and integrated information theory. *Neuroscience Research*, 107, 1–7.
- [24] Mediano, P. A., et al. (2022). Greater than the parts: A review of the information decomposition approach to causal emergence. *Phil. Trans. R. Soc. A*, 380(2227), 20210246.
- [25] Tegmark, M. (2016). Improved measures of integrated information. *PLoS Comp. Bio.*, 12(11), e1005123.
- [26] Amari, S. (2016). *Information Geometry and Its Applications*. Springer.