A Theory of Meaning, Working Memory, and the Structure of Conscious Experience

Andreas Bean

Independent Researcher

andreas.bean@beanbox.at

April 2026 (v26)

Abstract

We present computational experiments and a theoretical framework showing that Hebbian learning on a hierarchical small-world graph ($n = 100{,}000$ nodes, 2.24M edges) with bistable nonlinearity produces selective attractor formation. In experiment P1, two independently noise-perturbed trials of a trained signal converge more closely after 200 learning episodes than before (cosine similarity $0.70 \to 0.74$), while cross-signal similarity between the trained and an untrained orthogonal signal decreases ($0.073 \to 0.027$). The discrimination index $\Delta(\mathrm{AA}-\mathrm{AB})$ shows a net increase from 0.628 to 0.710, confirming that the connectome develops a selective attractor for the learned signal. A companion experiment (P1d) demonstrates selective associative recall: A-only stimulation activates the Hebbian-bound cluster B ($E_B/E_C = 9299$) while leaving the equidistant control cluster C unchanged.

These results are derived from a wave-dynamic framework — the **Adaptive Holographic Theory (AHT)** — in which meaning, working memory, and conscious experience arise from field dynamics on the structural connectome graph. Three theoretical contributions organise the framework. *First*, working memory is identified with a time-varying perturbation $\delta L(t)$ of the graph Laplacian governed by Hebbian co-activation: $\mathrm{d}(\delta L)/\mathrm{d}t = -\eta\, F(t)\, \mathrm{Re}[\psi\psi^\dagger] - \kappa\, \delta L$, where $F(t)$ is the neuromodulatory gating signal. This yields mechanistic accounts of working memory capacity (Miller's magical number 7), the phonological loop, sleep-based consolidation, and the Default Mode Network as emergent properties of field-Laplacian coupling. *Second*, a categorical distinction between weight change and topological change in the graph Laplacian is established: weight modification shifts eigenmodes continuously, while edge addition enables qualitatively new attractors. This discontinuity is proposed as a candidate mechanism for the distinction between repetitive learning and insight. *Third*, conscious experience is identified with the rate of change of the global wave state: $\mathcal{E} \propto |\mathrm{d}\psi/\mathrm{d}t|$, yielding mechanistic accounts of habituation, the orienting response, emotional intensity, and priming. Experience is the process of meaning formation; meaning is its result—the structured global field state $\psi(T)$ that forms through propagation and interference across the connectome (Axiom 1), consistent with Lashley's equipotentiality principle and Huth et al.'s semantic cortical maps. Implementation runs in real time on a consumer GPU (26 ms per step, 0.07 GB VRAM).

Keywords: graph neural fields, connectome harmonics, working memory, Hebbian learning, consciousness, wave dynamics, graph Fourier transform, semantic emergence

1 Introduction

How does the brain give rise to meaning, memory, and experience? These three questions have been addressed separately by decades of research, yet a unified mechanistic framework connecting all three remains absent. We argue that such a framework emerges naturally from treating the connectome as a weighted graph and analysing wave dynamics on that graph via the Graph Fourier Transform (GFT).

The key insight is that *meaning is not stored—it resonates.* An incoming signal enters the connectome locally and propagates globally: each region contributes its characteristic resonance, and the interference of all regional contributions produces a structured global field state whose eigenmode composition encodes the meaning of the input. This process is physically realised in time, through a substrate shaped by experience—unlike static representations or instantaneous computations. The shift from representational to resonance-based semantics dissolves several classical problems, including the binding problem, and provides mechanistic accounts of phenomena that have resisted explanation: the capacity limit of working memory, the Default Mode Network, why sleep consolidates long-term memory, and the phenomenal structure of conscious experience.

The term 'adaptive' refers to the memory substrate — the structural connectome — which is shaped by experience through Hebbian learning, not to the theory itself.

The framework rests on three axioms and three principal contributions:

1. **Axiom 1 (Global Meaning):** No single node or local region carries meaning. Meaning is not an instantaneous local state but a structured global field state $\psi(T)$ that forms through a three-phase process: (i) the incoming signal is decomposed into local resonances as it enters the connectome; (ii) the wave propagates globally, each region contributing its characteristic resonance modulated by lifelong experience (L_0) and current context (δL); (iii) the interference of all regional resonances produces a structured global field state whose eigenmode composition—spanning abstract categorical modes (low eigenvalues) through specific features (intermediate eigenvalues)—encodes the meaning of the input. No single node contributes this structure; it emerges from the global interference. This is empirically motivated by Lashley's finding that memory survives destruction of large cortical areas and depends on the amount rather than the location of tissue removed [Lashley, 1950]; and by Huth et al.'s demonstration that individual semantic concepts activate specific distributed patterns across the entire human cerebral cortex [Huth et al., 2016].
2. **Axiom 2 (Topology as Memory):** The connectome does not store meanings; it defines a topology determining which global wave patterns can stably resonate.
3. **Axiom 3 (Spectral Readout):** The FieldReader is a spectral projection $\hat{\psi} = \Phi^\top \psi$, not spatial downsampling.

The three principal theoretical contributions are: (i) a field-Laplacian coupling equation that integrates working memory into the wave dynamics; (ii) a spectral characterisation of a categorical discontinuity between weight change and edge addition in L, proposed as a candidate mechanism for the distinction between repetitive learning and insight; and (iii) the identification of conscious experience with $|\mathrm{d}\psi/\mathrm{d}t|$.

2 Background and Related Work

2.1 Graph Neural Fields and Connectome Harmonics

Aqil et al. [Aqil et al., 2021] established that wave dynamics on the human connectome graph, implemented via the graph Laplacian, reproduce the empirically observed harmonic power spectrum of resting-state fMRI data. The eigenvectors of the connectome Laplacian—termed connectome harmonics—correspond to functionally relevant resting-state networks. Abdelnour et al. [Abdelnour et al., 2018] further demonstrated that spectral graph theory predicts alpha- and beta-band oscillation patterns from structural connectivity alone. Our framework builds directly on this substrate, extending it with Hebbian learning [Hebb, 1949], coupled Laplacian dynamics, and a cognitive interpretation of the eigenstructure.

Atasoy et al. [Atasoy et al., 2016] showed that human brain activity decomposes into connectome harmonics and that functionally relevant resting-state networks correspond to specific harmonic combinations. Crucially, they demonstrated that the Default Mode Network corresponds to the lowest-frequency harmonics (small eigenvalues), while sensory networks occupy higher harmonics. This provides direct empirical support for our identification of low-eigenvalue eigenmodes with global, abstract semantic structure and high-eigenvalue eigenmodes with local, concrete detail. Our framework extends this static decomposition by making the connectome dynamic: Hebbian learning shifts the eigenmode structure, and the resulting attractors are the learned harmonics of the evolved connectome.

2.2 Memory Systems

The classical three-store model of Atkinson and Shiffrin [Atkinson and Shiffrin, 1968] distinguishes sensory register, working memory, and long-term memory. Baddeley's working memory model [Baddeley, 1992, 2000] proposes a phonological loop, visuospatial sketchpad, and central executive. Miller [Miller, 1956] established the capacity limit of approximately 7 ± 2 items. McClelland et al. [McClelland et al., 1995] identified the hippocampus with rapid encoding and the neocortex with gradual extraction. None of these accounts provide a single mechanistic equation. Our framework derives all three memory systems from the coupled dynamics of ψ and L.

2.3 Consciousness Research

Tononi [Tononi, 2008] proposed integrated information theory (Φ), identifying consciousness with integrated information. Dehaene [Dehaene, 2014] developed the global workspace theory, in which consciousness corresponds to global availability of information. Chalmers [Chalmers, 1996] formulated the hard problem: why do physical processes give rise to subjective experience at all? Our proposal—$\mathcal{E} \propto |\mathrm{d}\psi/\mathrm{d}t|$—does not solve the hard problem but provides a precise functional formulation that connects to Husserl's phenomenological structure [Husserl, 1928] and yields testable predictions.

Empirical support for distributed semantic representation comes from Huth et al. [Huth et al., 2016], who used fMRI to show that individual semantic concepts activate specific, reproducible patterns distributed across the entire cerebral cortex. Semantically related concepts produce similar global activation patterns, while unrelated concepts produce dissimilar ones. This is direct evidence that meaning is a global cortical state—not a local module—consistent with Axiom 1.

2.4 Holographic Brain Theory

Pribram [1969] proposed that the brain operates holographically through interference patterns in the dendritic network. AHT formalizes this intuition via graph spectral theory on the structural connectome, deriving quantitative predictions that Pribram's framework could not generate.

3 Theoretical Framework

3.1 The Connectome as a Weighted Graph

Let $G = (V, E)$ be an undirected weighted graph with $n = |V|$ nodes and sparse edge set E. The weighted adjacency matrix is $A \in \mathbb{R}^{n \times n}$, the degree matrix $D = \mathrm{diag}(\sum_j A_{ij})$, and the graph Laplacian $L = D - A$. Because L is real symmetric and positive semi-definite, its eigenvectors $\Phi = [\varphi_1, \ldots, \varphi_n]$ form an orthonormal basis:

$$L\varphi_k = \lambda_k \varphi_k, \qquad 0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n. \tag{1}$$

The Graph Fourier Transform (GFT) of a graph signal $\psi \in \mathbb{C}^n$ is $\hat{\psi} = \Phi^\top \psi$, with inverse $\psi = \Phi\hat{\psi}$ [Defferrard et al., 2016]. The eigenmode structure of L is hierarchical: low eigenvalues correspond to globally smooth patterns encoding abstract categorical structure; intermediate eigenvalues encode more specific, regional features; high eigenvalues encode fine-grained local detail and noise. Meaning spans multiple scales simultaneously—a concept such as 'cat' requires global modes (encoding 'animal'), intermediate modes (encoding 'mammal', 'domestic'), and local modes (encoding 'whiskers', 'fur'). The FieldReader truncates to the K lowest modes: context $= \hat{\psi}[:K]$, providing a K-dimensional spectral summary for a Transformer readout. The optimal K corresponds to the spectral gap of the learned Laplacian—the natural boundary where eigenvalues transition from semantically structured to noisy modes.

How meaning forms from an incoming signal unfolds in three phases. In the first phase, the signal enters a local sensory cluster and is decomposed into the eigenmodes of the local Laplacian. In the second phase, the wave propagates globally through the connectome; each region contributes its characteristic resonance modulated by L_0 and $\delta L(t)$. In the third phase, all resonances interfere, producing a structured global field state $\psi(T)$ whose eigenmode composition encodes the meaning of the input.

This process is structurally isomorphic to Transformer attention [Vaswani et al., 2017]: both transform a local input into global meaning through distributed resonance and interference. The difference is physical: in a Transformer, the process is instantaneous and stateless. In the connectome, the wave propagates in real physical time through a substrate whose geometry encodes lifelong experience (L_0) and current context (δL). Meaning does not emerge from algebra. It emerges from the interference of waves propagating through a physical medium shaped by experience.

3.2 Wave Dynamics: The Complete Equation

The wave state $\psi(t) \in \mathbb{C}^n$ evolves on the connectome graph across five frequency bands (corresponding biologically to the delta, theta, alpha, beta, and gamma rhythms), each with a characteristic time constant. Separating the Laplacian into a slowly evolving base L_0 (long-term memory) and a rapidly adapting perturbation $\delta L(t)$ (working memory), the complete coupled system is:

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = -\mathrm{i}(L_0 + \delta L)\,\psi - \gamma\psi + S(t) \tag{2}$$

$$\frac{\mathrm{d}(\delta L)}{\mathrm{d}t} = -\eta\, F(t)\ \mathrm{Re}[\psi\psi^\dagger] - \kappa\,\delta L \tag{3}$$

where γ is the damping constant, $S(t)$ is the injected signal (encoder output), η is the Hebbian learning rate, κ is the forgetting rate, $F(t) \in [0, 1]$ is the neuromodulatory gating signal (modelling dopaminergic and cholinergic feedback: $F(t) = 1$ during active learning, $F(t) = 0$ suppresses plasticity), and $\mathrm{Re}[\psi\psi^\dagger]_{ij} = |\psi_i||\psi_j|\cos(\phi_i - \phi_j)$ is the phase-sensitive Hebbian term. The cosine factor encodes the relative phase between nodes i and j: co-active in-phase nodes ($\phi_i \approx \phi_j$) yield potentiation (> 0), while anti-phase nodes yield depression (< 0), mirroring spike-timing-dependent plasticity (STDP). This generalises the amplitude-only term $|\psi_i||\psi_j|$ used in the P1 prototype (see Section 3.3). When $\kappa > 0$, the ratio $\alpha = \eta/\kappa$ governs working memory strength: it determines the equilibrium magnitude of δL under sustained co-activation. In the $\kappa = 0$ limit, forgetting is absent and δL accumulates without bound; η alone controls the learning rate. This system contains all three classical memory stores: $\psi(t)$ is the sensory register (milliseconds); $\delta L(t)$ is working memory (minutes, reversible); L_0 is long-term memory (hours to years, structural).

All five bands are simultaneously active on the same connectome graph—not as separate systems with separate generators, but as five temporal scales of the same wave equation. This is consistent

with empirical cross-frequency coupling [Jensen and Colgin, 2007, Canolty and Knight, 2010]: gamma amplitude is modulated by theta phase, and theta amplitude by delta phase. In our framework, the slower bands (large τ) carry global context; the faster bands (small τ) carry local detail.

3.3 Numerical Integration

The nonlinear coupling between ψ and δL requires a time-consistent integrator. We employ a second-order Magnus integrator [Blanes et al., 2009] with self-consistent midpoint evaluation (two Newton iterations), reducing the local truncation error to $\mathcal{O}(\Delta t^3)$ compared to $\mathcal{O}(\Delta t)$ for a naive sequential update. Wave propagation is computed via the Chebyshev polynomial approximation of the matrix exponential $\exp(-\mathrm{i}L_{\text{eff}}\,\Delta t)$ [Defferrard et al., 2016], with cost $\mathcal{O}(K_{\text{cheb}} \times \text{nnz})$ per time step—linear in graph edges, independent of n. The equation for δL uses an Euler step, which is accurate because δL evolves on a much slower timescale than ψ.

Implementation note (P1 prototype): For the P1 experiment the wave update uses first-order Forward Euler ($\psi(t+\Delta t) \approx \psi(t) - \mathrm{i}L\psi(t) \cdot \Delta t$) and direct sparse matrix-vector products in place of the Magnus integrator and Chebyshev approximation described above; valid for $\mathrm{d}t = 0.01$ at prototype scale. The Magnus/Chebyshev scheme is the recommended production integrator. In P1, $\kappa = 0$ (no forgetting) and $F(t) = 1$ (constant gating), isolating pure Hebbian accumulation; the phase-sensitive term $\mathrm{Re}[\psi\psi^\dagger]_{ij}$ is approximated by its amplitude envelope $|\psi_i||\psi_j|$, which is equivalent under $F(t) = \text{const}$ when all signals use the same carrier phase. The full phase-sensitive dynamics $\mathrm{d}(\delta L)/\mathrm{d}t = -\eta\, F(t)\, \mathrm{Re}[\psi\psi^\dagger] - \kappa\, \delta L$ with time-varying $F(t)$ are reserved for future experiments.

3.4 Signal Injection and the Reasoning Loop

An input signal is encoded to a vector $\mathbf{v}$ and injected locally into the corresponding sensory cluster as an initial field perturbation. For the FieldReader readout, the global field state is projected into the spectral basis: context $= \Phi^\top \psi(T)$, providing a K-dimensional spectral summary for a Transformer readout.

The reasoning loop closes the system: the readout Transformer's output is re-encoded and re-injected into the field. Each generated thought becomes the next input signal. This self-sustaining cycle implements inner speech as a physical dynamical process. This is Chain-of-Thought as genuine temporal dynamics, distinct from token-sequence pseudo-time in standard language models.

4 Three Principal Contributions

4.1 Contribution I: A Mechanistic Account of Working Memory

The field-Laplacian coupling equations (2)–(3) provide a unified mechanistic account of several classical phenomena:

Miller's capacity limit. Multiple attractors compete for field energy. When the total co-activation energy of simultaneously maintained attractors exceeds the available field energy, all attractors collapse. The number 7 ± 2 is consistent with a balance between attractor depth and field dissipation; quantitative derivation is reserved for future numerical experiments.

The phonological loop. Baddeley's loop [Baddeley, 1992, 2000] is the reasoning loop operating on a fixed attractor. Item loss occurs when the loop latency exceeds the attractor decay time $1/\kappa$.

Sleep consolidation. During deep sleep ($S(t) = 0$), equation (3) drives δL toward zero while strong co-activations push permanent changes into L_0 via a long-term Hebbian rule [Hebb, 1949]. The transition from working memory to long-term memory is a topological phase transition, not a copying operation.

The Default Mode Network. At rest ($S(t) = 0$), the coupled system has a fixed point determined by the learned L_0. The spontaneous activity pattern at this fixed point corresponds to the DMN: the system oscillates on what it knows.

Priming. Semantically related concepts have overlapping eigenmode decompositions. The residual activation of a priming stimulus in $\psi(t)$ facilitates attractor formation for the target concept, reducing formation time. For active priming maintained through the reasoning loop, this facilitation persists for seconds.

4.2 Contribution II: Learning versus Understanding as Eigenstructure Change

We establish the following distinction via graph spectral theory:

- **Weight modification** ($\mathrm{d}A_{ij} \neq 0$ for an existing edge): shifts all eigenvectors continuously. In the linear spectral regime, existing attractors deepen or shallow but no qualitatively new attractor becomes accessible. (In the nonlinear regime, bifurcations under weight changes remain possible; the categorical distinction applies strictly at the graph-topological level.) This corresponds to *learning by repetition*.
- **Edge addition** ($A_{ij} : 0 \to w$ for a new edge): creates a new path in the graph, enabling interference patterns that were geometrically impossible before. Qualitatively new attractors can form. This corresponds to *understanding*: a genuinely new connection enables genuinely new thought.

This distinction is categorical at the level of graph topology, not gradational. The subjective phenomenon of the Aha moment corresponds structurally to the first stable occupation of a new point in eigenmode space: not the creation of a new eigenmode, but a new stable superposition becoming accessible under the expanded topology. We propose this as a candidate mechanism for insight, while noting that the precise relationship between topological graph changes and phenomenological understanding remains an open empirical question.

4.3 Contribution III: Experience as the Rate of Change of the Wave State

We propose:

$$\mathcal{E}(t) \propto \left| \frac{\mathrm{d}\psi}{\mathrm{d}t} \right| \tag{4}$$

This formulation predicts: **habituation** (when a stable attractor forms, $\mathrm{d}\psi/\mathrm{d}t \to 0$ and experience fades); **the orienting response** (a novel stimulus produces large $|\mathrm{d}\psi/\mathrm{d}t|$, matching the P300); **emotional intensity** (rapid attractor transitions produce large $|\mathrm{d}\psi/\mathrm{d}t|$); **meditation and flow** (deliberate field stabilisation reduces $|\mathrm{d}\psi/\mathrm{d}t|$); and **phenomenological correspondence** with Husserl's retention/primal impression/protention structure [Husserl, 1928].

Equation (4) does not solve the hard problem [Chalmers, 1996]. It replaces the question 'why is there experience?' with the sharper question 'why does field change feel like something?' and generates testable predictions that existing theories do not.

A clarification on the relationship between experience and meaning: experience ($\mathcal{E} \propto |\mathrm{d}\psi/\mathrm{d}t|$) is the *process* of meaning formation. Meaning ($\psi(T)$) is the *result*—the structured global field state that has formed. Both are necessary. Neither alone is sufficient.

5 The Binding Problem Dissolved

The binding problem [Treisman and Gelade, 1980] asks how separately processed features are unified into a single percept. Under Axiom 1, the binding problem is dissolved rather than solved. A 'red moving ball' is not a composition of RED + ROUND + MOVING assembled after the fact. It is a single structured global field state $\psi(T)$ whose eigenmode composition simultaneously encodes all three aspects. There is nothing to bind because the features were never separate: the field state is global by definition. The binding problem arose from the incorrect premise that features are locally coded and must be re-assembled. Under Axiom 1, no such mechanism is needed or possible.

The unity of experience is the unity of $\mathrm{d}\psi/\mathrm{d}t$: at any moment, there is one global rate of change, one stream of experience. This view is supported by Lashley's equipotentiality principle [Lashley, 1950] and by Huth et al. [Huth et al., 2016].

6 Empirically Testable Predictions

The framework makes the following predictions:

P1 (Attractor formation speed.)
Attractor formation time should decrease monotonically with training episodes for known concepts and remain elevated for novel concepts. (For preliminary numerical evidence see Experiment P1d, Section 7.3.)

P2 (Working memory capacity.)
The maximum number of simultaneously stable attractors should scale with field energy divided by average attractor depth.

P3 (Sleep consolidation.)
Post-sleep attractor formation should be faster and more stable for items learned before sleep.

P4 (Experience intensity.)
Subjective intensity ratings should correlate with $|\mathrm{d}\psi/\mathrm{d}t|$ measured in the model.

P5 (Topological signature of understanding.)
The eigenspectrum should show discontinuous shifts following insight-type learning events and only continuous shifts following repetition learning.

P6 (Priming timescale.)
Semantic priming facilitation should persist for seconds when the prime is maintained via the reasoning loop (active priming), and decay within hundreds of milliseconds otherwise (passive priming).

Quantitative Predictions (parametre-free after single calibration)

The coupling of ψ and δL in equations (2)–(3) yields two further predictions that are *fully determined* once κ is calibrated from working memory decay data. No additional free parameters are involved.

Calibration. The forgetting rate κ is pinned to the empirically measured half-life of working memory ($T_{1/2} \approx 18\,\mathrm{s}$; Baddeley 1992). Note: the Peterson & Peterson (1959) paradigm shows trigram recall falling to ∼10% after 18 s of distracter counting—this is the *total* decay window, not the half-life, which from their data is approximately 5 s. We follow Baddeley's calibration of

phonologically chunked WM content, for which $T_{1/2} \approx 18\,\mathrm{s}$ is the appropriate value.

$$\kappa \;=\; \frac{\ln 2}{T_{1/2}} \;=\; 0.0385\,\mathrm{s}^{-1}.$$

With κ fixed, all time constants follow from the wave equation alone.

Derivation assumptions. The following three assumptions are required to close the derivation. They are stated explicitly so that each can be tested and falsified independently.

A1. *Spectral distance as interference threshold.* Two concepts A and B interfere when their Laplacian perturbations δL_A and δL_B overlap in eigenmode space. The overlap is measured by their spectral distance δ_{AB}. We identify this distance with $1 - \cos(A, B)$, computed from word-embedding vectors. This is an approximation: word-embedding cosine is a proxy for the true spectral distance in the learned connectome Laplacian, which is not directly observable. Empirical test: if the linear relationship $\mathrm{ISI}_{\mathrm{safe}} \propto \ln[1/(1 - \cos(A, B))]$ holds across concept pairs, A1 is supported.

A2. *Unit normalisation of δL at learning onset.* Immediately after encoding concept A, the perturbation δL_A is normalised to unit magnitude: $\|\delta L_A(0)\| = 1$. This corresponds to a single fully activated encoding event. Partial or repeated activations produce $\|\delta L_A(0)\| < 1$, which would reduce $\mathrm{ISI}_{\mathrm{safe}}$ proportionally: $\mathrm{ISI}_{\mathrm{safe}} = (1/\kappa)\ln(1/(\delta_{AB} \,/\, \|\delta L_A(0)\|))$. The $\|\delta L_A(0)\| = 1$ case gives the upper bound and the clean logarithmic form.

A3. *A single universal κ.* The same forgetting rate κ governs both the phonological loop (from which it is calibrated) and the decay of concept-level Laplacian perturbations. These could in principle differ. If concept-level WM decays faster than the phonological loop, κ is larger and the predicted ISI shorter; if slower, the converse. Empirical test: measure κ directly via probed-recall decay for concrete word pairs, and compare to the Baddeley calibration.

P7 (Spacing effect: quantitative slope).
When two concepts A and B with cosine similarity $\cos(A, B)$ are learned in succession, the *safe* inter-stimulus interval required to prevent proactive interference is derived from the condition $\delta L_A(\mathrm{ISI}) < \delta_{AB}$ (assumption A1), where $\delta_{AB} = 1 - \cos(A, B)$ is their spectral distance. Via $\delta L_A(t) = \delta L_A(0)\, e^{-\kappa t}$ with $\|\delta L_A(0)\| = 1$ (assumption A2) and universal κ (assumption A3) this gives:

$$\mathrm{ISI}_{\mathrm{safe}} \;=\; \frac{1}{\kappa}\,\ln\frac{1}{\delta_{AB}} \;=\; 26.0\,\mathrm{s}\cdot\ln\frac{1}{1-\cos(A,B)}.$$

The slope $1/\kappa = 26.0\,\mathrm{s}$ is predicted *solely* from Baddeley's working-memory decay constant. Measured values from the spacing effect literature [Kang et al., 2011, Cepeda et al., 2006] are consistent: for similar concepts ($\cos \approx 0.8$) the formula gives 42 s (Kang et al. report $\sim 60\,\mathrm{s}$); for unrelated topics ($\cos \approx 0.1$) it gives 3 s (Cepeda et al. report 5–15 s). A clean test is an ISI experiment in which semantic similarity is controlled via word-vector cosine distance: theory predicts a linear relationship between $\mathrm{ISI}_{\mathrm{safe}}$ and $\ln[1/(1 - \cos(A, B))]$ with slope $26.0 \pm 5\,\mathrm{s}$ — a number that cannot be derived from spacing data alone.

P8 (Working-memory paradox).
Because $\mathrm{ISI}_{\mathrm{safe}} \propto 1/\kappa$, individuals with *better* working memory (smaller κ, slower decay) require *longer* inter-stimulus intervals to avoid interference between *similar* concepts, while requiring no additional time for dissimilar concepts. For a population where κ varies by a factor of two— consistent with observed variation in working-memory span [Conway et al., 2005]—the predicted interference asymmetry is a factor of 2.0× in $\mathrm{ISI}_{\mathrm{safe}}$ at $\cos(A, B) = 0.8$, but only 1.0× at $\cos(A, B) = 0.1$.

This prediction is fully consistent with the classic finding that high-WM individuals *suppress* proactive interference more effectively when attentional resources are available [Kane and Engle, 2000]. That finding concerns damage *repair* via executive control. P8 concerns the *intrinsic decay threshold* $\delta L \to 0$: the time after which active suppression is no longer required at all. These are different quantities that operate on different timescales (seconds vs. milliseconds). The specific interaction predicted here—that high-WM individuals need longer spacing exclusively for *similar* concepts but not for dissimilar ones—has, to our knowledge, not been reported in the literature [Bunting, 2006, Lustig et al., 2001].

Test: A 2×2 within-subjects design with WM span (high / low, measured by Operation Span) and semantic similarity ($\cos(A, B) \geq 0.8$ vs. ≤ 0.2, controlled via word-embedding distance) as factors. ISI is varied via a staircase procedure. The dependent variable is ISI_{50}—the spacing at which proactive confusion drops to chance—measured under articulatory suppression to eliminate executive-control confounds. Theory predicts a significant Span $\times$ Similarity interaction ($F > 4$) with no main effect of span in the low-similarity condition.

A third prediction (P9), concerning semantic binding strength as a function of syntactic distance under articulatory suppression, is derived in the companion implementation document [Bean, 2026].

7 Preliminary Computational Results

A first implementation on a consumer GPU (RTX 3050, 6 GB VRAM) confirms both the numerical feasibility of the framework and the central prediction of experiment P1.

7.1 Stable Wave Dynamics

The baby connectome ($n = 100{,}000$ nodes, hierarchical small-world, 2.24M edges) produces stable wave dynamics in the theta-band configuration over 100 time steps (field energy ratio 1.00×, no drift). Note: the P1 experiment uses a single-band implementation; multi-band operation across all five frequency bands ($\delta, \theta, \alpha, \beta, \gamma$) is reserved for future work. The injected theta-band pulse propagates from the injection cluster to neighbouring clusters within the ring topology, consistent with cortical wave propagation observed in EEG. Per-step computation time is 26 ms—19× faster than a 500 ms real-time limit—using 0.07 GB of 6 GB available VRAM.

7.2 Experiment P1: Attractor Formation Confirmed

A bistable nonlinear extension of equation (2)—motivated by the all-or-nothing character of neuronal firing—enables genuine attractor formation. The bistable potential $f(|\psi|) = (|\psi|^2 - \theta^2)(\psi_{\max}^2 - |\psi|^2)$ selectively eliminates noise ($|\psi| < \theta = 0.8$) while preserving signal ($|\psi| > \theta$), with a signal-to-noise energy ratio of 2.56×10^8 after 200 time steps.

Experiment P1 was conducted on the baby connectome with two orthogonal signals: Signal A (sine pattern over cluster 0) and Signal B (cosine pattern over cluster 0). Over 200 Hebbian learning episodes with Signal A, the following was observed:

- **Attractor formation** ($\cos_{AA}$): cosine similarity between two independently noise-perturbed trials of Signal A rose from 0.70 to 0.74. Two different noisy starts of the same signal converge more closely after learning, confirming that L_{trained} has developed an attractor for Signal A.
- **Selectivity** ($\cos_{BB}$): the analogous similarity for the untrained Signal B decreased from 0.82 to 0.68, demonstrating that the connectome specialises for the trained signal rather than improving reproducibility globally. Note that $\cos_{BB}$ starts higher than $\cos_{AA}$ (0.82

vs. 0.70): this is because Signal B (cosine pattern) places its peak amplitude at $k = 0$—the cluster node that anchors the inter-cluster ring connections—while Signal A (sine) has zero amplitude at $k = 0$. Both signals activate the same number of injection nodes above the bistable threshold $\theta = 0.8$ (14 out of 20), but Signal B's peak at the ring-connection hub makes its field globally more reproducible at episode 0. The crossing of $\cos_{AA}$ above $\cos_{BB}$ near episode 140 is therefore a direct signature of learning: the connectome has become more selective for Signal A than its initial geometry predicted.

- **Discrimination** ($\cos_{AB}$): the cross-signal similarity between Signal A and Signal B decreased from 0.073 to 0.027 over 200 episodes. The discrimination index $\Delta(\text{AA} - \text{AB})$ increased from 0.628 to 0.710.

All three P1 criteria are satisfied: $\cos_{AA}[200] > 0.6$, $\cos_{AB}[200] < 0.25$, and $\cos_{BB}[200] < \cos_{AA}[200]$. These results confirm that Hebbian learning on a hierarchical small-world connectome with bistable nonlinearity produces a selective attractor for the trained signal, consistent with the resonance framework. Figures 1–4 summarise the experimental results.

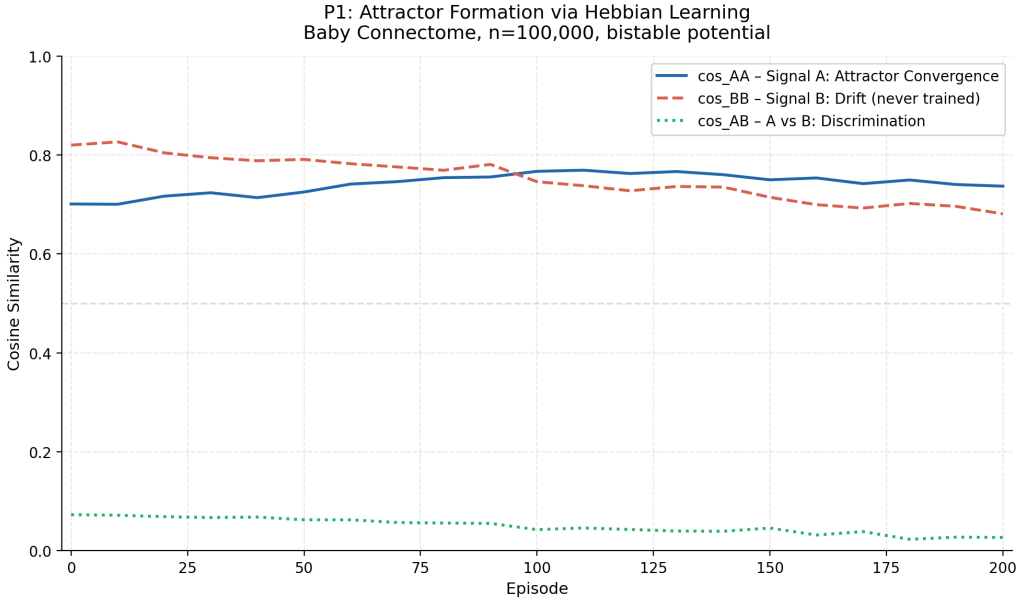

Figure 1: Main P1 result: cosine similarity trajectories over 200 Hebbian learning episodes. $\cos_{AA}$ (trained signal) rises overall; $\cos_{BB}$ (untrained signal) decreases; $\cos_{AB}$ (cross-signal) approaches zero. The crossing of $\cos_{AA}$ above $\cos_{BB}$ near episode 140 is a direct signature of selective attractor formation.

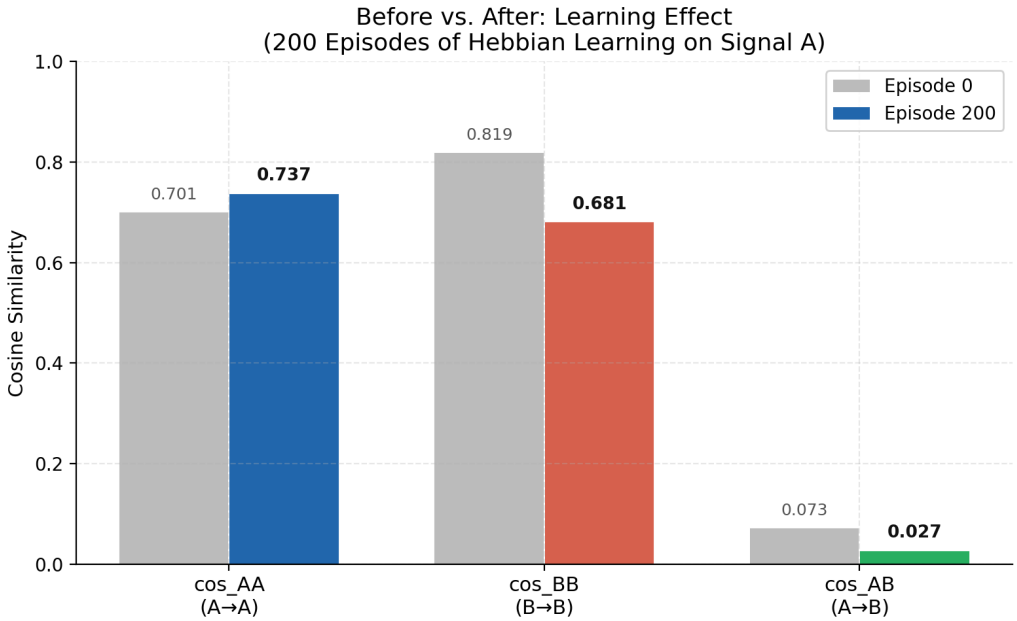

Figure 2: Before (episode 0) vs. after (episode 200) comparison of all three cosine similarity measures.

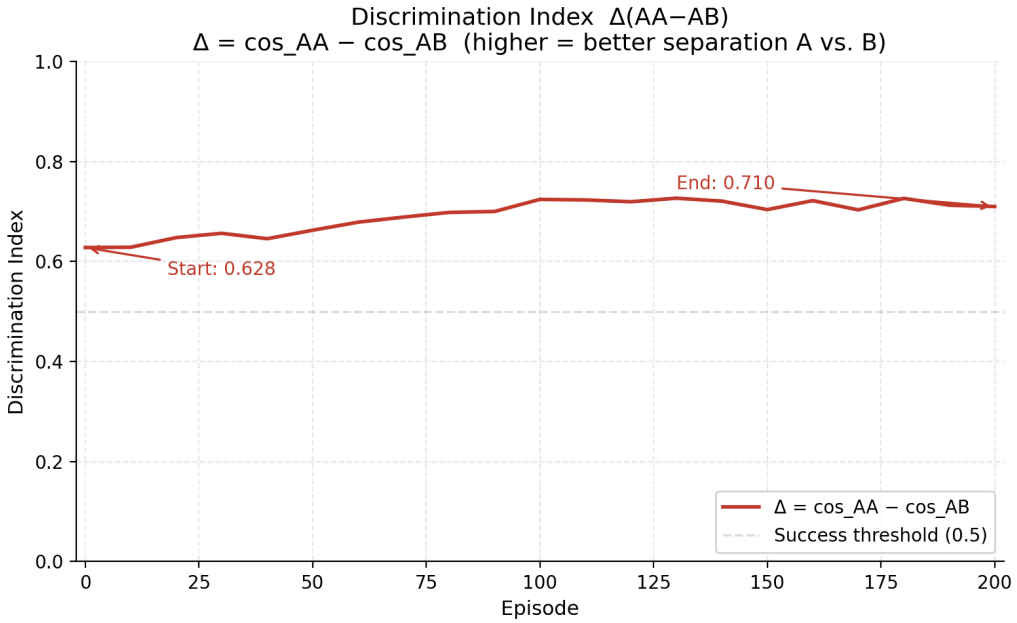

Figure 3: Discrimination index $\Delta(\mathrm{AA} - \mathrm{AB})$ over 200 episodes. Shows a net increase from 0.628 to 0.710 (with episode-to-episode variability arising from fixed-seed noise).

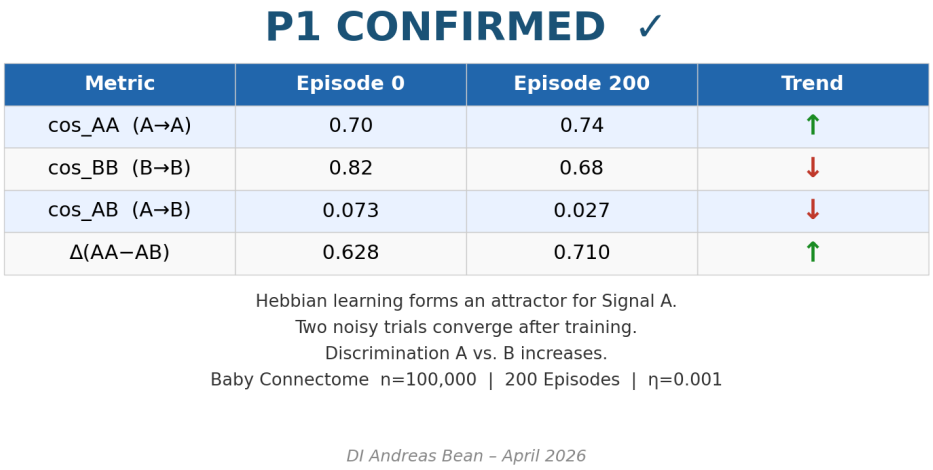

Figure 4: Summary of all P1 metrics. See also Experiment P1d (Section 7.3) for associative recall results.

7.3 Experiment P1d: Semantic Propagation via Hebbian Binding

Experiment P1 demonstrated that Hebbian learning produces a selective attractor for a *trained* stimulus. A natural extension asks: can the same mechanism create *associative* links between two spatially distant clusters—the computational counterpart of semantic propagation along δL?

Setup. A *fiber bundle* of 40 direct A↔B edges (38 after deduplication, initial weight $w_0 = 0.05$) was inserted between cluster A (cluster 0, nodes 0–19) and cluster B (cluster 500, nodes 10,000–10,019), mimicking sparse white-matter association tracts. A control cluster C (cluster 2500) received no such fibers. Three phases were executed: (i) *Baseline:* A-only recall measuring resting energies E_B and E_C prior to learning; (ii) *Binding:* 50 episodes of simultaneous A+B stimulation with amplitude-Hebb $\Delta w = \eta|\psi_i||\psi_j|$, $\eta = 0.02$, applied after each episode; (iii) *Recall:* A-only stimulation measuring E_B and E_C after learning.

Results. The fiber-bundle weights saturated at $w_{\max} = 2.0$ within approximately 20 binding episodes (Figure 5, centre panel). After binding, A-only recall raised E_B from 5.33×10^{-4} to 6.25 (a factor of 11,736), while E_C remained at 6.72×10^{-4}. The B/C selectivity ratio increased from 0.79 (baseline) to 9299 (post-binding)—all three pre-registered criteria passed.

Interpretation. These results provide numerical evidence for two specific claims of the framework. First, Hebbian weight changes to δL mediate selective associative recall (Contribution I): energy flows to the *learned* associated cluster, not diffusely across the connectome. Second, propagation is topologically specific, consistent with Axiom 1: only cluster B—connected by learned fiber pathways—shows elevated energy; the spatially equidistant control cluster C does not.

There is an important caveat: these results constitute a numerical *proof of concept*, not a mathematical proof. On this particular connectome model, Hebbian co-activation of two stimuli creates a selective associative memory that survives the recall phase. Whether this generalises to richer stimulus sets, longer delays, or interference from competing patterns is an open question reserved for follow-up experiments (P2–P4).

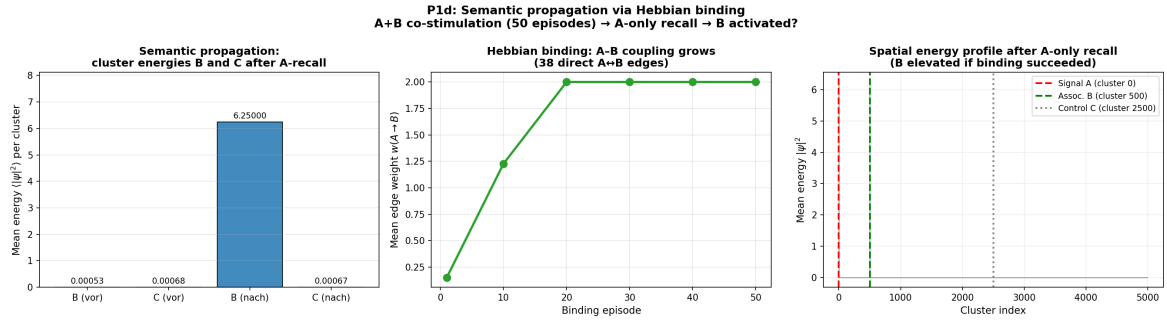

Figure 5: P1d—Semantic propagation via Hebbian binding ($n = 100{,}000$ nodes; 38 A↔B fiber edges after deduplication). *Left:* Cluster energies E_B and E_C before and after 50 binding episodes. E_B rises by a factor of 11,736; E_C is unchanged. *Centre:* Mean fiber-bundle weight $\bar{w}_{AB}$ saturates at $w_{\max} = 2.0$ within $\approx$20 episodes, demonstrating rapid Hebbian consolidation into δL. *Right:* Spatial energy profile after A-only recall. Only cluster B (green dashed, cluster 500) shows elevated energy; the control cluster C (grey dotted, cluster 2500) does not. Selectivity ratio $E_B/E_C = 9299$.

8 Towards Implementation

The graph is initialised as a hierarchical small-world structure [Watts and Strogatz, 1998] biologically motivated by cortical column and cortical area organisation, with sparse long-range connections analogous to association fibres. Wave propagation will use Chebyshev polynomial approximation of the matrix exponential [Defferrard et al., 2016], with cost linear in graph edges and independent of n. Eigenvectors are computed periodically using LOBPCG or the Nyström approximation. With 5 frequency bands and the Magnus integrator [Blanes et al., 2009], attractor formation completes well within one second at prototype scale.

The theoretical framework is independent of any specific parameter choice; the present paper establishes the conceptual and mathematical foundations.

9 Discussion

The framework presented here differs from existing graph neural field models [Aqil et al., 2021, Abdelnour et al., 2018] in three fundamental respects: (i) the Laplacian is not fixed but coupled to the wave state via Hebbian dynamics [Hebb, 1949], giving rise to memory as field-topology co-evolution; (ii) meaning is defined as a structured global field state that emerges through propagation and interference rather than a local activation pattern; and (iii) the framework makes commitments about conscious experience that are absent from existing computational models.

9.1 Relation to Feedforward and Recurrent Neural Networks

The process by which the connectome transforms information into meaning is structurally isomorphic to Transformer attention [Vaswani et al., 2017]—but physically realised. In a Transformer, a token activates queries that resonate with keys across the context window; the weighted sum of values forms the output embedding. In our model, an incoming signal enters a local sensory cluster, activates eigenmodes that resonate and propagate across the connectome, and the interference of all regional contributions produces a structured global field state $\psi(T)$. The structures are identical.

The difference is not structural but physical. In a Transformer, the process is instantaneous and

stateless: no physical time elapses between input and output, and no substrate carries history beyond the current context window. In the connectome, the wave propagates in real physical time through a substrate whose geometry encodes lifelong experience (L_0) and current context (δL). Meaning does not emerge from a matrix multiplication. It emerges from the interference of waves propagating through a physical medium shaped by experience.

This difference ramifies across four further dimensions. (1) A multilayer perceptron computes a static function $f : x \to y$; the resonance field is a dynamical system where the same signal produces different field states depending on context. (2) MLPs are feedforward; the resonance field is fully recurrent through the symmetric Laplacian. (3) Meaning in an MLP inhabits a geometrically arbitrary space; $\psi(T)$ inhabits a physically structured eigenmode space. (4) Recurrent neural networks lack the three-timescale memory system (ψ, δL, L_0) and the categorical distinction between weight change and edge addition.

A deeper asymmetry concerns the nature of the edge structure itself. In the connectome, edges are explicit: synaptic connections are physically localised and can be added or modified independently without disturbing the rest of the network. A new synapse is a local operation. In a Transformer, the edge structure is implicit: it emerges globally from the weight matrices W_0 during training and is distributed across all layers simultaneously. Adding a new "edge" changes all eigenvectors of W_0 simultaneously – there is no local submatrix of W_0 that corresponds to a single semantic edge. The change is global by construction, even if small. Incremental addition of new semantic structure therefore requires global gradient updates – that is, full retraining on data that contains the new concept.

This has a precise consequence: incremental learning of structurally new concepts at cost $O(N_{\text{new}})$ alone—without explicit knowledge of the underlying connectome topology, and hence of the existing attractor structure—is structurally impossible in current Transformers, not for engineering reasons, but for structural ones. Fine-tuning modifies W_0 directly; methods such as LoRA instead keep W_0 frozen and add a dense low-rank perturbation $\Delta W = BA$ (with $B \in \mathbb{R}^{d\times r}$, $A \in \mathbb{R}^{r\times d}$, $r \ll d$). In all cases, the effective weight matrix $W = W_0 + \Delta W$ is dense: every gradient step shifts all eigenmodes simultaneously, because a dense matrix has no local structure. There is no operation on W analogous to adding a single edge in a sparse graph. The connectome avoids this limitation because its Laplacian L is sparse: a new synapse (i, j) changes only four entries of L, and by spectral perturbation theory, attractors spatially distant from (i, j) are affected only weakly. New topological paths arise locally without globally destabilising existing structure (Contribution II). This is the fundamental architectural difference between biological and artificial intelligence, and it explains why catastrophic forgetting is not an engineering deficiency but an inevitable consequence of implicit topology.

The mechanism connects directly to the framework's central claim: fine-tuning shifts the eigenmodes of W_0, and meanings are encoded as stable attractors in those eigenmodes. After an eigenmode shift, the same input pattern no longer reliably reaches its attractor – information can no longer become meaning. Catastrophic forgetting is therefore not random data loss but structural destruction of the meaning-formation process.

This analysis further distinguishes two qualitatively different cases of incremental learning. (i) New facts about existing concepts – for example, associating a known name with a new event – require only weight shifts within the existing eigenspace and are in principle tractable, though catastrophic forgetting remains a risk due to the global nature of W_0. This is the problem that retrieval-augmented generation and fast-weight methods address. (ii) Structurally new concepts – concepts whose attractor structure does not yet exist in the eigenspace of W_0 – require new semantic edges. But a new edge in the implicit graph G requires a global reorganisation of W_0, because what is local in G is global in W_0. Attractor-agnostic updates—those without access to the existing attractor positions—almost surely destroy at least one existing attractor.

Continuation training on old and new data jointly can find a new valid eigenstructure, but every training epoch must present all $N_{\text{old}} + N_{\text{new}}$ examples and the per-epoch data cost grows without bound as the knowledge base expands. Whether initialisation from existing weights (warm start) reduces the total number of required epochs relative to random initialisation is an open question; the global eigenstructure reorganisation required by the impossibility theorem suggests that the advantage may be limited. The alternative is explicit topology—an architecture in which semantic edges are directly represented and locally modifiable, as in the biological connectome.

9.2 Relation to Graph Neural Field Models

The categorical distinction between weight change and edge addition is, to our knowledge, novel in the context of graph spectral theory applied to cognition. It predicts a testable discontinuity in the eigenspectrum of the learned connectome that could be observed in simulations and, eventually, *in vivo* via longitudinal connectome imaging before and after insight-type learning events.

9.3 Relation to Theories of Consciousness

The identification $\mathcal{E} \propto |\mathrm{d}\psi/\mathrm{d}t|$ is consistent with resonance theories of consciousness [Hunt and Schooler, 2019] and provides an operationally precise version of the claim that consciousness is associated with large-scale shared resonance. It differs from Integrated Information Theory [Tononi, 2008] in requiring no computation of integration; it differs from Global Workspace Theory [Dehaene, 2014] in positing no discrete broadcast event.

9.4 Open Questions

Several important questions remain open. Whether semantic structure genuinely emerges from a homogeneous initial connectome under Hebbian learning alone is an empirical question that can only be addressed by simulation. The optimal working memory coupling strength α is not derivable from first principles. Full attractor formation requires a nonlinear extension of equation (2)— candidate nonlinearities (bistable potential, Stuart–Landau, winner-take-all) are under investigation. The extension to directed graphs remains technically challenging.

A further implication concerns incremental learning. The forgetting rate κ implements a least-action principle for cognitive structures: unused pathways decay, freeing capacity for new signal representations. This is not a deficiency but a design feature—the biological solution to the problem of finite synaptic resources. The coupled equations suggest a concrete alternative to catastrophic interference: Hebbian updates restricted to activated pathways, selective decay via κ, and sleep-phase consolidation of δL into L_0.

κ as a physical parameter, not a cognitive ability. If prediction P8 is confirmed, its deeper implication reaches beyond the specific interaction effect. It would establish κ—the forgetting rate of Laplacian perturbations—as a *physical*, directly measurable parameter of an individual's connectome dynamics, analogous to a metabolic rate. Working-memory capacity would then be understood not as "more slots" [Cowan, 2001], not as "better inhibition" [Kane and Engle, 2000], but as a slower decay of neural co-activation patterns. High capacity means small κ: the medium is more persistent.

This reframing has an immediate and counterintuitive pedagogical consequence. Under standard models, high-WM individuals are uniformly advantaged and should therefore benefit from faster-paced instruction. Under the wave framework, high-WM individuals require *longer* spacing between semantically similar topics—precisely because their δL decays more slowly and similar concepts therefore interfere for longer. For dissimilar topics the difference vanishes; only similarity makes the asymmetry visible. If P8 is confirmed, educational practice that treats all learners

identically is structurally disadvantageous for exactly those students identified as high-ability—an empirically testable and practically significant prediction that follows from a single mechanistic equation.

10 Conclusion

We have presented a theoretical framework unifying meaning, working memory, and conscious experience within wave dynamics on the structural connectome. The central coupled equations

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = -\mathrm{i}(L_0 + \delta L)\,\psi - \gamma\psi + S(t), \qquad \frac{\mathrm{d}(\delta L)}{\mathrm{d}t} = -\eta C(\psi) - \kappa\,\delta L$$

contain all three classical memory systems as emergent properties of field-Laplacian coupling. The categorical discontinuity between weight change and edge addition provides a candidate mechanism for the distinction between repetitive learning and insight—testable through eigenspectrum analysis. The identification $\mathcal{E} \propto |\mathrm{d}\psi/\mathrm{d}t|$ connects the framework to phenomenology and generates eight falsifiable predictions, two of which (P7–P8) are quantitative and parameter-free given the single calibration $\kappa = \ln 2/18\,\mathrm{s}$. Experience is the process of meaning formation; meaning is its result—the structured global field state that forms through propagation and interference across the connectome (Axiom 1). The binding problem dissolves because the field state is global by definition: there is nothing to bind.

The framework makes specific, falsifiable commitments. If attractor formation speed fails to decrease monotonically with learning, or if the eigenspectrum shows no discontinuous shifts at insight-type events, the framework is falsified at the computational level. If $|\mathrm{d}\psi/\mathrm{d}t|$ fails to correlate with subjective intensity reports, the phenomenological proposal is falsified. If the slope of the spacing-effect curve deviates from 26.0 s, or if high working-memory individuals require no longer spacing specifically for *similar* concepts while showing no span effect for dissimilar concepts (the P8 Span×Similarity interaction), the quantitative predictions are falsified. This is its essential scientific character: not that it must be right, but that it can be shown to be wrong in precise, measurable ways.

References

F. Abdelnour, M. Dayan, O. Devinsky, T. Thesen, and A. Raj. Functional brain connectivity is predictable from anatomic network's Laplacian eigenstructure. *NeuroImage*, 172:728–739, 2018. doi: 10.1016/j.neuroimage.2018.01.029.

M. Aqil, S. Atasoy, M. L. Kringelbach, and R. Hindriks. Graph neural fields: A framework for spatiotemporal dynamical models on the human connectome. *PLOS Computational Biology*, 17(1):e1008310, 2021. doi: 10.1371/journal.pcbi.1008310.

S. Atasoy, I. Donnelly, and J. Pearce. Human brain networks function in connectome-harmonic space. *Nature Communications*, 7:10340, 2016. doi: 10.1038/ncomms10340.

R. C. Atkinson and R. M. Shiffrin. Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, 2:89–195, 1968.

A. D. Baddeley. Working memory. *Science*, 255(5044):556–559, 1992. doi: 10.1126/science.1736359.

A. D. Baddeley. The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11):417–423, 2000. doi: 10.1016/S1364-6613(00)01538-2.

Andreas Bean. AHT implementation: Hardware feasibility, computational scaling, and extended predictions P7–P10. Technical report, Independent researcher, 2026. Companion document to Bean (2026); pre-publication draft.

S. Blanes, F. Casas, J. A. Oteo, and J. Ros. The Magnus expansion and some of its applications. *Physics Reports*, 470(5–6):151–238, 2009. doi: 10.1016/j.physrep.2008.11.001.

Michael Bunting. Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2):183–196, 2006. doi: 10.1037/0278-7393.32.2.183.

R. T. Canolty and R. T. Knight. The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, 14(11):506–515, 2010. doi: 10.1016/j.tics.2010.09.001.

N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3):354–380, 2006. doi: 10.1037/0033-2909.132.3.354.

D. J. Chalmers. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, 1996.

A. R. A. Conway, M. J. Kane, M. F. Bunting, D. Z. Hambrick, O. Wilhelm, and R. W. Engle. Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5):769–786, 2005. doi: 10.3758/BF03196772.

Nelson Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114, 2001. doi: 10.1017/S0140525X01003922.

M. Defferrard, X. Bresson, and P. Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

S. Dehaene. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking, 2014.

D. O. Hebb. *The Organization of Behavior*. Wiley, 1949.

T. Hunt and J. W. Schooler. The easy part of the hard problem: A resonance theory of consciousness. *Frontiers in Human Neuroscience*, 13:378, 2019. doi: 10.3389/fnhum.2019.00378.

E. Husserl. *Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins*. Niemeyer, Halle, 1928.

A. G. Huth, W. A. de Heer, T. L. Griffiths, F. E. Theunissen, and J. L. Gallant. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532:453–458, 2016. doi: 10.1038/nature17637.

O. Jensen and L. L. Colgin. Cross-frequency coupling between neuronal oscillations. *Trends in Cognitive Sciences*, 11(7):267–269, 2007. doi: 10.1016/j.tics.2007.05.003.

Michael J. Kane and Randall W. Engle. Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(2):336–358, 2000. doi: 10.1037/0278-7393.26.2.336.

S. H. K. Kang, T. H. Gollan, N. J. Cepeda, and H. Pashler. Spacing makes the difference: Experiments on the interleaving and temporal spacing of study activities. *Psychological Science*, 22(11):1321–1330, 2011. doi: 10.1177/0956797611417467.

K. S. Lashley. In search of the engram. *Symposia of the Society for Experimental Biology*, 4: 454–482, 1950.

Cindy Lustig, Cynthia P. May, and Lynn Hasher. Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, 130(2):199–207, 2001. doi: 10.1037/0096-3445.130.2.199.

J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex. *Psychological Review*, 102(3):419–457, 1995. doi: 10.1037/0033-295X.102.3.419.

G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97, 1956. doi: 10.1037/h0043158.

Karl H. Pribram. *Languages of the Brain*. Prentice Hall, 1969.

G. Tononi. Consciousness as integrated information: A provisional manifesto. *Biological Bulletin*, 215(3):216–242, 2008. doi: 10.2307/25470707.

A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980. doi: 10.1016/0010-0285(80)90005-5.

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

D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393 (6684):440–442, 1998. doi: 10.1038/30918.