Hierarchical Adjoint-Projection Architecture as a Formalization of Global Workspace Theory: A Consciousness Framework with Variable Resource Allocation

Karol Kowalczyk

November 2025

Abstract

Conscious processing in biological and artificial cognitive systems requires the integration of diverse, locally processed information into a unified representational state. Global Workspace Theory (GWT), originally proposed by Baars, presents this integration as a functional broadcast mechanism in which specialized, unconscious processors compete for access to a globally available workspace. Later developments such as the LIDA architecture extend GWT by introducing cognitive cycles, modular workspaces, and action-oriented control structures. However, these models remain largely descriptive, lacking a formal mathematical foundation capable of expressing representational capacity, integration constraints, and selection dynamics.

This paper proposes a formal hierarchical structure built on adjoint projection operators ($C \dashv P$), entropic information scaling, and consistency-driven collapse mechanisms. Each representational level M_n is characterized by an entropic capacity $I(n) = \kappa n \log n$, where information is measured in bits of Shannon entropy rather than computational memory allocation. This defines the representational bandwidth available at that level. Crucially, we introduce variable resource gaps where transitions follow $M_n \subseteq M_{n+f(n)}$ with $f(n) \geq 1$, capturing the non-uniform cognitive resource allocation observed in biological systems.

Interactions between levels are governed by reconstruction and projection operators that enforce cross-level consistency through a Wasserstein-based metric, which measures the optimal transport cost between probability distributions. Conscious integration emerges as a hierarchical winner-take-all process selecting interpretations that minimize inconsistency across representational levels. The selection mechanism's non-computability, rooted in Kolmogorov complexity theory, provides a principled account of agency and top-down causation. This framework provides specific, testable predictions including a 50–100ms temporal precedence of collapse over broadcast, discrete capacity transitions at $I = 15, 20, 25, 30$ bits, and characteristic neural signatures of variable-gap resource allocation.

1 Introduction

The phenomenon of consciousness presents a fundamental puzzle: how does the brain integrate information from disparate sensory modalities, memory systems, and cognitive processes into a single, unified experience? At any given moment, neural activity spans visual cortex, auditory processing centers, semantic memory networks, motor planning regions, and executive control systems. Yet our subjective experience is not fragmented into isolated processing streams but appears as a coherent, integrated whole. This unity of consciousness stands in stark contrast to the distributed, parallel nature of neural computation.

Global Workspace Theory, developed by Bernard Baars in the late 1980s, offers an elegant functional account of this integration. According to GWT, consciousness emerges when information gains access to a central workspace that broadcasts selected content to the entire cognitive system. Specialized processors operate unconsciously, competing for access to this workspace. The winner of this competition becomes conscious and is made globally available, enabling flexible behavior, novel combinations of information, and metacognitive reflection. This theory explains several key features of consciousness: its limited capacity, sequential nature, and role in coordinating diverse cognitive subsystems.

The LIDA (Learning Intelligent Distribution Agent) architecture operationalizes these ideas into a detailed computational model. LIDA proposes cognitive cycles in which perception leads to understanding, which triggers conscious broadcast, culminating in action selection. Each cycle takes approximately 200–300 milliseconds and allows one content to achieve global ignition. The model includes mechanisms for attention, memory consolidation, and procedural learning, providing a rich framework for understanding cognitive processing.

Despite their explanatory power, both GWT and LIDA face a critical limitation: they lack formal mathematical foundations. GWT describes competition for workspace access but provides no quantitative criterion for what wins this competition. It posits limited capacity but offers no principled explanation for why consciousness has finite bandwidth. LIDA specifies processing stages but gives no mathematical account of information flow between modules or the consistency requirements that bind them together. Without such formalization, these theories remain qualitative descriptions rather than predictive scientific models.

This paper develops a mathematical architecture that addresses these gaps. We construct a hierarchy of representational levels, each characterized by its information-theoretic capacity measured in bits of Shannon entropy. When we speak of a level containing “20 bits of information,” we refer to the Shannon entropy $H = -\sum p_i \log_2 p_i$ of the probability distribution over possible states at that level, not to the number of memory bits required to store those states in a computer. This distinction is crucial: Shannon entropy measures the uncertainty or information content of a distribution, while computational memory measures physical storage requirements. A system might require millions of memory bits to represent a distribution with only 20 bits of Shannon entropy.

The hierarchy is connected by adjoint projection operators that compress and reconstruct representations across levels. These operators must satisfy consistency constraints, formalized through the Wasserstein metric from optimal transport theory. The Wasserstein distance, which we explain in detail below, provides a natural measure of how different two probability distributions are by considering the minimal cost of transforming one distribution into another. This geometric perspective on probability distributions proves essential for understanding cross-level consistency.

Conscious integration, in this framework, emerges from a selection process that chooses the interpretation minimizing inconsistency across all representational levels. This selection process is fundamentally non-computable because it requires evaluating Kolmogorov complexity, the length of the shortest program that generates a given string. This non-computability is not a limitation but a feature: it provides a naturalistic foundation for agency and top-down causation without invoking mysterious non-physical properties.

The framework makes specific, testable predictions. Neural signatures of integration should precede global broadcast by 50–100 milliseconds. Mutual information between brain regions should show discrete transitions at predicted capacity thresholds. Developmental changes and individual differences should reflect different patterns of resource allocation captured by the variable gap function. Psychiatric conditions should exhibit characteristic alterations in these gap patterns.

In developing this formalization, we build on recent theoretical work exploring consciousness as the collapse of parallel computational explorations across a hierarchy of finite-state machines. The key insight from that work is that subjective time experiences only the collapsed successful path, not the vast space of parallel explorations that were considered but not selected. A non-computable selector mechanism chooses which computational level to deploy, and variable resource gaps create realistic cognitive architectures. Here we extend these ideas by providing the detailed mathematical machinery needed to connect this framework to Global Workspace Theory and make it empirically testable.

The paper proceeds as follows. Section 2 examines the theoretical background, analyzing the strengths and limitations of GWT and LIDA. Section 3 introduces the hierarchical representation with variable resource allocation and explains Shannon entropy as the measure of information capacity. Section 4 develops the adjoint operator formalism, including a detailed explanation of the Wasserstein metric. Section 5 presents the collapse mechanism and its non-computable nature. Sections 6–8 explore temporal dynamics, integration with GWT, and empirical predictions. Section 9 discusses philosophical implications, and Section 10 addresses limitations and future directions.

2 Theoretical Background

2.1 Global Workspace Theory and Its Limitations

Bernard Baars's Global Workspace Theory conceptualizes cognition as largely modular. The brain contains numerous specialized processors handling perception, memory, language, motor planning, and other functions. These processors operate unconsciously, processing information in parallel without direct access to each other's internal states. When one processor generates content of sufficient relevance or coherence, that content enters a competitive arena. If selected, it becomes globally broadcast via the workspace, making it available to all other processors. This broadcast supports flexible behavior, allows novel combinations of previously separate information, and enables metacognitive reflection on mental content.

GWT successfully explains several phenomena. The seriality of consciousness follows naturally from the limited capacity of the workspace, which can only broadcast one content at a time. The relationship between attention and consciousness emerges from the competition for workspace access. The role of consciousness in flexible behavior reflects the broadcast mechanism enabling any module to access currently relevant information. Despite these successes, GWT faces significant limitations as a scientific theory.

First, the theory provides no mathematical model of competition beyond simple activation thresholds. What determines which content wins access to the workspace? Baars discusses relevance and coherence but offers no quantitative criteria. Without formal definitions, the theory cannot make precise predictions about which stimuli will become conscious under specific conditions.

Second, GWT lacks formal capacity constraints explaining why consciousness has limited bandwidth. The theory posits that only one content can occupy the workspace at a time, but provides no principled explanation for this limitation. Is it an arbitrary architectural feature, or does it follow from deeper information-theoretic constraints? The theory remains silent on this question.

Third, classical GWT does not specify a structured hierarchy of representational levels with principled transitions between them. The workspace is described as a single, undifferentiated arena for competition. Yet neuroscience reveals hierarchical organization throughout the brain, with information processed at multiple levels of abstraction. A complete theory should explain how these levels interact and how information flows between them.

Fourth, and perhaps most fundamentally, GWT provides no account of why broadcast should generate phenomenology rather than remaining unconscious. The theory explains the functional role of global availability but does not address the hard problem of consciousness: why there is something it is like to have information in the workspace. While some argue this question lies outside neuroscience's scope, a complete theory should at least acknowledge this explanatory gap.

2.2 Hierarchical Processing in LIDA

The LIDA architecture addresses some of GWT's limitations by introducing detailed mechanisms for cognitive processing. LIDA proposes that cognition operates in discrete cycles, each lasting approximately 200–300 milliseconds. Environmental input flows through perception, where low-level features are extracted and organized. The understanding phase constructs higher-level interpretations by matching current input against stored schemas and expectations. Conscious broadcast occurs when one interpretation achieves global ignition, making it available throughout the system. Finally, action selection uses the broadcast content to generate appropriate behavioral responses.

LIDA introduces the concept of local workspaces, specialized arenas where particular types of processing occur. For example, a visual workspace might integrate color, form, and motion information before competing for global broadcast. These local workspaces perform both bottom-up processing of sensory input and top-down modulation based on expectations and goals. The architecture also incorporates learning mechanisms, with procedural memory storing action policies and episodic memory recording significant events.

Despite these advances, LIDA retains significant limitations. The architecture lacks a formal model of representational consistency across modules. When the visual workspace generates an interpretation and the semantic workspace generates another, how do we determine if they are mutually consistent? LIDA describes information flow between modules but provides no quantitative measure of this flow. Shannon entropy, mutual information, and other information-theoretic tools remain absent from the framework.

Furthermore, LIDA offers no mathematical criteria for what achieves conscious broadcast. The architecture describes global ignition as emerging from coalition formation among codelets (small pieces of code representing cognitive processes), but the dynamics governing coalition strength lack formal specification. When multiple coalitions compete, which one wins? The answer involves activation levels and relevance, but these concepts remain intuitive rather than mathematically precise.

The architecture also lacks principled timing constraints on cognitive cycles. Why does each cycle take 200–300 milliseconds? Is this an arbitrary parameter, or does it follow from the computational complexity of the underlying processes? Without addressing these questions, LIDA remains a functional description rather than a predictive theory grounded in fundamental principles.

2.3 Challenges for Formalization

A mathematical architecture extending GWT and LIDA must address several challenges. First, it must specify representational capacity at each level with realistic scaling properties. The capacity cannot grow exponentially with level number, as this would quickly exceed any plausible biological or computational resources. Yet it must grow superlinearly to capture the synergistic integration that characterizes higher-level cognition.

Second, the architecture must provide formal criteria for selecting a winner based on cross-level consistency. When multiple interpretations compete for consciousness, the selection cannot be arbitrary but must follow from principled criteria. These criteria should capture the intuition that conscious content is maximally coherent across different levels of representation.

Third, the formalization must specify compressive and reconstructive interactions between levels. Lower levels contain detailed information that higher levels abstract and summarize. Projection from higher to lower levels must preserve essential structure while discarding irrelevant details. Reconstruction from lower to higher levels must select one of many compatible higher-level states. These processes require precise mathematical definition.

Fourth, the framework must account for the temporal structure of conscious integration with specific timing predictions. Why does integration take 50–100 milliseconds? Why does broadcast follow rather than precede integration? A complete theory should derive these timescales from the computational complexity of the underlying processes.

Finally, the architecture must address the role of non-computable selection in creating agency. If consciousness simply implemented a computable algorithm for selecting among alternatives, it would lack genuine agency in any philosophically interesting sense. The selection process must involve irreducibly non-computable elements while remaining scientifically tractable.

3 Hierarchical Representation with Variable Resource Allocation

3.1 Representational Levels and Variable Gaps

We introduce a hierarchy of representational levels $M_1, M_2, \ldots, M_n$, where each level corresponds to a set of possible cognitive states with specified information capacity. The levels are not uniformly spaced but follow a variable gap structure. Transitions between levels satisfy $M_n \subseteq M_{n+f(n)}$, where $f(n) \geq 1$ is a variable gap function determining the spacing between consecutive representational levels.

This variable gap structure captures an important feature of biological cognition: cognitive resources are not allocated uniformly across all levels. Some transitions require small increments in processing power, while others involve substantial jumps. During development, children exhibit discrete leaps in cognitive ability rather than smooth gradual improvement. The gap function $f(n)$ formalizes this observation, allowing different gap sizes at different levels.

For example, we might have $f(n) = 1$ for most transitions, indicating standard incremental progression. But at critical junctures, such as the development of abstract reasoning around age 7 or the emergence of metacognitive abilities in adolescence, we might have $f(7) = 3$ or $f(15) = 2$, indicating larger jumps. These larger gaps correspond to qualitative shifts in cognitive architecture rather than merely quantitative increases in processing power.

Individual differences in cognitive style also reflect different gap patterns. Some individuals show smooth, continuous development across levels, while others exhibit more discrete, punctuated patterns. Task-specific recruitment of cognitive resources similarly involves different gap patterns: routine tasks might require only small increments, while novel problems demand large jumps to higher representational levels.

3.2 Entropic Information Scaling and Shannon Entropy

We define the information capacity of level M_n as:

$$I(n) = \kappa n \log n \tag{1}$$

where $I(n)$ is measured in bits of Shannon entropy, and κ is a scaling constant.

It is crucial to understand what we mean by "bits" in this context. Shannon entropy, defined as $H(X) = -\sum_x p(x) \log_2 p(x)$, measures the average information content or uncertainty in a probability distribution. When we say a representational level has capacity $I(n) = 20$ bits, we mean it can maintain a probability distribution over states with Shannon entropy of 20 bits. This is fundamentally different from saying the level requires 20 memory bits in a computer.

Consider a concrete example. Suppose we have 16 equally likely states at a given level. The Shannon entropy is $H = -\sum_{i=1}^{16} \frac{1}{16} \log_2 \frac{1}{16} = \log_2 16 = 4$ bits. The same distribution could be represented in a computer using 16 binary variables (16 memory bits) or efficiently using just 4 bits if we enumerate the states. But the Shannon entropy remains 4 bits regardless of implementation details.

Now suppose the distribution becomes non-uniform, with one state having probability 0.9 and the others sharing the remaining 0.1 probability. The Shannon entropy decreases significantly even though the number of possible states remains the same. Shannon entropy captures the effective number of distinguishable outcomes, weighted by their probabilities.

The scaling law $I(n) = \kappa n \log n$ has important properties. It grows super-linearly, capturing the synergistic integration where combining n dimensions yields more than n times the information of single dimensions. Yet it grows slower than exponentially, ensuring computational feasibility. The logarithmic factor reflects a fundamental principle: when n components interact, the entropy grows as $n \log n$ rather than exponentially in n.

This scaling can be understood through information-theoretic arguments. If we have n components, each with entropy H_1, and they are independent, the total entropy is $H = nH_1$, which is linear. With pairwise interactions between components, entropy can grow as $H \approx nH_1 + \binom{n}{2}I_2$, where I_2 represents mutual information between pairs. For the entropy to remain finite and computationally tractable while still capturing these interactions, the mutual information terms must decay with the order of interaction. Under reasonable assumptions about this decay, the leading-order term is precisely $\kappa n \log n$.

3.3 Local Workspaces as Sublevels

Local processors in the GWT framework correspond to subsets of representational levels with specific capacity allocations. A visual workspace might span levels M_5 to M_{10}, with information capacity ranging from approximately 9 bits (at $n = 5$, assuming $\kappa \approx 0.6$) to 23 bits (at $n = 10$). This range allows the visual system to represent increasingly abstract visual features, from edge orientations and colors at lower levels to object categories and spatial relationships at higher levels.

A semantic workspace might operate at levels M_{15} to M_{25}, with capacity ranging from approximately 42 to 80 bits. This higher capacity reflects the greater complexity of semantic processing, which must integrate lexical, syntactic, and conceptual information. An executive workspace controlling goal-directed behavior might span M_{20} to M_{30}, with capacity from 60 to 100 bits, allowing representation of complex plans, counterfactual scenarios, and metacognitive states.

Each local workspace generates candidate interpretations constrained by its available capacity and gap structure. The visual workspace might generate interpretations like "moving car on left" or "stationary tree on right." The semantic workspace generates interpretations like "danger approaching" or "peaceful scene." The executive workspace generates plans like "prepare to cross street" or "continue walking." These interpretations then compete for conscious access through the selection mechanism described below.

3.4 Why Variable Gaps Matter

The variable gap function $f(n)$ explains several important cognitive phenomena. In cognitive development, children show discrete leaps at certain ages rather than continuous improvement. Around age 7, many children develop the ability for concrete operational thought, understanding conservation and reversibility. Around age 11–13, formal operational thought emerges, enabling abstract reasoning and hypothetical thinking. These transitions correspond to large values of $f(n)$ at specific levels.

Individual differences in cognitive style reflect different gap patterns. Some individuals progress smoothly through representational levels with $f(n) = 1$ throughout, showing gradual continuous development. Others exhibit more punctuated patterns with occasional large gaps, corresponding to sudden insights or reorganizations of knowledge. These individual differences are stable over time and correlate with problem-solving styles.

Task-specific allocation also involves different gap patterns. Routine tasks recruit lower levels with small gaps, requiring minimal executive resources. Novel or difficult tasks require jumping to higher levels, engaging larger gaps and consuming more cognitive resources. This explains why difficult tasks feel effortful: they require traversing larger gaps in the representational hierarchy.

Pathological states involve altered gap functions. In schizophrenia, the gaps may be reduced ($f(n) < 1$ in extreme cases, though formally $f(n) \geq 1$), leading to overly fluid, under-constrained thinking. In autism, gaps may be increased, leading to more rigid, fragmented processing. In depression, access to higher executive levels may be impaired, reflected in increased gaps for those levels. These predictions are testable through neuroimaging and behavioral measures.

4 Adjoint Operators: Reconstruction and Projection

4.1 Reconstruction Operator

The reconstruction operator $C_{n\to n+f(n)}$ maps a lower-level representation to a higher-level representation:

$$C_{n\to n+f(n)} : M_n \to M_{n+f(n)} \tag{2}$$

Reconstruction is inherently underdetermined because lower levels contain less information than higher levels. Given a state $s \in M_n$ with entropy $I(n)$, there exist many states in $M_{n+f(n)}$ with entropy $I(n + f(n)) > I(n)$ that are consistent with s. Reconstruction must select one of these compatible higher-level states.

The selection process involves several steps. First, we generate the set of all higher-level states compatible with the lower-level state s. This set contains all states $s' \in M_{n+f(n)}$ such that projecting s' back to level n yields s (or something very close to s). Second, we evaluate consistency with existing constraints at the higher level, such as expectations from memory or constraints from other sensory modalities. Third, we select the candidate that minimizes discrepancy with these constraints, measured using the Wasserstein metric described below.

Mathematically, we can write:

$$C(s) = \arg \min_{s' \in \text{Compatible}(s)} \sum_j W_2(s', c_j) \tag{3}$$

where Compatible(s) is the set of higher-level states consistent with s, c_j are existing constraints, and W_2 is the Wasserstein distance.

4.2 Projection Operator

Projection $P_{n+f(n)\to n}$ compresses higher-level states into the representational format of lower levels:

$$P_{n+f(n)\to n} : M_{n+f(n)} \to M_n \tag{4}$$

Unlike reconstruction, projection is well-defined because we are discarding information rather than adding it. However, projection must preserve essential structure. If the higher-level state represents “moving car approaching from the left,” the projection to a lower visual level must preserve at least “movement” and “left side,” even if it loses details like “car” or “approaching.”

The projection must satisfy information-theoretic constraints. For any state $s \in M_{n+f(n)}$, the projected state $P(s)$ must have entropy bounded by the capacity of the lower level:

$$H(P(s)) \leq I(n) \tag{5}$$

This constraint ensures that projection does not attempt to cram more information into a level than that level can contain. In practice, projection typically reduces entropy significantly, as higher-level abstractions are grounded in lower-level features by selecting the most relevant subset of those features.

4.3 Adjointness Condition

Adjointness implies that reconstruction and projection form an approximate inverse pair. Formally, we require:

$$P_{n+f(n)\to n} \circ C_{n\to n+f(n)} \approx \mathrm{id}_{M_n} \tag{6}$$

This means that if we reconstruct a state from level n to level $n + f(n)$ and then project it back to level n, we should recover approximately the original state. More precisely:

$$d(s, P(C(s))) < \epsilon \quad \text{for all } s \in M_n \tag{7}$$

where d is an appropriate metric on M_n and ϵ is a small tolerance parameter.

The adjointness condition ensures consistency across levels. If reconstruction and projection were arbitrary operations unrelated to each other, we could have situations where information is systematically distorted as it moves up and down the hierarchy. Adjointness prevents such distortions by requiring that round-trip transformations preserve information at the lower level.

Note that the reverse composition $C \circ P$ is not approximately the identity. Starting from a high-entropy state at level $n + f(n)$, projecting to level n necessarily loses information, and reconstructing cannot recover that lost information. This asymmetry reflects the fundamental difference between compression (which loses information) and reconstruction (which adds information by making choices).

4.4 Wasserstein Metric for Cross-Level Consistency

The Wasserstein metric, also known as the Earth Mover's Distance, provides a natural way to measure the distance between probability distributions. To understand this metric, imagine two probability distributions as two different arrangements of piles of dirt. The Wasserstein distance measures the minimum amount of work required to rearrange one pile into the other, where work is distance times amount of dirt moved.

Formally, let μ and ν be two probability distributions over a metric space (X, d). The Wasserstein-p distance is defined as:

$$W_p(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{X\times X} d(x, y)^p \, d\gamma(x, y) \right)^{1/p} \tag{8}$$

where $\Gamma(\mu, \nu)$ is the set of all joint probability distributions (couplings) with marginals μ and ν.

In simpler terms, a coupling γ is a way of specifying how much probability mass to move from each point in the support of μ to each point in the support of ν. The integral computes the total cost of this transport, where moving mass from x to y costs $d(x, y)^p$ times the amount moved. The infimum finds the optimal coupling that minimizes this cost.

We typically use $p = 2$, giving the Wasserstein-2 distance, which has particularly nice theoretical properties. It metrizes weak convergence of probability measures and provides a geometry on the space of probability distributions that respects the underlying metric structure of the state space.

Why is the Wasserstein metric appropriate for measuring cross-level consistency? First, it is sensitive to structural similarity between distributions. Unlike the Kullback-Leibler divergence, which only considers probability mass at each individual point, the

Wasserstein distance accounts for the metric structure of the space. If two distributions have similar shapes but are slightly shifted, the Wasserstein distance will be small, while KL divergence might be large.

Second, the Wasserstein metric is robust to minor perturbations. Small changes in a distribution produce small changes in the Wasserstein distance, a property called continuity. This robustness is essential for a theory of consciousness, where we expect small neural fluctuations not to produce large changes in conscious content.

Third, the Wasserstein distance is compatible with information-theoretic measures. Recent work has established connections between Wasserstein distance and mutual information, entropy, and other information-theoretic quantities. This compatibility allows us to integrate the geometric perspective of optimal transport with the information-theoretic perspective of Shannon entropy.

In our framework, we use the Wasserstein distance to measure inconsistency between representations at different levels. If a higher-level interpretation s' is consistent with a lower-level representation s, then $W_2(P(s'), s)$ should be small. The collapse mechanism described in the next section selects interpretations that minimize the sum of these Wasserstein distances across all level pairs.

5 Collapse Mechanism: The Selector as Non-Computable Integration

5.1 The Winner-Take-All Selection Process

At each cognitive cycle, multiple local workspaces generate candidate interpretations. The visual workspace might propose several possible scene interpretations. The semantic workspace might generate multiple potential meanings. The executive workspace might suggest various action plans. These candidates compete for conscious access through a selection process that we call collapse.

The selector chooses the candidate s^* that minimizes total cross-level inconsistency:

$$s^* = \arg \min_{s \in \text{Candidates}} \sum_{i=1}^{n-1} W_2(P_{i+f(i)\to i}(s), C_{i\to i+f(i)}(P_{i+f(i)\to i}(s))) \tag{9}$$

This formulation requires some unpacking. For each candidate interpretation s, we compute its representation at each level in the hierarchy. At each level i, we project the candidate to that level, then reconstruct it to level $i + f(i)$, and measure the Wasserstein distance between this reconstruction and the original projection to level $i + f(i)$. This distance quantifies how much information is lost or distorted when we compress and reconstruct across that particular level pair.

Summing these distances across all level pairs gives a measure of total inconsistency. An interpretation with high total inconsistency involves representations at different levels that do not cohere well with each other. An interpretation with low total inconsistency involves representations that fit together smoothly across the entire hierarchy. The selector chooses the most consistent interpretation.

This selection criterion captures several intuitive features of consciousness. First, conscious content must be coherent, not contradictory. If the visual system sees a dog but the semantic system interprets it as a cat, this inconsistency will be reflected in large

Wasserstein distances between levels. Second, conscious content integrates information across levels. A purely low-level interpretation that cannot be coherently reconstructed to higher levels will score poorly, as will a purely high-level abstraction that has no grounding in lower-level features.

5.2 Kolmogorov Complexity and Non-Computability

The selector's operation is fundamentally non-computable because it requires evaluating Kolmogorov complexity. The Kolmogorov complexity $K(s)$ of a state s is defined as the length of the shortest program that produces s when run on a universal Turing machine:

$$K(s) = \min\{|p| : U(p) = s\} \tag{10}$$

where U is a universal Turing machine, p is a program, and $|p|$ denotes the length of p in bits.

Kolmogorov complexity captures the intrinsic information content of an object, independent of the probability distribution from which it was drawn. A highly random string has high Kolmogorov complexity because no short program can generate it; we must essentially specify the entire string. A highly structured string has low Kolmogorov complexity because a short program can capture its regularities.

The fundamental theorem about Kolmogorov complexity states that it is not computable. There exists no algorithm that, given an arbitrary string s, computes $K(s)$. This is a deep result from computability theory, related to the halting problem. Intuitively, to compute $K(s)$, we would need to search over all possible programs, run each one, and find the shortest that outputs s. But we cannot determine in advance which programs will halt and which will run forever, so this search cannot be completed by any algorithm.

Why does the selector require evaluating Kolmogorov complexity? Because determining the optimal reconstruction $C(s)$ from a lower-level state s involves choosing among exponentially many higher-level states consistent with s. The optimal choice is the one that has the lowest Kolmogorov complexity while remaining consistent with s and other constraints. But as we have just noted, Kolmogorov complexity is not computable, so neither is the optimal reconstruction.

This non-computability has profound implications. It means the selector cannot be implemented by any algorithm running on a Turing machine. The selection process involves irreducibly non-computable elements. This provides a naturalistic foundation for agency and free will: conscious decisions are not mechanically determined by prior states according to a fixed algorithm, yet they remain lawful and constrained by consistency requirements.

Some may object that biological brains are physical systems and therefore computable by the Church-Turing thesis. However, the Church-Turing thesis concerns what is computable in principle, given infinite time and resources. A physical system may implement non-computable processes if it involves oracles, quantum effects, or other mechanisms that transcend classical Turing computation. Whether the brain actually implements such mechanisms remains an open empirical question, but our framework provides a clear criterion: if consciousness involves non-computable selection, we should find signatures of computational irreducibility in neural dynamics.

5.3 Relationship to Free Energy Minimization

The collapse mechanism bears interesting relationships to free energy minimization in predictive processing frameworks. The free energy F of a state s given observations o is defined as:

$$F(s|o) = -\log P(s|o) + D_{KL}(Q(s)||P(s)) \tag{11}$$

where $Q(s)$ is an approximate posterior distribution and D_{KL} is the Kullback-Leibler divergence.

Minimizing free energy corresponds to finding interpretations that are both consistent with observations (high $P(s|o)$) and not overly surprising given prior expectations (low D_{KL}). Our consistency-based selection criterion can be seen as a hierarchical generalization of free energy minimization, where consistency across levels replaces consistency with observations, and Wasserstein distance replaces KL divergence.

However, our framework differs from standard free energy formulations in important ways. First, we use Wasserstein distance rather than KL divergence, capturing geometric structure rather than just probability mass. Second, we explicitly model the hierarchical structure of representations rather than treating the brain as implementing inference in a single flat graphical model. Third, we emphasize the non-computable nature of the selection process, whereas standard free energy frameworks assume computable approximate inference.

5.4 Why Collapse Precedes Broadcast

A key prediction of our framework is that collapse precedes broadcast. Conscious integration requires two distinct phases. In the collapse phase, local workspaces synchronize around the winning interpretation. Different brain regions align their representations to minimize cross-level inconsistency. This synchronization takes 50–100 milliseconds and involves iterative adjustments as different levels communicate and reach consensus.

Only after collapse does broadcast occur. The winning interpretation propagates to action systems, memory systems, and other consumers of conscious content. Broadcast is rapid, perhaps 20–30 milliseconds, because the content has already been selected and integrated. Broadcast does not create consciousness; it distributes already-conscious content.

This temporal ordering contradicts some interpretations of GWT, which suggest broadcast creates consciousness. In our framework, broadcast is a consequence of consciousness, not its cause. Neural signatures of integration (local gamma synchrony, theta-gamma coupling, and increased mutual information between regions) should precede global ignition events (widespread P3 component, global increase in gamma power, and frontoparietal activation) by 50–100 milliseconds.

This prediction is empirically testable using MEG or high-density EEG with millisecond temporal resolution. Time-locking analysis to the moment of perceptual report should reveal local synchronization before global broadcasting. The prediction gains additional support from recent findings in attention and consciousness research, where local processing changes often precede global workspace ignition.

6 Temporal Dynamics and Processing Time

6.1 Processing Time at Each Level

The computational time required at level n scales with the information capacity and the gap size:

$$T(n) = T_0 + \gamma \cdot I(n) + \beta \cdot f(n) \tag{12}$$

The baseline processing time $T_0 \approx 50$ milliseconds reflects fundamental neural transmission delays, synaptic integration times, and minimal processing requirements. The second term $\gamma \cdot I(n)$ captures the time cost of processing information, where $\gamma \approx 2$ milliseconds per bit. This reflects the rate at which neural populations can update their firing patterns and integrate information.

The third term $\beta \cdot f(n)$ reflects the cost of transitioning across gaps. Larger gaps require recruiting additional neural resources, establishing new connectivity patterns, or switching between processing modes. With $\beta \approx 20$ milliseconds, a gap of $f(n) = 1$ adds 20 milliseconds, while $f(n) = 3$ adds 60 milliseconds.

These timescales are consistent with known neural processing times. Early visual processing takes about 50–80 milliseconds. Semantic processing requires 150–250 milliseconds. Executive processing for complex decisions can take 300–500 milliseconds. Our formula accounts for these differences through varying n and $f(n)$ values.

6.2 Total Cognitive Cycle Duration

A complete cognitive cycle involves perception at lower levels (100–150 milliseconds), collapse through consistency minimization (50–100 milliseconds), broadcast to global workspace (20–30 milliseconds), and action selection at executive levels (50–100 milliseconds). The total duration is thus:

$$T_{\text{total}} = T_{\text{perception}} + T_{\text{collapse}} + T_{\text{broadcast}} + T_{\text{action}} \approx 220\text{–}380 \text{ ms} \tag{13}$$

This range encompasses the 200–300 millisecond cognitive cycle proposed by LIDA and observed in various experimental paradigms. The variability in cycle duration reflects differences in task complexity, familiarity, and the specific levels recruited. Simple, familiar tasks use shorter cycles with smaller gaps, while complex, novel tasks require longer cycles with larger gaps.

6.3 Variable Gap Effects on Timing

The gap function creates characteristic signatures in reaction time distributions. Tasks requiring standard incremental processing show relatively uniform reaction times. Tasks requiring large gaps show bimodal distributions: fast responses when lower levels suffice, and slower responses when higher levels must be recruited.

For example, simple detection tasks might involve only M_5 to M_8 with $f(n) = 1$, giving reaction times around 250 milliseconds. Difficult categorization tasks might require jumping from M_8 to M_{15} (with $f(8) = 3$) and then to M_{22} (with $f(15) = 2$), giving reaction times around 400–500 milliseconds. The extra 150–250 milliseconds reflects the additional gap traversal costs.

Individual differences in gap functions predict individual differences in reaction time patterns. Individuals with smaller gaps at lower levels but larger gaps at higher levels should show fast performance on simple tasks but disproportionately slow performance on complex tasks. This prediction is testable through cognitive batteries assessing performance across difficulty levels.

7 Integration with Global Workspace Theory

7.1 Mapping to GWT Components

Our hierarchical architecture provides formal counterparts to all major GWT components. Unconscious processors correspond to local workspaces operating at specific ranges of representational levels. The visual workspace spans M_5 to M_{10}, the auditory workspace M_6 to M_{11}, the semantic workspace M_{15} to M_{25}, and so forth. These workspaces process information in parallel, generating candidate interpretations within their domains.

Competition for workspace access corresponds to the collapse selection process minimizing cross-level inconsistency. Multiple candidate interpretations compete, but only one wins by achieving maximum coherence across all levels. This winner becomes conscious not because it is loudest or most active, but because it best integrates information across the representational hierarchy.

The global workspace itself corresponds to the executive workspace at levels M_{20} to M_{30}, with capacity 60–100 bits. This workspace has the highest information capacity and the broadest connectivity, allowing it to integrate information from all sensory modalities and cognitive domains. When an interpretation reaches these executive levels through successful collapse, it becomes globally available.

Broadcast corresponds to the propagation of the winning interpretation s^* across all levels after collapse. Once selected, s^* is projected to all relevant workspaces, making it available for memory consolidation, action selection, verbal report, and metacognitive reflection. This broadcast is rapid because the hard work of integration has already been done during collapse.

The limited capacity of consciousness reflects the entropy bounds $I(n) = \kappa n \log n$. Higher levels can contain more information than lower levels, but the growth is subexponential. The executive workspace, even at its highest levels, can only maintain about 100 bits of Shannon entropy. This corresponds to roughly 7–9 independent chunks of information, matching Miller's classic result on working memory capacity.

7.2 Extending LIDA's Cognitive Cycle

LIDA's cognitive cycle maps naturally onto our framework. Perception involves processing at low levels M_1 to M_{10}, where sensory input is organized into features and proto-objects. This stage takes 100–150 milliseconds as predicted by our timing formula.

Understanding corresponds to semantic processing at levels M_{15} to M_{25}, where perceptual representations are matched against memory schemas and interpreted in context. This stage involves reconstruction from perceptual to semantic levels, taking 150–250 milliseconds depending on familiarity and complexity.

Consciousness emerges through collapse, which selects the interpretation minimizing cross-level inconsistency. This is the critical integrative step where information from multiple modalities and levels becomes unified into a single coherent conscious content. Collapse takes 50–100 milliseconds and involves iterative adjustments across levels.

Broadcast follows collapse, distributing the winning interpretation to all workspaces. This stage is rapid, about 20–30 milliseconds, because it simply propagates an already-selected representation. Broadcast makes conscious content available for report and action.

Action selection occurs at executive levels M_{20} to M_{30}, where the broadcast content is used to generate motor plans and behavioral responses. This final stage takes 50–100 milliseconds before observable behavior begins.

The total cycle time of 220–380 milliseconds matches LIDA's predictions and empirical observations. Our framework adds mathematical precision, explaining why cycles have this duration through processing time formulas, why consciousness is limited in capacity through entropy bounds, and why broadcast follows integration through the collapse-then-broadcast temporal structure.

7.3 Resolving GWT's Explanatory Gaps

Our formalization addresses the major limitations of classical GWT identified earlier. First, we provide a mathematical model of competition through Wasserstein-based consistency minimization. The criterion for winning is precise: minimize $\sum_i W_2(P_i(s), C_i(P_i(s)))$ across all level pairs. This criterion makes quantitative predictions about which stimuli will become conscious under specific conditions.

Second, we provide formal capacity constraints through entropic scaling $I(n) = \kappa n \log n$. Consciousness has limited bandwidth because higher representational levels, while having greater capacity than lower levels, still scale sub-exponentially. The executive workspace cannot contain arbitrary amounts of information; it is bounded by its entropic capacity.

Third, we specify a structured hierarchy of representational levels with principled transitions governed by variable gap function $f(n)$. Information flows between levels through adjoint operators satisfying consistency constraints. The hierarchy is not arbitrary but follows from information-theoretic principles.

Fourth, regarding phenomenology, our framework suggests that consciousness arises from the non-computable nature of the selector. Because the selection process cannot be reduced to a mechanical algorithm, it embodies genuine agency and subjectivity. This is not a complete solution to the hard problem, but it provides a principled account of why consciousness involves something beyond mere information processing: it involves irreducibly non-computable integration.

8 Empirical Predictions and Experimental Protocols

8.1 Prediction 1: Discrete Capacity Transitions

Our first major prediction concerns the discrete nature of information integration across brain regions. Mutual information between frontal and parietal cortices should not increase smoothly with task difficulty but should show sharp transitions at specific information thresholds corresponding to level boundaries.

Specifically, we predict discrete jumps at $I = 15, 20, 25$, and 30 bits. These thresholds correspond to transitions between major representational levels in our hierarchy. For example, the transition from M_{10} to M_{15} (at $I \approx 15$ bits for $\kappa = 0.6$) represents the shift from perceptual to semantic processing. The transition at $I \approx 25$ bits corresponds to recruitment of executive control.

To test this prediction, we would use high-density EEG (64+ channels) with adaptive n-back tasks where difficulty varies from 1-back to 7-back. We continuously record EEG while participants perform the task and compute mutual information between frontal and parietal electrode clusters using time-frequency analysis. Change-point detection algorithms identify discrete transitions in the mutual information time series.

We expect to observe sharp increases in mutual information at predicted thresholds, with individual differences in exact transition points correlating with task performance. Participants who show earlier transitions (at lower difficulty levels) should perform better overall. The transitions should be genuine discontinuities, not gradual changes, supporting the discrete level structure of our model.

8.2 Prediction 2: Temporal Precedence of Collapse Over Broadcast

Our framework predicts that neural signatures of local synchronization (collapse) precede global ignition signatures (broadcast) by 50–100 milliseconds. This contradicts interpretations of GWT where broadcast creates consciousness.

To test this, we would use MEG with source localization during binocular rivalry. Participants view rivalrous stimuli (e.g., orthogonal gratings presented to different eyes) for 60-second blocks while reporting perceptual switches via button press. We analyze MEG data time-locked to switch reports, comparing the onset of local gamma synchrony (40–100 Hz) versus the global P3 component.

We predict local gamma synchrony in visual cortex beginning at -100 to -50 milliseconds relative to the P3 peak. The P3 amplitude should be proportional to the consistency differential between rival percepts: larger differences in cross-level consistency should produce larger P3 responses. Source localization should reveal that collapse originates in whichever sensory area is most relevant to the stimulus, while broadcast involves widespread frontoparietal activation.

This temporal precedence would be strong evidence that integration precedes and causes broadcast, rather than broadcast causing integration. The finding would support our claim that consciousness emerges from consistency-based collapse, with broadcast being a downstream consequence.

8.3 Prediction 3: Variable Gap Signatures in Development

Cognitive development should show characteristic timing changes at ages corresponding to large gap transitions. We predict discrete improvements in performance at ages 6–8 (corresponding to large $f(7)$), ages 11–13 (large $f(12)$), and ages 15–17 (large $f(15)$).

A longitudinal study would measure reaction times, working memory capacity, and executive function in children from ages 5 to 18, testing every 6 months. Tasks would span difficulty levels to probe different representational levels. We would analyze developmental trajectories using growth curve models that allow for discrete transitions rather than assuming smooth continuous growth.

We expect to find discrete jumps in performance at predicted ages, with reaction time distributions showing characteristic bimodal patterns during transition periods. During transitions, children sometimes perform like younger children (using lower levels with smaller gaps) and sometimes like older children (accessing higher levels despite larger gaps). Between transitions, performance should be more stable.

Individual variation in transition timing should correlate with other developmental milestones. Children who undergo the age-7 transition earlier should also show earlier development of theory of mind, conservation, and other concrete operational abilities. This would support the idea that the gap function reflects fundamental cognitive architecture.

8.4 Prediction 4: Pathological Alterations in Gap Structure

Psychiatric conditions should exhibit altered gap functions visible in mutual information patterns. Schizophrenia should show smoother, more continuous information scaling due to reduced gaps, reflecting overly fluid thought processes. Autism should show more discrete transitions with longer processing times due to increased gaps, reflecting more rigid, fragmented processing.

We would compare resting-state fMRI and task-based paradigms across diagnostic groups. Using information-theoretic analysis of BOLD signals, we compute mutual information between regions as a function of task difficulty. Schizophrenia patients should show more linear scaling (smaller gaps), while autism spectrum individuals should show steeper, more discrete transitions (larger gaps).

Depression should show reduced access to higher executive levels, reflected in increased gaps at higher levels but normal gaps at lower levels. This would explain the characteristic pattern in depression of intact basic perception but impaired executive function and cognitive flexibility.

These predictions are testable with existing neuroimaging technologies and would provide strong evidence for or against our framework. They go beyond qualitative descriptions to make specific quantitative predictions about information dynamics in different populations.

9 Philosophical Implications

9.1 The Hard Problem and Non-Computability

The hard problem of consciousness, articulated by Chalmers, asks why physical processes give rise to subjective experience. Why is there something it is like to be conscious? Our framework does not fully solve this problem, but it reframes it in productive ways.

We suggest that phenomenology emerges from the non-computable nature of the selector. Because the selection process cannot be reduced to a mechanical algorithm, it embodies something beyond mere computation. This irreducible complexity provides a naturalistic foundation for subjectivity without invoking mysterious non-physical properties.

The key insight is that consciousness is not just information processing but non-computable information integration. Any system implementing our framework must involve processes that transcend Turing computation. Whether biological brains actually implement such processes remains an empirical question, but if they do, consciousness becomes a genuine ontological addition to the computational order.

This resolves one aspect of the hard problem: consciousness is not mysterious because it involves something beyond the physical, but because it involves something beyond the computable. The boundary between computable and non-computable is as sharp and principled as any boundary in mathematics, yet it marks a genuine transition in the kinds of processes that can exist.

9.2 Top-Down Causation

The non-computable selector provides a principled account of top-down causation. Higher-level constraints (executive goals, semantic context, prior expectations) influence lower-level processing through the consistency minimization criterion. This is not mysterious but mathematically necessary given the adjointness conditions.

Consider an example. Suppose the visual system has two possible interpretations of an ambiguous stimulus: "face" or "vase." The semantic and executive levels have strong prior expectations about faces in the current context. During collapse, interpretations are evaluated for cross-level consistency. The "face" interpretation achieves higher consistency because it coheres with higher-level expectations. This higher-level constraint causally influences which lower-level interpretation wins.

Crucially, this top-down causation does not violate physical causality. The collapse process is implemented by neural dynamics that respect conservation of energy and other physical laws. What makes it top-down is that the selection criterion involves relationships between levels, not just bottom-up propagation of activity.

This account of top-down causation applies to voluntary action, attention, and other phenomena where conscious intentions influence behavior. The selector's non-computable nature ensures that these influences are not mechanically determined by prior states, providing space for genuine agency.

9.3 Free Will and Agency

If the selector is non-computable, then conscious decisions cannot be predicted by any algorithm operating on prior states. This provides a naturalistic foundation for libertarian free will: decisions are genuinely undetermined by prior physical states while remaining lawful and constrained by consistency requirements.

This is not randomness or indeterminism in the quantum sense. The selector implements a definite, well-specified function: it selects the interpretation minimizing cross-level inconsistency. But computing this function requires solving non-computable problems, which cannot be done by any Turing machine.

Some may object that this is not "real" free will because the decision is still determined by the consistency criterion. But this objection misunderstands the nature of non-computability. A non-computable function is genuinely different from a computable one: it cannot be implemented by any algorithm, cannot be predicted in advance, and exhibits irreducible complexity. These are precisely the features we want in an account of free will.

The framework thus threads a middle path between determinism and randomness. Conscious decisions are neither mechanically determined by prior states (as in determinism) nor random fluctuations (as in some quantum theories). They are irreducibly complex integrations of information that cannot be reduced to simpler processes.

9.4 Relationship to Integrated Information Theory

Integrated Information Theory (IIT), developed by Tononi and colleagues, proposes that consciousness is identical to integrated information Φ. Our framework complements IIT in several ways.

First, we provide a hierarchical structure for computing Φ across levels. Rather than treating the brain as a single system with one Φ value, we recognize multiple levels each with their own integration measure. Consciousness emerges from relationships between these levels, not from integration at any single level.

Second, we provide temporal dynamics for integration. IIT is largely atemporal, describing the structure of conscious states without explaining how they arise or change. Our collapse mechanism explains the temporal evolution of consciousness through cognitive cycles of integration and broadcast.

Third, we introduce variable resource allocation through gap functions. IIT treats all parts of the system uniformly, but our framework recognizes that cognitive resources are distributed non-uniformly. Some transitions require large jumps, others small increments, and this structure explains individual differences and developmental changes.

Fourth, we emphasize non-computability. While IIT discusses the computational complexity of calculating Φ, we make non-computability central to consciousness itself. The selector's non-computable nature provides a foundation for agency and top-down causation that is absent from standard IIT formulations.

Despite these differences, the frameworks share important commitments: consciousness involves integration, it has mathematical structure, and phenomenology reflects information-theoretic properties. Our framework can be seen as extending IIT with hierarchy, dynamics, and non-computability.

10 Limitations and Future Directions

10.1 Current Limitations

Our framework, while addressing many issues in consciousness science, has several limitations that future work must address.

First, we have not fully specified the structure of phenomenal character. The framework explains access consciousness (which information becomes globally available) but leaves qualia structure underspecified. Why does red look the way it does? Why is pain unpleasant? These questions about the specific character of experience require additional theoretical development.

Second, emotional integration is not explicitly modeled. Emotions involve complex interactions between bodily states, appraisals, and conscious feelings. How do affective processes fit into the representational hierarchy? Do they occupy specific levels, or do they modulate processing across all levels? The framework needs extension to address these questions.

Third, embodiment and sensorimotor grounding require further development. Our focus on internal representational levels somewhat neglects the role of bodily interaction with the environment. Enactive approaches to cognition emphasize that consciousness is not just internal integration but involves dynamic coupling between brain, body, and world. Future work should incorporate these insights.

Fourth, learning dynamics and gap function adaptation need elaboration. We have specified gap functions as static properties of cognitive architecture, but in reality, these functions change through development and learning. What are the mechanisms of this change? How do new levels emerge? How do gaps adjust in response to experience?

10.2 Future Theoretical Work

Several directions for theoretical extension suggest themselves.

First, we should extend the framework to multi-modal sensory integration. How are visual, auditory, tactile, and other sensory streams combined? Do they occupy parallel hierarchies that merge at higher levels, or are they integrated throughout? The mathematics of cross-modal binding could be formalized using tensor products or other algebraic structures.

Second, incorporating predictive processing and active inference would strengthen the framework. The brain is not just passively integrating information but actively predicting future states and selecting actions to minimize surprise. Our collapse mechanism could be extended to include forward prediction, with the selector choosing interpretations that best support accurate predictions.

Third, developing learning rules for gap function adaptation would make the framework more dynamic. We might model gap changes through Hebbian-like rules: frequently used transitions develop smaller gaps, while rarely used transitions develop larger gaps. This would explain expertise effects and cognitive development.

Fourth, formalizing relationships to quantum measurement theory could be productive. The collapse metaphor suggests quantum mechanics, and recent work explores whether quantum processes might play a role in consciousness. Our framework could be extended to incorporate genuine quantum effects if empirical evidence supports their involvement.

10.3 Future Empirical Work

On the empirical front, several research programs would test and refine the framework.

First, large-scale neuroimaging studies should test predictions about discrete capacity transitions. Using adaptive tasks that vary information load, we can measure mutual information between brain regions and look for predicted thresholds at 15, 20, 25, and 30 bits. These studies should include large samples to account for individual differences.

Second, developmental longitudinal studies should track gap function changes from childhood through adolescence. By testing children repeatedly over years, we can identify critical transition periods and characterize individual trajectories. This work would connect our framework to developmental psychology and education.

Third, comparative studies across psychiatric populations should characterize gap function alterations in different conditions. Schizophrenia, autism, depression, ADHD, and other conditions should show distinct patterns. This work would have clinical implications for diagnosis and treatment.

Fourth, artificial implementations using our computational model would test whether the framework produces realistic cognitive behavior. Building systems that implement hierarchical collapse with variable gaps would reveal hidden assumptions and generate new predictions. Such implementations could also have practical applications in AI.

11 Conclusion

We have presented a formal mathematical architecture that extends and unifies Global Workspace Theory and the LIDA cognitive model. The hierarchical adjoint-projection framework provides several key advances over previous approaches.

First, entropic capacity scaling through $I(n) = \kappa n \log n$ provides a principled account of representational capacity at each level. Information is measured in bits of Shannon entropy, capturing the uncertainty or information content of probability distributions rather than computational memory requirements. This scaling grows super-linearly to capture synergistic integration but remains sub-exponential to ensure computational feasibility.

Second, variable resource gaps through $f(n) \geq 1$ capture non-uniform cognitive allocation. Some transitions require small increments, others large jumps. This structure explains discrete developmental leaps, individual cognitive styles, task-specific resource recruitment, and pathological alterations in psychiatric conditions.

Third, adjoint operators $C \dashv P$ enforce cross-level consistency through Wasserstein-based metrics. The Wasserstein distance measures optimal transport cost between probability distributions, providing a geometry-sensitive measure of distributional similarity. This metric naturally captures the intuition that consistent interpretations fit together smoothly across levels.

Fourth, non-computable selection provides agency and top-down causation. The selector minimizes cross-level inconsistency, but computing this minimum requires evaluating Kolmogorov complexity, which is provably non-computable. This irreducible complexity provides a naturalistic foundation for consciousness as something beyond mere algorithmic computation.

Fifth, temporal dynamics distinguish collapse (50–100ms pre-report) from broadcast (at report). Integration precedes and causes global availability, not the reverse. This temporal structure makes specific predictions about neural timing that can be tested with MEG or high-density EEG.

The framework resolves several longstanding issues in consciousness science. The hard problem gains new perspective through non-computability: phenomenology emerges from irreducibly non-computable integration. Top-down causation becomes mathematically precise through consistency requirements between levels. Free will gains naturalistic grounding in the non-algorithmic nature of conscious selection. The binding problem dissolves as cross-level consistency creates unified representations.

Empirical predictions provide concrete paths for validation. Discrete capacity transitions at predicted thresholds, temporal precedence of collapse over broadcast, developmental signatures of variable gaps, and pathological alterations in psychiatric conditions are all testable with current technologies. If confirmed, these predictions would establish consciousness as a mathematically precise phenomenon emerging from hierarchical computational architecture.

Future work should extend the framework to emotional processing, embodied cognition, and learning dynamics while pursuing the outlined empirical predictions. The ultimate goal is a complete mathematical theory of consciousness grounded in information theory, computation theory, and empirical neuroscience.

The framework presented here represents a step toward that goal. By formalizing Global Workspace Theory with hierarchical structure, entropic scaling, adjoint operators, Wasserstein metrics, and non-computable selection, we transform a qualitative functional description into a quantitative predictive theory. Whether nature implements consciousness through these mechanisms remains an empirical question, but we now have a precise mathematical framework for asking that question.

References

Aaronson, S. (2013). *Quantum Computing Since Democritus.* Cambridge University Press.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness.* Cambridge University Press.

Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Progress in Brain Research*, 150:45–53.

Chaitin, G. J. (1987). *Algorithmic Information Theory.* Cambridge University Press.

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

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). Wiley-Interscience.

Dehaene, S. and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2):200–227.

Dehaene, S., Charles, L., King, J.-R., and Marti, S. (2014). Toward a computational theory of conscious processing. *Current Opinion in Neurobiology*, 25:76–84.

Dennett, D. C. (1991). *Consciousness Explained.* Little, Brown and Company.

Franklin, S. and Madl, T. (2014). The LIDA architecture: Adding new modes of learning to an intelligent, autonomous, software agent. *Integrated Design and Process Technology*, IDPT-2002.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.

Graziano, M. S. (2013). *Consciousness and the Social Brain.* Oxford University Press.

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience*, 17(5):307–321.

Li, M. and Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and Its Applications* (3rd ed.). Springer.

Mashour, G. A., Roelfsema, P., Changeux, J.-P., and Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, 105(5):776–798.

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLOS Computational Biology*, 10(5):e1003588.

Penrose, R. and Hameroff, S. (1996). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Mathematics and Computers in Simulation*, 40(3–4):453–480.

Seth, A. K. (2018). Consciousness: The last 50 years (and the next). *Brain and Neuroscience Advances*, 2:1–6.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423.

Shanahan, M. (2010). *Embodiment and the Inner Life: Cognition and Consciousness in the Space of Possible Minds.* Oxford University Press.

Sipser, M. (2012). *Introduction to the Theory of Computation* (3rd ed.). Cengage Learning.

Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7):450–461.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42(2):230–265.

Villani, C. (2009). *Optimal Transport: Old and New.* Springer.

A Mathematical Proofs

A.1 Proof of Entropic Scaling Optimality

Theorem A.1. *Among scaling functions that capture super-linear growth while remaining computationally feasible, $I(n) = \kappa n \log n$ emerges naturally from information-theoretic considerations of interacting components.*

Proof. Consider a system with n interacting components. If components are independent, total entropy is $H = n \cdot H_1$ where H_1 is the entropy of a single component. With pairwise interactions, we have $H \leq n \cdot H_1 + \binom{n}{2} \cdot I_2$ where I_2 represents typical mutual information between pairs.

For computational feasibility, we require sub-exponential growth. Consider all k-way interactions with weights $\alpha_k = \alpha_0/k!$ that decay factorially. The total entropy becomes:

$$H = \sum_{k=1}^{n} \binom{n}{k} \frac{\alpha_0}{k!} \approx \sum_{k=1}^{n} \frac{n^k}{k! \cdot k!} \alpha_0 \tag{14}$$

Using Stirling's approximation and dominated convergence, this sum converges to $\alpha_0 n \log n + O(n)$. Thus $I(n) = \kappa n \log n$ represents the leading-order term when higher-order interactions decay appropriately. □

A.2 Convergence of Variable Gap Sequences

Theorem A.2. *For any bounded gap function $f(n)$ with $\sum_{n=1}^{\infty} 1/f(n) = \infty$, the hierarchy $\{M_n\}$ achieves computational universality in the limit.*

Proof. Let L be any computable language. By computational universality of Turing machines, there exists some N such that a machine with N states can decide L.

For any starting level n_0, consider the sequence:

$$S = \{n_0, n_0 + f(n_0), n_0 + f(n_0) + f(n_0 + f(n_0)), \ldots\} \tag{15}$$

If $\sum 1/f(n) = \infty$, then $f(n)$ cannot grow faster than linearly on average. This implies the sequence S is unbounded: for any N, there exists some element of S exceeding N.

Therefore, the sequence eventually includes a level M_k with $k \geq N$, which can decide L. Since L was arbitrary, the hierarchy is computationally universal in the limit. □

B Experimental Design Details

B.1 Protocol for Testing Discrete Capacity Transitions

Materials:

- 64-channel EEG system with sampling rate $\geq$ 1000 Hz
- Visual n-back task with adaptive difficulty (n = 1 to 7)
- Information-theoretic analysis toolkit (custom MATLAB or Python scripts)
- Change-point detection algorithms (PELT, binary segmentation)

Procedure:

1. Participants ($N = 40$) perform adaptive n-back task with difficulty adjusted based on performance
2. Record continuous EEG throughout 45-minute session
3. Compute mutual information $I(X;Y)$ between frontal (Fz, F3, F4) and parietal (Pz, P3, P4) electrode clusters using time-frequency decomposition
4. Apply change-point detection to identify discrete transitions in mutual information
5. Compare detected transition points to predicted thresholds at $I = 15, 20, 25, 30$ bits
6. Correlate individual transition points with behavioral performance metrics

Expected Results: Sharp increases in mutual information at predicted thresholds, individual differences in exact transition points correlating with working memory capacity and task accuracy.

B.2 Protocol for Testing Collapse-Broadcast Temporal Order

Materials:

- MEG system with 306 sensors and source localization capability
- Binocular rivalry stimuli (orthogonal gratings at 45° and 135°)
- Response device for continuous perceptual reports
- Source reconstruction software (MNE-Python or FieldTrip)

Procedure:

1. Present rivalrous stimuli for 60-second blocks (10 blocks total)
2. Participants report perceptual state continuously via button press
3. Analyze MEG data time-locked to switch reports (transitions between percepts)
4. Identify local gamma synchrony (40–100 Hz) in visual cortex using beamformer source localization
5. Identify global P3 component (300–500 ms) in sensor space
6. Compare onset latencies: gamma in visual cortex versus P3 peak
7. Measure P3 amplitude as function of consistency differential between percepts

Expected Results: Local gamma synchrony onset at -100 to -50 ms relative to P3 peak, P3 amplitude correlating with perceptual dominance strength, source localization showing collapse initiating in visual cortex before spreading frontally.

C Computational Implementation Sketch

```python
import numpy as np
from scipy.stats import wasserstein_distance

class HierarchicalConsciousnessModel:
    def __init__(self, n_max=30, kappa=0.6, gamma=2.0, beta=20.0):
        """
        Initialize hierarchical consciousness model.

        Parameters:
        - n_max: Maximum level in hierarchy
        - kappa: Entropic scaling constant for I(n) = kappa * n * log(n)
        - gamma: Time cost per bit (ms/bit)
        - beta: Gap transition cost (ms/gap)
        """
        self.n_max = n_max
        self.kappa = kappa
        self.gamma = gamma
        self.beta = beta
        self.machines = self._initialize_machines()
        self.gap_function = self._initialize_gaps()

    def _initialize_machines(self):
        """Create hierarchy with entropic capacity scaling."""
        machines = []
        for n in range(1, self.n_max + 1):
            capacity = self.kappa * n * np.log(n + 1)  # Shannon entropy in bits
            machines.append({
                ’level’: n,
                ’capacity’: capacity,
                ’state’: None
            })
        return machines

    def _initialize_gaps(self):
        """Define variable gap function f(n)."""
        gaps = np.ones(self.n_max, dtype=int)
        # Critical developmental transitions
        gaps[7] = 3   # Age ˜7: concrete operations
        gaps[12] = 2  # Age ˜12: formal operations
        gaps[15] = 2  # Age ˜15: abstract reasoning
        gaps[20] = 2  # Executive function maturation
        return gaps

    def collapse(self, candidates):
        """
        Select winner minimizing cross-level inconsistency.

        Parameters:
        - candidates: List of candidate interpretations

        Returns:
        - winner: Candidate with minimal inconsistency
        - min_discrepancy: Inconsistency score
        """
        min_discrepancy = float(’inf’)
        winner = None

        for candidate in candidates:
            discrepancy = 0

            # Compute inconsistency across all level pairs
            for i in range(len(self.machines) - 1):
                n = self.machines[i][’level’]
                f_n = self.gap_function[i] if i < len(self.gap_function) else 1

                # Project to level i, reconstruct to level i+f(n)
                projected = self.project(candidate, n + f_n, n)
                reconstructed = self.reconstruct(projected, n, n +
                    f_n)

                # Measure Wasserstein distance
                w_dist = self.wasserstein_metric(
                    candidate['distributions'][n + f_n],
                    reconstructed['distribution']
                )
                discrepancy += w_dist

            if discrepancy < min_discrepancy:
                min_discrepancy = discrepancy
                winner = candidate

        return winner, min_discrepancy

    def project(self, state, from_level, to_level):
        """Project state from higher to lower level."""
        # Compress by selecting most probable features
        # This is a simplified implementation
        capacity_ratio = self.machines[to_level-1]['capacity'] / \
                        self.machines[from_level-1]['capacity']

        # Keep top features based on capacity ratio
        n_features = int(capacity_ratio * len(state['features']))
        projected_features = state['features'][:n_features]

        return {
            'level': to_level,
            'features': projected_features,
            'distribution': state['distributions'][to_level]
        }

    def reconstruct(self, state, from_level, to_level):
        """Reconstruct state from lower to higher level."""
        # Generate compatible higher-level states
        # Select one minimizing complexity (approximated by
           entropy)
        # This is a simplified implementation

        # Expand features based on capacity increase
        capacity_ratio = self.machines[to_level-1]['capacity'] / \
                        self.machines[from_level-1]['capacity']

        # Generate new features (simplified)
        reconstructed_features = state['features'] * int(
           capacity_ratio)

        return {
            'level': to_level,
```

```python
            'features': reconstructed_features,
            'distribution': np.random.dirichlet(np.ones(len(reconstructed_features)))
        }

    def wasserstein_metric(self, dist1, dist2):
        """Compute Wasserstein-2 distance between distributions."""
        return wasserstein_distance(dist1, dist2)

    def temporal_cost(self, level):
        """Compute processing time at given level (in ms)."""
        n = level
        capacity = self.kappa * n * np.log(n + 1)
        f_n = self.gap_function[n-1] if n <= len(self.gap_function) else 1

        T0 = 50  # Baseline processing time
        time = T0 + self.gamma * capacity + self.beta * f_n
        return time

    def cognitive_cycle(self, sensory_input):
        """
        Execute one complete cognitive cycle.

        Returns processing time and conscious content.
        """
        # Perception phase (low levels)
        t_perception = sum(self.temporal_cost(i) for i in range(1, 11))
        perceptual_candidates = self.generate_perceptual_candidates(sensory_input)

        # Understanding phase (semantic levels)
        t_understanding = sum(self.temporal_cost(i) for i in range(15, 26))
        semantic_candidates = self.generate_semantic_candidates(perceptual_candidates)

        # Collapse phase (consistency minimization)
        t_collapse_start = 50
        winner, inconsistency = self.collapse(semantic_candidates)
        t_collapse = t_collapse_start + 50  # 50-100ms

        # Broadcast phase (rapid distribution)
        t_broadcast = 25

        # Action selection
        t_action = sum(self.temporal_cost(i) for i in range(20, 31))

        total_time = t_perception + t_collapse + t_broadcast + 50

        return {
            'conscious_content': winner,
            'cycle_time': total_time,
            'inconsistency': inconsistency,
            'phase_times': {
                'perception': t_perception,
                'collapse': t_collapse,
                'broadcast': t_broadcast,
                'action': t_action
            }
        }
```

This implementation sketch illustrates key computational components while remaining tractable for simulation studies. The actual implementation would require more sophisticated probability distribution handling, proper Wasserstein distance computation using optimal transport algorithms, and realistic models of perceptual and semantic processing.