The Mathematical Foundations of Constrained Object Hierarchies-A Universal Framework for World Modeling, General Intelligence and Agentic Systems

Harris Wang *

Posted Date: 11 February 2026

doi: 10.20944/preprints202602.0521.v2

Keywords: artificial general intelligence; constrained object hierarchies; neuroscience-grounded AI; category theory; dynamical systems; hybrid intelligent systems; constraint satisfaction; hierarchical learning

 doi:10.20944/preprints202602.0521.v2

Article

The Mathematical Foundations of Constrained Object Hierarchies-A Universal Framework for World Modeling, General Intelligence and Agentic Systems

Harris Wang

School of Computing and Information Systems, Athabasca University, Athabasca, Alberta, Canada; harrisw@athabascau.ca

Abstract

Constrained Object Hierarchies (COH) offer a unified theoretical foundation for artificial general intelligence (AGI), rooted in neuroscience principles and developed with full mathematical rigor. This paper presents the complete formalization of COH theory, showing how intelligence emerges from hierarchically composed structures governed by adaptive optimization constraints. We introduce precise definitions, establish core properties of soundness and completeness, and situate COH within established mathematical frameworks including category theory, dynamical systems, and information theory. Building on this foundation, we prove three central theorems that demonstrate COH's practical significance for AGI: guaranteeing high-fidelity world modeling, preventing jagged or non-smooth intelligence behaviors, and enabling the construction of coherent agentic systems. These results provide quantitative bounds on representational accuracy, generalization performance, and decision-making complexity. Collectively, the findings show that COH delivers a mathematically rigorous, interpretable, and scalable basis for modeling intelligent systems across six heterogeneous domains, while preserving the flexibility required for general intelligence and ensuring explicit guarantees for safety and transparency.

Keywords: artificial general intelligence; constrained object hierarchies; neuroscience-grounded AI; category theory; dynamical systems; hybrid intelligent systems; constraint satisfaction; hierarchical learning

1. Introduction

Artificial General Intelligence (AGI) research seeks a principled account of how intelligent behavior can emerge across diverse tasks, environments, and representational settings. Although significant advances have been made in symbolic reasoning, neural computation, reinforcement learning, cognitive architectures, and probabilistic inference, existing approaches remain fragmented. Each captures an important facet of intelligence—structure, learning, prediction, or control—but none offers a unified mathematical framework capable of integrating these facets into a coherent theory of general intelligence.

Constrained Object Hierarchies (COH) (Wang, 2025) address this gap by grounding intelligence in three neuroscience-inspired principles: hierarchical organization, predictive optimization under uncertainty, and constraint-driven regulation, which together support both stability and adaptive behavior. Contemporary research in cortical hierarchy, predictive coding, and homeostatic regulation (Rao et al, 2024; Miyashita, 2024; Lin et al, 2025) supports the view that intelligent systems can be modeled as structured, adaptive hierarchies whose behavior emerges from interactions among components, goals, and constraints.

 doi:10.20944/preprints202602.0521.v2

COH formalizes these principles through a minimal, orthogonal 9-tuple that defines components, attributes, methods, neural learning mechanisms, goals, constraints, daemons, and semantic embeddings within a single mathematical structure. This paper develops the complete theoretical foundation of COH, establishes its internal consistency, and demonstrates its expressive power across multiple scientific domains. In addition, we extend the core framework with three novel theoretical contributions: (1) quantitative bounds on world model fidelity, (2) formal guarantees against jagged intelligence through constraint-aligned generalization, and (3) constructive proofs for building verifiably safe agentic systems with polynomial-time decision complexity.

Contributions

This work makes the following key contributions:

- A complete mathematical formalization of Constrained Object Hierarchies (COH) as a minimal and orthogonal 9-tuple specifying hierarchical structure, semantic composition, constraint systems, adaptive optimization, and learning mechanisms.
- A unified category-theoretic foundation, demonstrating that COH supports products, coproducts, morphisms, functorial mappings, and limit constructions required for rigorous compositional reasoning.
- A constrained hybrid dynamical-systems formulation, establishing stability properties, admissible constraint regions, and connections to information-theoretic principles within a coherent dynamical interpretation.
- A unified learning and adaptation theory, showing that COH functions as both a universal hierarchical approximation architecture and a constraint-regulated optimization system capable of supporting reinforcement learning and hierarchical credit assignment.
- Formal guarantees of soundness, completeness, expressive minimality, and orthogonality, proving that COH is internally consistent, irreducible, and suitable as a general mathematical basis for intelligence.
- Three AGI-relevant theorems providing quantitative guarantees:
  (a) bounds on world-model fidelity as a function of constraint expressiveness;
  (b) cross-hierarchical smoothness conditions that prevent jagged intelligence behaviors; and
  (c) constructive methods for building coherent agentic systems with bounded decision complexity and verifiable safety properties.
- Cross-domain demonstrations across six heterogeneous scientific and engineering systems, showing that COH supports stable modeling, interpretability, and simulatability across physical, biological, cyber-physical, economic, and cultural-computational domains.
- Integration with major theories of intelligence, including hierarchical reinforcement learning, cognitive architectures, category-theoretic compositionality, and active inference, positioning COH as a unified and mathematically rigorous foundation for AGI.

Paper Organization

Section 2 reviews prior work across neuroscience, mathematical modeling, constrained optimization, and AGI research. Section 3 formalizes the COH 9-tuple. Section 4 develops its category-theoretic foundations, and Section 5 presents the dynamical-systems perspective. Section 6 introduces learning and adaptation mechanisms. Section 7 establishes soundness, completeness, and minimality. Section 8 applies COH to six scientific domains. Section 9 relates COH to existing theories of intelligence, Section 10 discusses implications for AGI development, and Section 11 concludes.

2. Literature Review

The pursuit of Artificial General Intelligence (AGI) necessitates a formal framework that integrates structural composition, adaptive learning, and constraint satisfaction. While existing paradigms offer valuable insights, they remain fragmented. This review synthesizes foundational work from cognitive neuroscience, mathematical frameworks for intelligence, and computational

 doi:10.20944/preprints202602.0521.v2

models of constrained systems, establishing the intellectual context and gap that Constrained Object Hierarchies (COH) theory aims to fill.

2.1. Neuroscience Foundations for Hierarchical Intelligence

A core inspiration for structurally grounded AGI comes from principles of neural organization. The hierarchical processing observed in the primate visual cortex provides a biological blueprint for compositional intelligence regulation (Rao et al, 2024; Miyashita, 2024; Lin et al, 2025). These three pillars—hierarchy, predictive optimization, and constraint satisfaction—form the neuroscientific cornerstone of the COH framework.

2.2. Mathematical and Theoretical Frameworks for Intelligence

Category theory provides a powerful language for describing compositional structures and relationships, relevant for defining COH morphisms (Fong and Spivak, 2019). However, these mathematical tools have rarely been integrated into a unified theory of general intelligence.

2.3. World Modeling and Representation Learning

The problem of learning accurate world models from data has seen significant advances in deep learning, with architectures like transformers demonstrating remarkable sequence modeling capabilities (Vaswani et al. 2017). However, current approaches lack formal guarantees on model fidelity and often fail to capture hierarchical structure explicitly. Research on conceptual spaces (Gärdenfors, 2014) and semantic embeddings provides foundations for the COH embedding component E, but typically without the constraint-driven semantics that COH provides for ensuring model-world alignment.

2.4. Generalization and Jagged Intelligence

The problem of robust generalization—where systems perform well on training distributions but fail on distributionally shifted tasks—has been identified as "jagged intelligence" in recent AI safety literature (Hendrycks et al., 2022). Theoretical work on out-of-distribution generalization (Arjovsky et al. 2019) and domain adaptation provides mathematical frameworks, but these typically lack the compositional, constraint-based approach that COH employs to ensure smooth capability generalization across task spaces.

2.5. Agentic Systems and Safe Autonomy

Research on autonomous agents spans cognitive architectures (Miyashita, 2024), hierarchical reinforcement learning (Pateria et al. 2024), and constrained decision-making (Altman 1999). While these provide important components for agent design, they lack a unified mathematical foundation that simultaneously addresses perception, action, goals, and ethics within a single formalism with complexity guarantees. Recent work on AI safety emphasizes constraint-based approaches (Amodei et al. 2016), but typically without the hierarchical composition and formal guarantees that COH provides.

2.6. The Role of Constraints in Intelligent Systems

Classical AI models constraints through CSPs (Russell & Norvig 2021), while constrained optimization frameworks such as CMDPs (Altman 1999) introduce bounds on expected costs. Neuroscience suggests constraints are fundamental—from homeostatic limits to syntactic structures—supporting the COH view that constraints must be first-class citizens.

 doi:10.20944/preprints202602.0521.v2

2.7. Synthesis and Gap Identification

Prior work provides essential but isolated components: neuroscientific principles of hierarchy and prediction, mathematical tools for composition and dynamics, computational models for learning and planning, and formalisms for constraint handling. The key gap is a unified, mathematically rigorous framework that cohesively integrates these elements. COH theory addresses this by proposing a minimal, orthogonal 9-tuple formalism that explicitly defines hierarchical objects (C), their properties (A), behaviors (M), and learning capacities (N), all under the governance of a multi-tiered constraint system (I, T, G) maintained by active processes (D), with a unifying semantic embedding (E). This bridges the expressiveness of symbolic systems, the adaptability of neural networks, and the principled governance of constraint-based reasoning, offering a foundational model for AGI.

3. Formal Definitions of COH Components

3.1. The 9-Tuple Formalism

Definition 3.1 (Constrained Object Hierarchy). A COH is defined as a 9-tuple:

$$\mathcal{O} = (C, A, M, N, E, I, T, G, D)$$

where each component has specific mathematical structure.

3.2. Component Formalization

Definition 3.2 (Components). Let **Obj** be the category of COH objects. Then:

$$C: \mathcal{O} \to \mathcal{P}(\mathbf{Obj})$$

where $\mathcal{P}$ denotes the power set, with the additional constraint that the component relation forms a directed acyclic graph (DAG).

Definition 3.3 (Attributes). Attributes are state variables:

$$A = \{a_i: \mathcal{O} \to \mathcal{V}_i\}_{i=1}^{n}$$

where each a_i maps the object to a value in domain $\mathcal{V}_i$, forming a measurable space $(\Omega, \mathcal{F})$.

Definition 3.4 (Methods). Methods are executable transformations:

$$M = \{m_j: \mathcal{S} \times \mathcal{A} \to \mathcal{S} \times \mathcal{R}\}_{j=1}^{k}$$

where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, and $\mathcal{R}$ is the reward/outcome space.

Definition 3.5 (Neural Components). Neural components are parameterized functions:

$$N = \{f_\theta: \mathcal{X} \to \mathcal{Y} \mid \theta \in \Theta\}$$

where each f_θ is a measurable function parameterized by θ in parameter space Θ.

Definition 3.6 (Embedding Component). The embedding function:

$$E: \mathcal{O} \to \mathcal{H}$$

maps the entire object to a Hilbert space $\mathcal{H}$, providing a unified semantic representation.

Definition 3.7 (Identity Constraints). Identity constraints are invariant conditions:

$$I = \{\phi_i: \mathcal{S} \to \mathbb{B}\}_{i=1}^{p}$$

where each ϕ_i is a predicate that must evaluate to true for all reachable states.

Definition 3.8 (Trigger Constraints). Trigger constraints are event-condition-action rules:

$$T = \{(e_i, c_i, a_i) \mid e_i \in \mathcal{E}, c_i: \mathcal{S} \to \mathbb{B}, a_i \in \mathcal{A}\}$$

where $\mathcal{E}$ is an event space.

Definition 3.9 (Goal Constraints). Goal constraints are optimization objectives:

$$G = \{g_j: \mathcal{S} \times \mathcal{A} \times \mathcal{T} \to \mathbb{R}\}_{j=1}^{q}$$

 doi:10.20944/preprints202602.0521.v2

where $\mathcal{T}$ is a time domain, and goals may be composed as:

$$G_{\text{total}} = \sum_{j=1}^{q} w_j \, g_j + \lambda R(\theta)$$

with regularization term $R(\theta)$.

Definition 3.10 (Constraint Daemons). Constraint daemons are continuous monitoring processes:

$$D = \{d_k : \mathcal{S}^{\mathcal{T}} \to \mathcal{A}\}_{k=1}^{r}$$

where each d_k operates on the state trajectory.

4. Category-Theoretic Foundations

4.1. COH as a Category

Definition 4.1 (Category of COH Objects). Let $\mathbf{COH}$ be the category where:

- Objects are COH 9-tuples $\mathcal{O}$
- Morphisms $f : \mathcal{O}_1 \to \mathcal{O}_2$ are structure-preserving maps that respect the hierarchical composition and constraints

Theorem 4.1 (Compositionality). The category $\mathbf{COH}$ has:

1. Finite products: $\mathcal{O}_1 \times \mathcal{O}_2$ representing parallel composition
2. Finite coproducts: $\mathcal{O}_1 + \mathcal{O}_2$ representing alternative compositions
3. Exponentials: $\mathcal{O}_2^{\mathcal{O}_1}$ representing function objects between COHs

Proof sketch[1]: The product COH is constructed by taking Cartesian products of each component set, with appropriate constraint conjunction. The coproduct takes disjoint unions with constraint disjunction. Exponentials are constructed from method spaces with appropriate currying.

4.2. Hierarchical Structure as a Functor

Definition 4.2 (Component Functor). The component structure defines a functor:

$$F : \mathbf{COH} \to \mathbf{Graph}$$

where $\mathbf{Graph}$ is the category of directed graphs, mapping each COH to its component DAG.

Theorem 4.2 (Hierarchy Preservation). For any morphism $f : \mathcal{O}_1 \to \mathcal{O}_2$ in $\mathbf{COH}$, there exists a graph homomorphism $F(f) : F(\mathcal{O}_1) \to F(\mathcal{O}_2)$ that preserves the hierarchical structure.

4.3. Constraint Systems as Limits

Definition 4.3 (Constraint Category). For a COH $\mathcal{O}$, define the constraint category $\mathbf{Con}(\mathcal{O})$ where:

- Objects are states $s \in \mathcal{S}$ satisfying all identity constraints I
- Morphisms are transitions generated by methods M that respect trigger constraints T

Theorem 4.3 (Constraint Satisfaction as Limit). The set of admissible states forms a limit in the diagram formed by constraint predicates:

$$\mathcal{S}_{\text{adm}} = \lim \left(\{\phi_i^{-1}(\text{true})\}_{i=1}^{p}\right)$$

[1] For complete proofs of all theorems presented in the paper, see Supplementary Appendix B.

 doi:10.20944/preprints202602.0521.v2

5. Dynamical Systems Perspective

5.1. COH as a Hybrid Dynamical System

Definition 5.1 (COH Dynamics). A COH defines a hybrid dynamical system:

$$\frac{ds}{dt} = f(s, a, \theta) \text{for continuous evolution}$$
$$s_{t+1} = g(s_t, a_t, \theta) \text{for discrete transitions}$$

with switching conditions defined by trigger constraints T.

Theorem 5.1 (Existence and Uniqueness). Under Lipschitz continuity assumptions on f and g, and measurability of trigger conditions, the COH dynamics admit unique solutions for given initial conditions and control policies.

5.2. Stability Analysis

Definition 5.2 (Constraint-Admissible Region). Let:

$$\mathcal{R} = \{s \in \mathcal{S} \mid \forall \phi \in I, \phi(s) = \text{true}\}$$

be the region satisfying identity constraints.

Theorem 5.2 (Constraint Stability). If the goal constraints G define a Lyapunov function $V(s)$ that decreases along trajectories within $\mathcal{R}$, then the system is asymptotically stable within the constraint-admissible region.

Proof: Standard Lyapunov theory applied to the restricted dynamics on $\mathcal{R}$, with daemons D ensuring boundary conditions are respected.

5.3. Information-Theoretic Perspective

Definition 5.3 (Intelligence as Constrained Optimization). The intelligence of a COH can be quantified as:

$$\mathcal{I}(\mathcal{O}) = \max_{\pi} \mathbb{E}\left[\sum_t \gamma^t r_t\right] - \lambda \cdot \mathrm{KL}(p_\pi \parallel p_{\text{constraints}})$$

where p_π is the policy distribution and $p_{\text{constraints}}$ encodes constraint satisfaction probabilities.

Theorem 5.3 (Free Energy Principle Correspondence). The COH optimization objective is equivalent to minimizing variational free energy under the interpretation of constraints as priors in a Bayesian model.

6. Learning and Adaptation Theory

6.1. Neural Components as Universal Approximators

Theorem 6.1 (Universal Approximation). For any measurable function $h: \mathcal{X} \to \mathcal{Y}$ and any $\epsilon > 0$, there exists a neural component $f_\theta \in N$ such that:

$$\| f_\theta(x) - h(x) \| < \epsilon \text{ almost everywhere}$$

provided the neural architecture has sufficient capacity.

Proof: Universal approximation results (Cybenko 1989; Hornik et al. 1989; Yarotsky 2017; Yun et al. 2020) ensure that neural components in COH can approximate arbitrary measurable functions.

6.2. Constrained Reinforcement Learning

Definition 6.1 (Constrained Policy Optimization). The learning problem for a COH is:

$$\max_{\pi_\theta} \mathbb{E}_{\tau \sim \pi_\theta}[G(\tau)] \text{subject to} \mathbb{E}_{\tau \sim \pi_\theta}[C_i(\tau)] \leq 0, \forall i$$

where C_i are constraint violation measures. The constrained optimization formulation aligns with classical CMDP theory (Altman 1999).

 doi:10.20944/preprints202602.0521.v2

Theorem 6.2 (Convergence of Constrained Learning). Under appropriate conditions (convex constraint functions, Lipschitz continuous gradients), the constrained optimization admits solutions that can be found via primal-dual methods.

6.3. Hierarchical Credit Assignment

Definition 6.2 (Hierarchical Value Decomposition). The value function for a COH decomposes as:

$$V(\mathcal{O}) = \sum_{\mathcal{O}' \in C(\mathcal{O})} w_{\mathcal{O}'} V(\mathcal{O}') + V_{\text{coordination}}(\mathcal{O})$$

where $V_{\text{coordination}}$ captures emergent properties from component interactions. The hierarchical value decomposition is consistent with multi-level RL formulations (Pateria et al. 2022).

Theorem 6.3 (Optimal Decomposition). There exists an optimal weight assignment $\{w_{\mathcal{O}'}\}$ that satisfies Bellman optimality while respecting constraint hierarchies.

7. Soundness and Completeness Proofs

7.1. Soundness: Internal Consistency

Definition 7.1 (Soundness). A COH theory is sound if:

1. All derivable statements are true in all models
2. The constraint system is satisfiable
3. The hierarchy is well-founded (no circular dependencies)

Theorem 7.1 (Constraint Soundness). If all identity constraints I are logically consistent (their conjunction is satisfiable), and trigger constraints T preserve this consistency, then the COH is sound.

Proof: By induction on the derivation length. Base case: Initial states satisfy I by definition. Inductive step: Each method application respects T, which by design preserves constraint satisfaction.

Theorem 7.2 (Hierarchical Soundness). The component DAG is well-founded (contains no cycles) iff the COH has a well-defined semantics.

Proof: Acyclicity ensures that component evaluation terminates, as each component can only depend on lower-level components. This is provable by topological ordering of the DAG.

7.2. Completeness: Expressive Power

Definition 7.2 (Completeness). A COH theory is complete if it can represent:

1. Any computable function (Turing completeness)
2. Any learning system with bounded rationality
3. Any constraint-satisfaction system with finite resources

Theorem 7.3 (Computational Completeness). The COH framework is Turing-complete.

Proof sketch: We can encode a Universal Turing Machine in a COH:

- C: Tape cells as sub-objects
- A: State and head position
- M: Transition function as methods
- I: Tape consistency constraints
- T: Step execution triggers
- N: Optional learning components
- The embedding E can encode the complete machine state

 doi:10.20944/preprints202602.0521.v2

Theorem 7.4 (Representational Completeness). For any intelligent system S with finite description, there exists a COH $\mathcal{O}$ that simulates S.

Proof: By structural induction on the components of S. Base components map to atomic COH objects. Composite systems map to hierarchical compositions. Learning components map to neural components with appropriate architectures.

Theorem 7.5 (Constraint Expressiveness). Any constraint system expressible in first-order logic with transitive closure can be encoded in COH constraints.

Proof: Identity constraints I capture universal quantifiers. Trigger constraints T capture implication. Hierarchical composition enables transitive relationships through component nesting.

7.3. Minimality and Orthogonality

Theorem 7.6 (Component Minimality). The 9-tuple is minimal: removing any component reduces expressive power.

Proof: For each component, we demonstrate a class of intelligent systems that cannot be represented without it:

- Without C: Cannot represent hierarchical systems
- Without A: Cannot maintain state
- Without M: Cannot perform actions
- Without N: Cannot learn adaptively
- Without E: Cannot have unified semantics
- Without I: Cannot maintain identity
- Without T: Cannot respond to events
- Without G: Cannot optimize behavior
- Without D: Cannot monitor constraints in real-time

Theorem 7.7 (Orthogonality). The nine components are orthogonal: no component can be expressed purely in terms of the others.

Proof: By demonstrating independence through information-theoretic measures of component contributions to system behavior.

8. Applications Across Six Domains

8.1. Introduction

A central claim of Constrained Object Hierarchies (COH) is that a single, mathematically grounded framework can model intelligent behavior across widely different domains. To demonstrate this generality, we apply the COH 9-tuple formalism to six complex systems: quantum computing, systems biology, smart-city infrastructure, aerospace mission planning, financial markets, and cultural heritage preservation. These domains span physical, biological, sociotechnical, and cultural systems, each with distinct constraints, dynamics, and optimization objectives. Despite their differences, all six applications instantiate the same COH structure, showing that COH provides a unified, domain-agnostic foundation for modeling intelligent systems. Before presenting these applications, it is necessary to discuss the implementability and simulability of the COH formalizations.

 doi:10.20944/preprints202602.0521.v2

8.1.1. Implementability of COH

Theorem 8.1.1 (COH Implementability). For any well-formed COH specification $\mathcal{O} = (C, A, M, N, E, I, T, G, D)$, there exists a constructive implementation $\mathcal{K}(\mathcal{O})$ in a universal computing model (Turing machine, lambda calculus, or physical computer) with the following properties:

1. **Semantic Preservation:** For every admissible state transition $s \xrightarrow{m} s'$ in $\mathcal{O}$, there exists a corresponding computational step in $\mathcal{I}(\mathcal{O})$ producing equivalent results.
2. **Constraint Enforcement:** All identity constraints I are maintained invariantly, trigger constraints T are evaluated before every relevant transition, and daemons D execute as continuous processes.
3. **Resource Bounds:** The implementation requires resources polynomial in the specification size:

$$\text{Time}(\mathcal{I}) = O\big(|\ C\ |^{k} \cdot \text{poly}(|\ I\ |, |\ T\ |, |\ M\ |)\big) \text{Space}(\mathcal{I}) = O\left(\sum_{i=1}^{n} |\ \mathcal{V}_i\ | + |\ \Theta\ | + \text{polylog}(|\ \mathcal{H}\ |)\right)$$

where k depends on the DAG depth of C, $|\ \mathcal{V}_i\ |$ are attribute domain sizes, $|\ \Theta\ |$ is neural parameter size, and $|\ \mathcal{H}\ |$ is embedding space dimension.

1. **Approximation Quality:** For neural components N, the implementation achieves approximation error bounded by universal approximation theorems, with quantization effects bounded by $\epsilon_{\text{quant}} = O(1/\text{bits})$.

Proof sketch: the proof takes 6 steps involving constructive implementation schema, semantic preservation, constraint enforcement, resource bounds analysis, approximation quality, proof of computational universality. the details can be found in supplementary appendix B.

8.1.2. Simulability of COH

The theoretical breadth of the COH framework, illustrated through its application across six distinct domains, must ultimately be supported by empirical validation to demonstrate practical utility. Because COH is proposed as a universal framework for world modelling and general intelligence, exhaustive real-world validation across all possible domains is neither feasible nor conceptually required. Simulation therefore becomes the primary methodological bridge between theory and practice, showing that COH models function as computational systems rather than purely abstract constructs. The Simulatability Theorem guarantees that every COH model admits a faithful computational realization, establishing simulation as a principled and sufficient validation pathway.

Theorem 8.1.1 (Simulatability). Any COH model can be simulated to arbitrary precision given sufficient computational resources.

Proof: Follows from the constructive nature of COH definitions and the computability of each component.

8.2. Quantum Computing Control Systems

8.2.1. Domain Overview

Quantum computing requires precise manipulation of quantum states under strict physical constraints such as unitarity, decoherence, and error-correction thresholds. The domain combines quantum mechanics, classical control theory, and real-time optimization—making it an ideal testbed for COH.

 doi:10.20944/preprints202602.0521.v2

8.2.2. COH Formalization[2]

$COH_{quantum}$ = (C, A, M, N, E, I, T, G, D)

C — Components
QubitArray, ControlElectronics, ErrorCorrectionUnit, Compiler

A — Attributes
qubit_states ∈ $\mathbb{C}^{2^n}$, decoherence_rates ∈ $\mathbb{R}^n$, gate_fidelities ∈ [0, 1]

M — Methods
apply_gate(gate_matrix), measure_qubit(index), initialize_state(state)

N — Neural Components
ErrorPredictionNN, DecoherenceCompensator, OptimalControlRL

E — Embedding
QuantumStateEmbedder: $\mathfrak{H} \to \mathbb{R}^{512}$
(maps high-dimensional Hilbert-space quantum states to a classical embedding)

I — Constraints
Entanglement limit: For all t, qubit_entanglement ≤ max_entanglement
Gate-error threshold: gate_errors < threshold

T — Triggers
If decoherence_rate > threshold → apply_dynamic_decoupling

G — Goals
Maximize algorithm_fidelity, Minimize error_correction_overhead

D — Daemons
CoherenceMonitor, ErrorSyndromeDaemon

8.2.3. Mathematical Guarantees

Theorem 8.2.1 (Quantum Correctness). For any quantum circuit C with n gates, the COH-controlled quantum system satisfies:

$$\text{fidelity}(Q(C), C_{\text{ideal}}) \geq 1 - n\varepsilon_{\text{gate}} - \varepsilon_{\text{meas}} - \varepsilon_{\text{decoh}},$$

where each error term is bounded by identity constraints and regulated by daemons.

8.3. Biological System Simulation

8.3.1. Domain Overview

Systems biology involves multi-scale modeling of molecular networks, cells, tissues, and organs. These systems exhibit nonlinear dynamics, stochasticity, and complex regulatory constraints.

8.3.2. COH Formalization

COH_{bio} = (C, A, M, N, E, I, T, G, D)

C — Components
MolecularNetwork, CellPopulation, TissueStructure, OrganSystem

A — Attributes
gene_expression ∈ $\mathbb{R}^m$, metabolite_concentrations ∈ $\mathbb{R}^n$, cell_states ∈ Enum^k

M — Methods
simulate_reaction_network(t), apply_drug_concentration(drug, c), mutate_gene(gene)

N — Neural Components
GeneRegulationPredictor, MetabolicFluxEstimator, CellFateClassifier

E — Embedding
BiologicalStateEncoder: multi-omic data → $\mathbb{R}^{1024}$ unified biological state representation

[2] Supplementary Appendix A outlines guidelines for translating COH formalizations of complex world systems into executable models suitable for real-world deployment or simulation.

 doi:10.20944/preprints202602.0521.v2

I — Constraints

Mass_**balance:** For all metabolites: production = consumption + accumulation

Energy_**conservation:** ATP_production ≥ cellular_maintenance

T — Triggers

If nutrient_depletion → activate_starvation_response

If DNA_damage_detected → activate_repair_pathways

G — Goals

Maximize model_accuracy(experimental_data), Minimize computational_cost

D — Daemons

HomeostasisDaemon, EnergyBalanceMonitor, ToxicityPredictor

8.3.3. Mathematical Guarantees

Theorem 8.3.1 (Biological Plausibility). For any dataset D, the COH biological model satisfies:

$$\mathrm{KL}(M_{\mathrm{pred}} \parallel D_{\mathrm{obs}}) \leq \varepsilon_{\mathrm{exp}} + \varepsilon_{\mathrm{model}} + \varepsilon_{\mathrm{stoch}},$$

with each term bounded by identity constraints.

8.4. Smart-City Infrastructure Management

8.4.1. Domain Overview

Smart cities integrate energy, transportation, water, and communication systems. These infrastructures require real-time optimization, resilience, and equitable service distribution.

8.4.2. COH Formalization

$\mathrm{COH}_{\mathrm{smartcity}}$ = (C, A, M, N, E, I, T, G, D)

C — Components

EnergyGrid, TransportationNetwork, WaterSystem, CommunicationInfrastructure

A — Attributes

energy_demand: TimeSeries, traffic_flow: Graph → ℝ, water_levels: SpatialField

M — Methods

route_vehicles(source, dest), allocate_power(region, amount), manage_wastewater()

N — Neural Components

DemandForecaster: Transformer, TrafficOptimizer: GNN, ResourceAllocator: MARL (Multi-Agent Reinforcement Learning)

E — Embedding

CityStateEncoder**:** multi-modal urban data → $\mathbb{R}^{2048}$ unified representation

I — Constraints

Flow conservation (∀ network): Σ inflows = Σ outflows

Utility service guarantee (∀ utility): service_level ≥ SLA_minimum

T — Triggers

If traffic_congestion > threshold → activate_alternate_routes

If power_outage_detected → reroute_energy_supply

G — Goals

Minimize carbon_emissions, Maximize resource_utilization_efficiency, Maximize citizen_quality_of_life

D — Daemons

ResilienceMonitor, SustainabilityDaemon, EquityEvaluator

8.4.3. Mathematical Guarantees

Theorem 8.4.1 (Urban Efficiency). The COH-managed city achieves Pareto-optimal solutions to multi-objective optimization problems:

 doi:10.20944/preprints202602.0521.v2

$$\min_{x \in X} [f_1(x), \ldots, f_k(x)]^T,$$

where X is the constraint-satisfying region.

8.5. Aerospace Mission Planning

8.5.1. Domain Overview

Aerospace missions involve hybrid dynamics, strict safety constraints, and long-horizon planning under uncertainty.

8.5.2. COH Formalization

COH_{space} = (C, A, M, N, E, I, T, G, D)

C — Components
OrbitalDynamics, PropulsionSystem, PowerManagement, ScientificInstruments

A — Attributes
position ∈ $\mathbb{R}^3$, velocity ∈ $\mathbb{R}^3$, fuel_mass ∈ $\mathbb{R}$, instrument_health ∈ $[0, 1]^m$

M — Methods
execute_orbital_maneuver(Δv), acquire_science_data(target), communicate_ground()

N — Neural Components
TrajectoryOptimizer: NeuralODE, AnomalyDetector: Autoencoder, ResourcePredictor: LSTM

E — Embedding
SpacecraftStateEncoder: combines orbital elements, resource states, and mission objectives.

I — Constraints
Orbital_energy_conservation: $E = v^2/2 - \mu/r$
Mass_balance: $m_{final} = m_{initial} - \int \dot{m}\, dt$

T — Triggers
If radiation_level > safe_limit → shield_instruments
If communication_loss > duration → enter_autonomous_mode

G — Goals
Maximize scientific_return
Minimize fuel_consumption
Ensure mission_success_probability > 0.99

D — Daemons
RadiationMonitor
SystemHealthDaemon
ContingencyPlanner

8.5.3. Mathematical Guarantees

Theorem 8.5.1 (Mission Safety). The probability of mission failure satisfies:

$$P(\text{failure}) \leq \sum_i \lambda_i T + \sum_j \int_0^T h_j(t)\, dt,$$

with hazard rates bounded by constraints and mitigated by daemons.

8.6. Financial Market Ecosystem Simulation

8.6.1. Domain Overview

Financial markets involve strategic agents, regulatory constraints, and systemic risk propagation.

 doi:10.20944/preprints202602.0521.v2

8.6.2. COH Formalization

$COH_{finance}$ = (C, A, M, N, E, I, T, G, D)

C — Components

InvestorPopulation, MarketMakerNetwork, ClearingHouse, RegulatoryBody

A — Attributes

asset_prices ∈ $\mathbb{R}^m$, portfolio_holdings ∈ $\mathbb{R}^{n \times m}$, market_microstructure = OrderBook

M — Methods

submit_order(asset, quantity, type), clear_market(), adjust_regulations(params)

N — Neural Components

PricePredictor: Transformer, AgentBehaviorModel: IRL, SystemicRiskDetector: GAN

E — Embedding

MarketStateEncoder**:** combines prices, volumes, sentiments, and network structures.

I — Constraints

No_arbitrage**:** For all portfolios, expected_return ≥ risk_free_rate + β · market_premium

Market_clearing**:** Σ buy_orders = Σ sell_orders at equilibrium_price

T — Triggers

If volatility > VIX_threshold → activate_circuit_breakers

If liquidity < minimum → inject_market_making

G — Goals

Maximize market_efficiency: 1 − |fundamental_value − market_price| / fundamental_value

Minimize systemic_risk: max_correlation_during_crisis

D — Daemons

FlashCrashDetector, ManipulationMonitor, LiquidityDaemon

8.6.3. Mathematical Guarantees

Theorem 8.6.1 (Market Stability). The COH-simulated market achieves efficient price discovery:

$$\text{Efficiency} = 1 - O\left(\frac{1}{\sqrt{N_{\text{informed}}}}\right),$$

and market impact and stability results align with empirical findings on the square-root law (Tóth et al. 2016; Benzaquen & Bouchaud 2018; Bucci et al. 2019).

8.7. Cultural Heritage Preservation

8.7.1. Domain Overview

Cultural heritage preservation requires balancing conservation, access, authenticity, and long-term risk management.

8.7.2. COH Formalization

$COH_{cultural\_heritage}$ = (C, A, M, N, E, I, T, G, D)

C — Components / Sub-objects

ArtifactScanner, MaterialAnalyzer , HistoricalContextDB, ConservationPlanner

A — Attributes / State Variables

artifact_state: S ∈ ImageSpace × MaterialSpace × StructuralSpace

material_composition: C ∈ $\mathbb{R}^{elements}$ (from spectroscopic analysis)

deterioration_rate: $\lambda \in \mathbb{R}^+$

Arrhenius form: $\lambda = A \cdot \exp(-E_a / (R \cdot T))$

historical_significance: σ ∈ [0, 1] (rarity × cultural_value × condition)

M — Methods / Executable Actions

digitize_artifact(resolution: {standard, high, ultra}) — creates digital surrogate

analyze_materials(technique: {XRF, Raman, FTIR}) — identifies composition

 doi:10.20944/preprints202602.0521.v2

recommend_conservation(treatment: {cleaning, consolidation, stabilization})
control_environment(parameter: {RH, T, light}, value: target)

N — Neural Components / Adaptive Models
StyleRecognizer — CNN for artistic-style classification
DeteriorationPredictor — physics-informed neural network for degradation forecasting
SemanticInterpreter — vision-language model for iconographic analysis
AuthenticityVerifier — GAN for detecting forgeries or tampering

E — Embedding Neural Component
CulturalArtifactEncoder

Fuses visual, material, historical, and contextual information:
Ecultural = CulturalTransformer(image, material, provenance, context, condition)

I — Identity Constraints / Fundamental Rules
Conservation ethics: Minimal_intervention ∧ Reversibility ∧ Documentation
Authenticity preservation: Original_material ≥ preservation_threshold
Access balance: Public_access ≥ minimum, Handling_damage ≤ maximum
Environmental standards:
Relative humidity (RH): 50% ± 5%
Temperature: 20 °C ± 2 °C
Light exposure: ≤ 50 lux for sensitive artifacts

T — Trigger Constraints / ECA Rules
If humidity > safe_range (40–60%) → activate_climate_control(setpoint = 50%)
If new_deterioration_detected → schedule_conservation(priority = urgency)
If handling_requested(artifact = fragile) → require_special_protocol(gloves = white)
If light_exposure > cumulative_limit → restrict_access(duration = cool_down)

G — Goal Constraints / Optimization Objectives
Maximize **information_preservation** = −ΔH(artifact_state), where H is entropy
Maximize **accessibility** = viewing_hours × visitor_count / preservation_cost
Minimize **degradation_risk** = Σ (λ_i × consequence$_i$ × vulnerability$_i$)
Maximize **cultural_value** = significance × condition × accessibility

D — Daemons / Real-time Monitors
EnvironmentalMonitor — tracks RH, T, light, pollutants
AuthenticityVerifier — checks for signs of forgery or tampering
AccessControlDaemon — logs and restricts handling
DeteriorationTracker — monitors changes in material properties

8.7.3. Mathematical Guarantees

Theorem 8.7.1 (Preservation Guarantee). For artifact A,

$$I(T) \geq I_0 \exp\left(-\int_0^T \lambda(t)\, dt\right),$$

where $\lambda(t)$ is the deterioration rate mitigated by conservation interventions.

8.8. Cross-Domain Analysis

8.8.1. Shared Mathematical Structure

Across all six domains, the same COH 9-tuple structure appears, demonstrating:

- universal hierarchical decomposition
- constraint satisfaction as a unifying principle
- consistent treatment of physical, biological, social, and cultural constraints
- modularity and compositionality across scales

 doi:10.20944/preprints202602.0521.v2

8.8.2. Flexibility Across Dimensions

COH supports:

- **scale adaptability** (from qubits to megacities)
- **temporal flexibility** (from femtoseconds to centuries)
- **uncertainty modeling** (quantum, stochastic, economic, environmental)
- **multi-objective optimization** (efficiency, safety, equity, preservation)

8.8.3. Unified Simulatability

The simulatability theorem (Theorem 7.1) holds across all domains, confirming that COH models are both theoretically expressive and practically implementable.

9. Connections to Existing Theories

9.1. Relationship to Free Energy Principle

Theorem 9.1 (COH as Generalized Free Energy Minimization). The COH optimization objective generalizes the free energy principle:

$$F(s) = \mathbb{E}_q[\log\, q(\psi \mid s) - \log\, p(\psi, s \mid \theta)] + \text{ConstraintViolation}(s)$$

where constraints act as additional priors.

9.2. Relationship to Hierarchical Reinforcement Learning

COH provides a formal foundation for hierarchical RL, with:

- Options corresponding to method sequences
- Subgoals corresponding to goal constraints
- Termination conditions corresponding to trigger constraints

9.3. Relationship to Cognitive Architectures

COH unifies aspects of:

- ACT-R (production rules as trigger constraints)
- SOAR (goal hierarchies as constraint hierarchies)
- CLARION (dual-process as neural vs. symbolic components)

10. Implications for AGI Implementation

10.1. World Modeling as Constrained Hierarchical Projection

Theorem 10.1 (World Model Fidelity). Let W be a real-world system and O_W its COH representation. For any finite observation sequence $o_{1:T}$ from W, there exists a COH model O_M such that the Kullback–Leibler divergence between the predicted and observed state distributions is bounded by the expressiveness of the constraint system:

$$D_{\mathrm{KL}}(P_W(s_t \mid o_{1:t}) \parallel P_M(s_t \mid o_{1:t})) \leq \alpha_H H(W) + \beta_C C(I) + \gamma_N \dim(N)$$

where $H(W)$ is the hierarchical complexity of W, $C(I)$ measures the completeness of identity constraints, and $\dim(N)$ represents the capacity of neural components. The constants $\alpha_H, \beta_C, \gamma_N$ are domain-dependent.

Proof sketch. Hierarchical decomposition C ensures structural alignment between model and world. Identity constraints I restrict deviations from physical laws. Neural components N provide universal approximation for residual dynamics. The bound follows by chaining approximation errors across hierarchical levels, with daemons D reducing residual divergence through constraint enforcement.

 doi:10.20944/preprints202602.0521.v2

10.2. Jagged Intelligence Avoidance Theorem

Theorem 10.2 (Smooth Capability Generalization). For a COH system O trained on tasks $T = \{T_1, \dots, T_m\}$, the performance gap on a new task T_{m+1} satisfies:

$$| R(T_{m+1}) - \frac{1}{m}\sum_{i=1}^{m} R(T_i) | \leq L \cdot d_C(S_{m+1}, \text{span}\{S_1, \dots, S_m\})$$

where d_C is the distance in constraint space, L is the Lipschitz constant of the policy with respect to constraints, and S_i is the constraint-admissible region for task T_i. The bound tightens as the constraint system becomes completer and more orthogonal.

Proof sketch. Jagged intelligence arises when systems perform well on training distributions but fail under distribution shift. COH reduces this risk through identity constraints I enforcing invariants, hierarchical composition C enabling systematic recombination of capabilities, and daemons D detecting constraint violations that indicate shift. The metric d_C quantifies alignment between task requirements, with smaller distances yielding smoother generalization.

10.3. Agentic System Implementation Theorem

Theorem 10.3 (Agent Construction from COH). Any autonomous agent specification

$$A = (\text{Percepts}, \text{Actions}, \text{Goals}, \text{Ethics})$$

can be implemented as a COH O_A with the following properties:

- **Perceptual grounding:** E maps raw percepts to semantic embeddings consistent with I.
- **Action realizability:** Each $a \in$ Actions corresponds to a method sequence $m_{i_1} \circ \cdots \circ m_{i_k} \in M^*$ respecting T.
- **Goal achievement:** Behavior converges to states maximizing G subject to I and T.
- **Ethical compliance:** Ethical principles are encoded as prioritized constraints $(I_{\text{ethics}}, T_{\text{ethics}})$ with lexical priority.

The agent's decision complexity grows as

$$O(\log | C | \cdot \text{poly}(| I |, | T |)),$$

rather than exponentially in state-space size.

Proof sketch. The architecture arises directly from COH components: perception as attribute updating A, world modeling via hierarchical decomposition C with neural prediction N, planning as constraint-satisfying method selection guided by G, and ethical behavior through prioritized constraints enforced by I, T, D. Hierarchical structure yields divide-and-conquer reasoning, while constraints prune the search space. Lexicographic ordering ensures ethical precedence (e.g., safety before efficiency).

10.4. Key Benefits of COH for AGI

Engineering Principles

- Modularity via hierarchical composition
- Safety through constraint enforcement
- Adaptability through neural components
- Interpretability through explicit constraint systems

Ethical Considerations

The constraint framework naturally encodes ethical principles as:

- Identity constraints for fundamental rights
- Trigger constraints for deontological rules
- Goal constraints for consequentialist optimization

Scalability Properties

 doi:10.20944/preprints202602.0521.v2

Theorem 10.4 (Compositional Scalability). The complexity of reasoning in a COH grows polynomially with the number of components, not exponentially.

Proof sketch. Hierarchical structure enables divide-and-conquer reasoning, while constraints limit cross-component interactions, preventing combinatorial explosion.

11. Conclusions

11.1. General Summary

This paper has presented a unified and mathematically rigorous foundation for **Constrained Object Hierarchies (COH)** as a theory of general intelligence. By formalizing intelligence as constrained hierarchical optimization, COH integrates structural composition, learning, prediction, and control within a single framework. Through its category-theoretic formulation, dynamical-systems interpretation, learning theory, and constraint-based semantics, COH establishes a coherent foundation for modeling adaptive, interpretable, and domain-general intelligent systems.

Beyond its theoretical development, COH demonstrates substantial **cross-domain generality**, capturing the structure and dynamics of systems as varied as quantum control, systems biology, smart-city infrastructure, aerospace missions, financial markets, and cultural heritage preservation. This breadth illustrates the expressive power of the framework and supports its potential as a unifying model for diverse forms of natural and artificial intelligence.

11.2. Future Directions

Several directions for further work emerge naturally from the present theory:

1. **Computational complexity and tractability** of inference, optimization, and hierarchical reasoning within COH-structured systems.
2. **Learning and adaptation of constraints**, including methods for discovering or refining identity, trigger, and goal constraints from data.
3. **Higher-order categorical generalizations**, extending COH to enriched categories, operadic structures, or multi-level semantic morphisms.
4. **Quantum extensions**, exploring how COH can incorporate quantum machine-learning components and hybrid classical–quantum intelligent architectures.

11.3. Final Remarks

COH provides a robust, neuroscience-grounded, and mathematically principled framework for understanding and engineering general intelligence. Its integration of hierarchical composition, constraint-driven regulation, predictive optimization, and adaptive learning positions it uniquely among existing theories of intelligence. By offering a unified account that spans symbolic, neural, dynamical, and probabilistic perspectives, COH lays a promising foundation for future AGI research, with implications for safety, interpretability, and scalable autonomous behavior.

References

Benzaquen, M., & Bouchaud, J.-P. (2018). Market impact with multi-timescale liquidity. *Quantitative Finance, 18*(11), 1781–1790.

Bucci, F., Benzaquen, M., Lillo, F., & Bouchaud, J.-P. (2019). Crossover from linear to square-root market impact. *Physical Review Letters, 122*(10), 108302. https://doi.org/10.1103/PhysRevLett.122.108302

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems, 2*(4), 303–314. https://doi.org/10.1007/BF02551274

Fong, B., & Spivak, D. I. (2019). *An invitation to applied category theory: Seven sketches in compositionality*. Cambridge University Press.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. MIT Press.

 doi:10.20944/preprints202602.0521.v2

Goebel, R., Sanfelice, R. G., & Teel, A. R. (2012). *Hybrid dynamical systems: Modeling, stability, and robustness*. Princeton University Press.

Hendrycks, D., Carlini, N., Schulman, J., & Steinhardt, J. (2022). *Unsolved problems in ML safety* (Version 5). arXiv. https://doi.org/10.48550/arXiv.2109.13916

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Lin, Z., Xuan, Y., Zhang, Y., Zhou, Q., & Qiu, W. (2025). Hypothalamus and brainstem circuits in the regulation of glucose homeostasis. *American Journal of Physiology–Endocrinology and Metabolism, 328*, E588–E598. https://doi.org/10.1152/ajpendo.00474.2024

Miyashita, Y. (2024). Cortical layer-dependent signaling in cognition: Three computational modes of the canonical circuit. *Annual Review of Neuroscience, 47*, 211–234. https://doi.org/10.1146/annurev-neuro-081623-091311

Pateria, S., Subagdja, B., Tan, A.-H., & Quek, C. (2022). Hierarchical reinforcement learning: A comprehensive survey. *ACM Computing Surveys, 54*(5), Article 3453160. https://doi.org/10.1145/3453160

Raji, I. D., & Dobbe, R. (2023). *Concrete problems in AI safety, revisited*. arXiv. https://doi.org/10.48550/arXiv.2401.10899

Rao, R. P. N., Gklezakos, D. C., & Sathish, V. (2024). Active predictive coding: A unifying neural model for active perception, compositional learning, and hierarchical planning. *Neural Computation, 36*(1), 1–32. https://doi.org/10.1162/neco_a_01627

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Tóth, B., Eisler, Z., & Bouchaud, J.-P. (2016). *The square-root impact law also holds for option markets*. arXiv. https://arxiv.org/abs/1602.03043

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008). Curran Associates, Inc.

Wang, H. (2025). Constrained object hierarchies as a unified theoretical model for intelligence and intelligent systems. *Computers, 14*, 478. https://doi.org/10.3390/computers14110478

Yarotsky, D. (2017). Error bounds for approximations with deep ReLU networks. *Neural Networks, 94*, 103–114. https://doi.org/10.48550/arXiv.1610.01145

Yun, C., Bhojanapalli, S., Rawat, A. S., Reddi, S. J., & Kumar, S. (2020). *Are transformers universal approximators of sequence-to-sequence functions?* arXiv. https://doi.org/10.48550/arXiv.1912.10077

Altman, E. (1999). *Constrained Markov decision processes*. Chapman & Hall/CRC.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). *Concrete problems in AI safety*. arXiv. https://doi.org/10.48550/arXiv.1606.06565

Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-Paz, D. (2019). *Invariant risk minimization*. arXiv. https://doi.org/10.48550/arXiv.1907.02893