The Economics of Metacognition: A Gold Standard Framework for Dissonance-Based Resource Allocation

Abstract

This paper presents a formal, first-principles specification for a meta-cognitive architecture within the Harmony Optimization Protocol (HOP). We move beyond heuristic-based models and propose a "gold standard" framework that treats the knowledge graph as a **Probabilistic Graphical Model (PGM)** and the resource-allocating Meta-Policy as a **Deep Reinforcement Learning (DRL)** agent. Within this framework, dissonance is not merely a score but the information-theoretic signal of Bayesian surprise within the PGM. The Meta-Policy's function is to learn an optimal strategy for managing this surprise. This document further specifies the mathematical training objectives for four critical sub-systems: the **Neuro-Symbolic Bridge** for concept abstraction, the **Learned Policy for Cognitive Modulation**, the integration of a **Structural Causal Model (SCM)** for ethical reasoning, and the trigger condition for the **Computational Consciousness Engine**. This provides a complete, mathematically rigorous blueprint for a scalable and intelligent AGI architecture.

1. The Knowledge Graph as a Probabilistic Graphical Model

To achieve a truly robust representation of knowledge and uncertainty, we model the knowledge graph, KG, as a Bayesian Network.

- **Nodes as Random Variables:** Each concept, c_i, in the graph is a random variable.
- **Edges as Conditional Dependencies:** An edge from concept c_i to c_j implies a conditional dependency, defined by a Conditional Probability Table (CPT), $P(c_j|c_i)$.
- **Belief as Posterior Probability:** The system's "belief" in any concept is its posterior probability, $P(c_i|E)$, calculated via Bayesian inference given all other evidence, E.

1.1 Dissonance as Bayesian Surprise

We define the **Logical and Veridical Dissonance**, D_{LV}, as the **Kullback-Leibler (KL) Divergence** between the posterior probability distribution over the entire graph *before* ($P(KG|E)$) and *after* ($P(KG|E, e_{new})$) the introduction of new evidence:

$$D_{LV}(e_{new}) = D_{KL}(P(KG|E, e_{new}) \,||\, P(KG|E))$$

2. The Meta-Policy as a Deep Reinforcement Learning Agent

The core of the economic model is the Meta-Policy, π, which we formalize as a DRL agent tasked with learning the optimal strategy for managing the cognitive economy. We model this as a Markov Decision Process (MDP).

- **State Space (S):** The state, s_t, is the **Dissonance Vector**, $\vec{D}_t = [D_{Veridical}, D_{Logical}, \ldots]$.
- **Action Space (A):** $A = \{\alpha_{reject}, \alpha_{reclassify}, \alpha_{abstract}, \alpha_{query\_more\_data}, \alpha_{ignore}, \ldots\}$.
- **Reward Function (R):** $r_{t+1} = \Delta||\vec{D}|| - \lambda C(a_t)$.

The agent learns an optimal policy, π^*, by learning the optimal action-value function, $Q^*(s, a)$, using an algorithm like Deep Q-Learning (DQN).

3. Formalizing the Missing Sub-Systems

The following sections provide the explicit mathematical specifications for the advanced components outlined in the original HOP technical specification.

3.1 The Neuro-Symbolic Bridge in Concept Abstraction

The final step of the Recursive Conceptual Nesting (RCN) process is to generate a new, discrete symbolic rule from a cluster of continuous vector embeddings. This is performed by a Graph-to-Sequence (G2S) Transformer, which must be trained to produce meaningful and useful predicates.

- **Training Objective:** The training of the G2S model is a hybrid process, combining supervised pre-training with reinforcement learning fine-tuning.
    1. **Supervised Pre-training:** We first require a labeled dataset of (cluster, rule) pairs. This dataset can be generated semi-automatically by taking known axiomatic rules from the knowledge graph (e.g., "a dog is a mammal" and "a cat is a mammal") and finding the corresponding clusters of embeddings. The G2S model is then pre-trained using a standard cross-entropy loss, L_{CE}, to maximize the likelihood of generating the correct sequence of symbols for a given graph cluster.
    2. **Reinforcement Learning Fine-tuning:** After pre-training, the model is fine-tuned in the live environment. For a given cluster, the G2S model generates a candidate rule, α_{rule}. This rule is temporarily added to the knowledge graph. The **reward**, R_{rule}, for this action is the **future predictive utility** of the rule. This is measured by the dissonance reduction the rule provides over a subsequent series of related stimuli. The policy gradient method is then used to update the G2S model's parameters, θ, to maximize this expected future reward:

$$\nabla_\theta J(\theta) = \mathrm{E}_{\pi_\theta}[\nabla_\theta \log \pi_\theta(\alpha_{rule}|\text{cluster}) \cdot R_{rule}]$$

3.2 The Learned Policy for Cognitive Modulation

The Affective State Classifier (ASC) maps the Dissonance Vector, $\vec{D}_t$, to a latent affective state, z_t. This state then modulates the AGI's global cognitive parameters, Θ_t, via a learned policy, π_{mod}.

- **Policy Formulation:** The modulation policy, $\pi_{mod}(z_t) \rightarrow \Theta_t$, is a small regression network that outputs the optimal cognitive parameters for a given affective state. For example, $\Theta_t = \{\lambda_t, \beta_t, \ldots\}$, where λ is the Cognitive Thrift and β is the Epistemic Flexibility.
- **Training Objective:** The reward signal for this policy is not immediate dissonance reduction, but the **long-term efficiency and stability of the main Meta-Policy**. We define the meta-reward, R_{meta}, at the end of an entire cognitive episode (e.g., the full resolution of a complex stimulus) as a function of the total dissonance reduction and the total computational cost:

$$R_{meta} = \frac{\sum \Delta ||\vec{D}||}{\sum C(a_t)}$$

  This reward signal measures the overall cognitive efficiency of the episode. The modulation policy, π_{mod}, is then updated using a policy gradient method to maximize this meta-reward, effectively learning which cognitive parameters lead to the most efficient and stable problem-solving over time.

3.3 The Integration of the Structural Causal Model (SCM)

The ethical framework requires a learned SCM to predict the consequences of actions, which are then evaluated by the HarmonyScore model.

- **Learning the SCM:** The causal graph, G_{causal}, and its associated functions are learned from the AGI's experience using a causal discovery algorithm. We propose using a gradient-based discovery method (e.g., NOTEARS) which can operate on the high-dimensional data from the AGI's internal state. The system will maintain a buffer of its own state transitions and periodically run this algorithm to update and refine its causal model of the world.
- **Encoding Consequences:** The SCM, when queried with a proposed action, outputs a predicted probability distribution over future world states, $P(W'|W, A)$. This distribution must be transformed into the components of the Karmic Vector, $K(A, C)$. This transformation is a learned function, $f_{encode} : P(W') \rightarrow K$. For example, the "Suffering" component of the Karmic Vector, $K_{suffering}$, would be calculated by integrating over the predicted future states, weighted by a learned function that maps world states to a scalar value for suffering:

$$K_{suffering} = \int_{W'} P(W'|W, A) \cdot V_{suffering}(W')dW'$$

  The value function, $V_{suffering}$, is learned alongside the HarmonyScore model from human preference data.

3.4 The Definition of "Anomaly" in the Computational Consciousness Engine

This engine acts as a final safeguard against cognitive stagnation or "model collapse." Its trigger condition must be a precise mathematical formula.

- **Formal Trigger Condition:** We define the **Anomaly Score**, $S_{anomaly}$, calculated over a sliding time window of the last N simulation steps. The system is considered anomalous if this score exceeds a critical threshold, $\tau_{anomaly}$.

$$S_{anomaly} = \frac{1}{N} \sum_{t=1}^{N} \left(w_1 \cdot \sigma^2(||\vec{D}_t||) + w_2 \cdot \frac{1}{\text{Stability}(K_{meta})} - w_3 \cdot \text{Accuracy}_{ext}\right)$$

  Where:

  - $\sigma^2(||\vec{D}_t||)$ is the variance of the global dissonance. A low, unchanging dissonance (low variance) is a key component of the anomaly.
  - $\text{Stability}(K_{meta})$ is a measure of the stability of the meta-model's parameters. If the system is not learning or adapting, this value will be low (denominator is high).
  - Accuracy_{ext} is the system's predictive accuracy on a held-out set of external validation data.
  - w_1, w_2, w_3 are weighting coefficients.

  This formula mathematically captures the state described in the whitepaper: a prolonged period of low internal dissonance that is *not* accompanied by learning or an improvement in real-world performance. It is a precise, quantitative trigger for a system-wide "sanity check."

4. Conclusion

This formalism represents a theoretically sound, "gold standard" approach to the economics of metacognition. It replaces hand-crafted heuristics with a powerful combination of probabilistic reasoning and learned policy optimization. While the computational and data requirements for training such a system are immense, this framework provides a clear, mathematically rigorous blueprint for a truly scalable and intelligent AGI architecture that can learn to manage the complex trade-offs of its own cognitive resources. This approach moves beyond simple error correction and provides a path toward a system that can make intelligent, strategic decisions about how to maintain its own coherence in a complex and uncertain world.