The Adaptive Cognitive Funnel (ACF): A Mathematical Specification

Abstract: This document provides a rigorous mathematical formalization of the Adaptive Cognitive Funnel (ACF) protocol. The ACF is a mechanism designed to mitigate reasoning myopia in dissonance-based AGI architectures by enabling scalable, budget-aware polling for latent global inconsistencies. We define the multi-stage selection and probing process, introduce the formal definition of Structural Tension (D_{ST}) as a new dissonance type, and specify its integration into the existing cognitive economy managed by the system's Meta-Policy.

1. Preliminaries and System Context

We assume the Harmony Optimization Protocol (H.O.P.) architecture as described in prior works. Let the system's knowledge be represented by a Knowledge Graph $KG = (V, E)$, where V is a set of concept nodes and E is a set of edges representing conditional dependencies. Each concept $v \in V$ has an associated vector embedding $\vec{e}_v \in \mathbb{R}^d$.

The system's state is driven by the Dissonance State Vector, $S_D^{(t)}$, which includes components such as Predictive Dissonance (D_P). A localized dissonance event identifies a subgraph $K_{sub} \subset KG$.

The system is governed by a Meta-Policy, π_{meta}, which learns to select actions, $a \in A$, to minimize global dissonance within a finite computational budget, C_{budget}. The reward function for π_{meta} is given by:

$$r_{t+1} = \frac{\Delta \text{ELBO}}{C(a_t)} - \lambda_{time}$$

The ACF protocol extends the action space A with a new composite action, a_{ACF}.

2. The Adaptive Cognitive Funnel (ACF) Protocol

The ACF is a sequential process composed of up to three stages. The decision to proceed to each subsequent stage is made by the Meta-Policy based on a cost-benefit analysis.

2.1. Stage 1: Heuristic Candidate Generation

- **Input:** A dissonant subgraph K_{sub} identified by a high local dissonance signal (e.g., $D_P(K_{sub})$).
- **Mechanism:**
  1. Compute the centroid embedding of the dissonant subgraph:

$$\vec{e}_{sub} = \frac{1}{|V_{sub}|} \sum_{v \in V_{sub}} \vec{e}_v$$

2. Perform an approximate k-Nearest Neighbors (k-NN) search over all concept embeddings in the main graph, $V \setminus V_{sub}$, using $\vec{e}_{sub}$ as the query vector.

- **Output:** A broad candidate set of k potentially related concept nodes:

$$V_{cand}^{(1)} = \{v_1, v_2, ..., v_k\} \subset V \setminus V_{sub}$$

- **Cost Function:** The computational cost of this stage, C_1, is dominated by the k-NN search.

$$C(a_{ACF,S1}) = C_{kNN}(|V|, d, k)$$

2.2. Stage 2: Learned Candidate Refinement

- **Input:** The candidate set $V_{cand}^{(1)}$ from Stage 1.
- **Mechanism:**

1. Employ a pre-trained Graph Attention Network (GAT), denoted $\text{GAT}_{influence}$, which has been trained to predict structural co-dissonance across the graph.
2. The GAT computes an influence score, $\alpha_{sub,j}$, for each candidate node $v_j \in V_{cand}^{(1)}$ relative to the source dissonance at K_{sub}. This score is derived from the learned attention mechanism of the GAT.

$$\alpha_{sub,j} = \text{Attention}(\text{GAT}_{influence}, K_{sub}, v_j)$$

- **Output:** A refined, ranked candidate set $V_{cand}^{(2)}$ of size $m \ll k$, containing only the nodes with the highest influence scores.

$$V_{cand}^{(2)} = \text{Top}_m(\{(v_j, \alpha_{sub,j}) \mid v_j \in V_{cand}^{(1)}\})$$

- **Cost Function:** The cost of this stage, C_2, is the cost of the forward pass through the GAT for each candidate.

$$C(a_{ACF,S2}) = k \cdot C_{GAT_{fwd}}$$

2.3. Stage 3: High-Fidelity Tension Probe

- **Input:** The refined candidate set $V_{cand}^{(2)}$ from Stage 2.
- **Mechanism:**
  1. For each candidate node $v_j \in V_{cand}^{(2)}$, define its local subgraph K_j.
  2. Perform a "Micro-ELBO" probe: execute a single gradient ascent step on the parameters λ_j of the variational distribution $Q_j(Z_j; \lambda_j)$ for the subgraph K_j.

$$\lambda_j' = \lambda_j + \eta \nabla_{\lambda_j} \mathrm{ELBO}(\lambda_j)$$

     where η is the learning rate.
  3. Measure the instantaneous change in the Evidence Lower Bound for that subgraph.
- **Output:** A set of scalar tension values, one for each probed node.

$$\{\Delta_j \mid \Delta_j = \mathrm{ELBO}(\lambda_j') - \mathrm{ELBO}(\lambda_j), \forall v_j \in V_{cand}^{(2)}\}$$

- **Cost Function:** The cost of this stage, C_3, is the sum of the costs for each probe, which involves a gradient calculation.

$$C(a_{ACF,S3}) = m \cdot C_{\nabla ELBO}$$

3. Formal Definition of Structural Tension (D_{ST})

The ACF protocol introduces a new dissonance type, Structural Tension (D_{ST}), which quantifies latent, non-local inconsistencies.

- **Definition:** For a given dissonant event at K_{sub} and a polled node v_j, the Structural Tension is the magnitude of the ELBO change from the Micro-ELBO probe, normalized by the cost of the probe itself to represent it as an efficiency.

$$D_{ST}(v_j | K_{sub}) = \frac{|\Delta_j|}{C_{\nabla ELBO}}$$

- **Aggregation:** The total Structural Tension for the cognitive cycle is the maximum observed tension across all probed nodes.

$$D_{ST}^{(t)} = \max_{v_j \in V_{cand}^{(2)}} \{D_{ST}(v_j | K_{sub})\}$$

- **Integration:** D_{ST} is incorporated into the global dissonance score via the learned meta-policy weights, $W^{(t)}$.

$$D_{global}^{(t)} = \sum_k w_k^{(t)} D_k^{(t)} + w_{ST}^{(t)} D_{ST}^{(t)}$$

4. Economic Integration with the Meta-Policy (π_{meta})

The ACF is not a mandatory process. The Meta-Policy, π_{meta}, must learn an optimal policy, π_{ACF}^*, for when and how deeply to engage the funnel.

- **Action Sub-Space:** The ACF action, a_{ACF}, is composed of a sequence of sub-actions $\{a_{S1}, a_{S2}, a_{S3}\}$.
- **Total Cost:** The total cost of an ACF action is the sum of the costs of the stages executed.

$$C(a_{ACF}) = \sum_{i \in \text{stages executed}} C(a_{ACF,Si})$$

- **Reward Signal:** The Meta-Policy is rewarded for using the ACF only if it leads to a more efficient reduction of long-term global dissonance. The reward function remains the same, but the Meta-Policy must now evaluate the trade-off between the immediate cost $C(a_{ACF})$ and the potential future dissonance reduction from resolving a latent conflict identified by D_{ST}. The policy must learn to estimate the expected value of information gained from polling.

$$Q^*(s, a_{ACF}) = \mathrm{E}\left[\frac{\Delta \mathrm{ELBO}_{\text{future}}}{C(a_{ACF})} - \lambda_{time} \mid s\right]$$

This formalizes the ACF as an economically managed cognitive process, where the system learns to invest computational resources in self-reflection only when the expected return in global coherence is sufficiently high.