A Graph-Theoretic and Active Inference Formulation of the Viable System Model

Jonas Hallgren

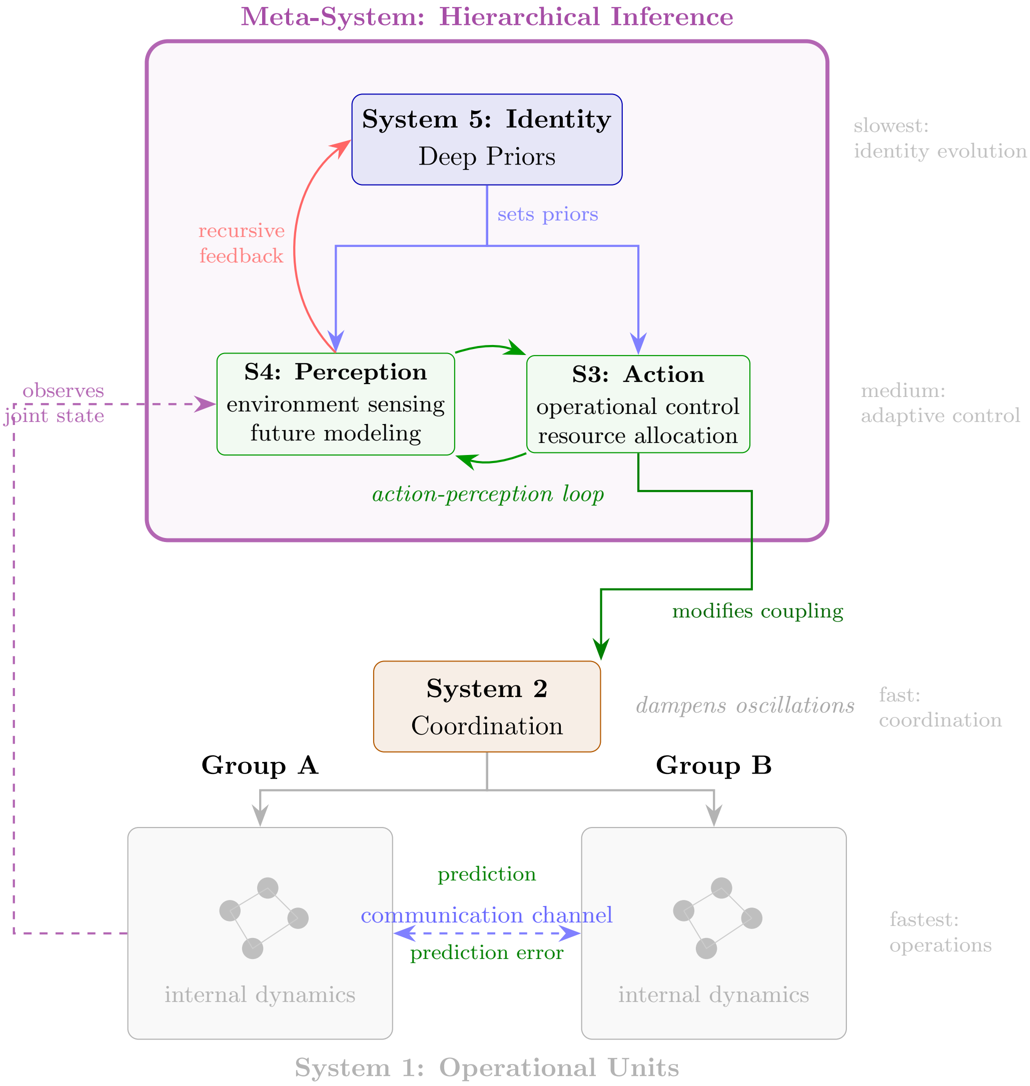
February 27, 2026

Abstract

The Viable System Model (VSM) by Stafford Beer provides a powerful, recursive blueprint for the organization of any persistent, autonomous system. Traditionally viewed through the lens of cybernetics and management science, its deep connections to modern computational and mathematical frameworks have remained largely implicit. This paper provides a formal description of the VSM from the combined perspectives of graph theory and the active inference framework, guided by the Free Energy Principle. We posit that the VSM is not merely a descriptive model but prescribes a necessary computational architecture for performing multi-scale variational inference. Each of the five VSM systems is defined in terms of its unique function in managing information flow and updating a collective generative model, thereby making the global problem of free energy minimization tractable for a complex, multi-agent system.

1 A Graph-Theoretic Visualization

The following diagram illustrates the information flows and structural relationships described below. It conceptualizes the VSM as a nested, multi-layered information processing architecture. The system boundary separates the organization from its environment, and information channels (arrows) show how each of the five systems contributes to the collective goal of minimizing free energy.



2 Foundational Concepts

2.1 The System as a Graph

We define a complex system as a collection of interacting agents. This structure is naturally represented as a graph $\mathcal{G} = (V, E)$, where V is the set of agents (vertices) and E is the set of interactions or communication channels (edges). The VSM imposes a specific, hierarchical structure on this graph to ensure viability.

2.2 Active Inference and Viability

Following the Free Energy Principle (FEP), we assert that for a system to be viable (i.e., to exist and resist dissolution), it must act to minimize its variational free energy, F. This quantity is a proxy for surprise or prediction error and is defined with respect to a generative model, $\mathcal{M}$, which encapsulates the system's beliefs about how its sensory inputs ($\tilde{y}$) are caused by hidden states of the world ($\tilde{x}$).

$$F = \underbrace{\mathbb{E}_{q(\tilde{x}|\mu)}[\ln q(\tilde{x}|\mu) - \ln P(\tilde{y}, \tilde{x}|\mathcal{M})]}_{\text{Variational Free Energy}} = \underbrace{D_{KL}[q(\tilde{x}|\mu)||P(\tilde{x}|\tilde{y}, \mathcal{M})]}_{\text{Complexity}} + \underbrace{-\mathbb{E}_{q(\tilde{x}|\mu)}[\ln P(\tilde{y}|\tilde{x}, \mathcal{M})]}_{\text{Accuracy (Surprise)}} \quad (1)$$

Here, $q(\tilde{x}|\mu)$ is an approximate posterior belief about the hidden states, parameterized by μ. Viability is achieved by continuously updating beliefs (μ) to minimize complexity and acting on the world (via policies π) to minimize surprise.

3 The VSM as a Multi-Scale Inference Architecture

We now formalize the five systems of the VSM as distinct, yet interconnected, computational modules solving different parts of the global free energy minimization problem.

3.1 System 1: Operations

System 1 comprises the fundamental operational units of the organization.

- **Graph-Theoretic View:** The graph $\mathcal{G}$ is partitioned into a set of disjoint subgraphs $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n\}$, where each $\mathcal{G}_i = (V_i, E_i)$ represents an operational unit. Each unit i interacts with its local environment, $\mathcal{E}_i$.
- **Active Inference Function:** Each unit $\mathcal{G}_i$ is an active inference agent in its own right. It possesses a local generative model $\mathcal{M}_i$ and minimizes a local free energy F_i based on its local observations $\tilde{y}_i$ from $\mathcal{E}_i$. Its primary task is to maintain a model of its immediate operational reality.
$$\forall i, \quad \mathcal{G}_i \text{ acts to minimize } F_i(\tilde{y}_i, \mu_i|\mathcal{M}_i). \tag{2}$$

3.2 System 2: Coordination

System 2 prevents harmful oscillations between the operational units of System 1.

- **Graph-Theoretic View:** System 2 is represented by a set of edges, E_2, that connect vertices across different subgraphs $\{\mathcal{G}_i\}$. These are the coordination channels.
- **Active Inference Function:** System 2 acts as a regularizer on the collective policy space. It introduces a constraint that minimizes the informational divergence between the beliefs of adjacent operational units. It ensures that the actions of one unit do not generate undue surprise for its neighbors. This can be framed as a constraint on the policies π_i, π_j for connected units i, j:
$$\text{minimize } D_{KL}[q_i(\tilde{x}_j|\mu_i)||q_j(\tilde{x}_j|\mu_j)], \quad \text{for } (\mathcal{G}_i, \mathcal{G}_j) \text{ connected by } E_2. \tag{3}$$

3.3 System 3: Control & Management

System 3 manages the internal homeostasis of the organization (the "Inside and Now").

- **Graph-Theoretic View:** System 3 is a metasystem, represented by a set of nodes V_3 with edges E_3 connecting to each operational unit $\mathcal{G}_i$. These edges represent channels for information compression (upwards) and control signals (downwards).
- **Active Inference Function:** System 3 does not observe the full state of System 1. Instead, each $\mathcal{G}_i$ sends a compressed signal, or sufficient statistic s_i, upwards. System 3 maintains a generative model of the entire internal organization, $\mathcal{M}_3$, conditioned on these statistics. Its function is to perform inference on the state of the whole. It minimizes a metasystem free energy F_3 and exerts control by setting priors on the policies or models of the System 1 units, $P(\pi_i|\mathcal{M}_3)$, to allocate resources and ensure synergy.

$$\text{Belief Update:} \quad \mu_3 \leftarrow \arg\min_{\mu_3} F_3(\{s_1, \ldots, s_n\}, \mu_3|\mathcal{M}_3) \tag{4}$$

$$\text{Control Action:} \quad P(\pi_i) \propto \exp(-F_3(\text{policy } \pi_i)) \tag{5}$$

3.4 System 4: Intelligence & Strategy

System 4 is concerned with the external environment and the future (the "Outside and Then").

- **Graph-Theoretic View:** System 4 is another metasystem, a set of nodes V_4 with edges representing information channels directed outwards to the global environment $\mathcal{E}_{\text{global}}$.
- **Active Inference Function:** The primary function of System 4 is to build and update a deep temporal model of the external world, $\mathcal{M}_4$. It performs inference on the hidden causes and dynamics of the environment, identifying long-term trends, opportunities, and threats. It is fundamentally engaged in model selection and learning over longer timescales than the other systems.

$$\mathcal{M}_4 \leftarrow \arg\max_{\mathcal{M}_4} P(\mathcal{M}_4|\text{historical observations from } \mathcal{E}_{\text{global}}) \tag{6}$$

3.5 System 5: Policy & Governance – Constraining Priors and Preferences

System 5 provides ultimate closure, identity, and strategic direction by defining the foundational context for the entire system's inference process.

- **Graph-Theoretic View:** System 5 is the apex node (or small clique) V_5 of the hierarchy, with edges connecting it to System 3 and System 4.
- **Active Inference Function:** System 5 performs inference at the highest level of abstraction, balancing the evidence of the current internal state from System 3 ($q(\text{internal}|\mu_3)$) with the model of the external future from System 4 ($\mathcal{M}_4$). Its role is twofold:

  1. **Constraining Generative Priors:** It sets the deepest, most abstract priors of the collective generative model, $P(\tilde{x}|\mathcal{M})$. These are the fundamental assumptions the system makes about itself and its place in the world—its identity.
  2. **Defining Preferences (Utility):** In active inference, preferences for certain outcomes are encoded as prior beliefs that those outcomes will be observed. System 5 defines these high-level preferences, $P(\tilde{y}|\mathcal{M})$, which constitute the system's core values or utility function. By defining what outcomes are least surprising, System 5 sets the ultimate objectives for the entire organization.

  This is achieved by setting the key hyper-parameters λ of the generative model that shape both beliefs and preferences.

$$\lambda^* = \arg\min_{\lambda} \mathbb{E}_{q(\text{states})}[F_{\text{global}}|\lambda] \tag{7}$$

  These hyper-parameters cascade down the hierarchy, providing the normative and existential constraints within which all subordinate systems operate.

4 Conclusion

Viewing the Viable System Model through the dual lens of graph theory and active inference reframes it from a static organizational chart into a dynamic, computational architecture. The five systems are not arbitrary divisions of labor but represent a necessary factorization of the global problem of existence. For a complex system to remain viable, it must implement these five fundamental information-processing functions: localized adaptation (S1), decentralized conflict resolution (S2), homeostatic internal regulation (S3), predictive environmental modeling (S4), and strategic goal-setting via the constraint of priors and preferences (S5). This decomposition makes the computationally intractable problem of minimizing a global free energy function tractable, providing a first-principles, physical justification for the VSM's enduring cybernetic wisdom.