Thoughtseeds: Evolutionary Priors, Nested Markov Blankets, and the Emergence of Embodied Cognition

Prakash Chandra Kavi[1*]

Gorka Zamora Lopez[1]

Daniel Ari Friedman[2]

[1]**Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain**
[2]**Active Inference Institute, Davis, California, USA**

*Corresponding Author

Abstract

The emergence of cognition requires a framework that bridges evolutionary principles with neurocomputational mechanisms. This paper introduces the "thoughtseed" framework, proposing that cognition arises from the dynamic interaction of self-organizing units of embodied knowledge called "thoughtseeds." We leverage foundational concepts from evolutionary theory, "neuronal packets," and the "Inner Screen" hypothesis within Free Energy Principle, and propose a four-level hierarchical/heterarchical model of the cognitive agent's internal states: Neuronal Packet Domains (NPDs), Knowledge Domains (KDs), thoughtseeds network, and meta-cognition. The dynamic interplay within this hierarchy, mediated by nested Markov blankets and reciprocal message passing, facilitates the emergence of thoughtseeds as coherent patterns of activity that guide perception, action, and learning. The framework further explores the role of the organism's Umwelt and the principles of active inference, especially the generative model at each nested level, in shaping the selection and activation of thoughtseeds, leading to adaptive behavior through surprise minimization. The "Inner Screen" is posited as the

locus of conscious experience, where the content of the dominant thoughtseed is projected, maintaining a unitary conscious experience. Active thoughtseeds are proposed as the fundamental units of thought that contribute to the "content of consciousness."

We present a mathematical framework grounded in active inference and dynamical systems theory. The thoughtseed framework represents an initial but promising step towards a novel, biologically-grounded model for understanding the organizing principles and emergence of embodied cognition, offering a unified account of cognitive phenomena, from basic physiological regulation to higher-order thought processes, and potentially bridge neuroscience and contemplative traditions.

Introduction

Embodied Cognition: An Evolutionary and Free Energy Perspective

Adaptation and knowledge transmission, driven by evolutionary mechanisms like the Baldwin effect and natural selection, are crucial for survival, allowing organisms to build on past experiences and innovate [59;5;19]. Modern perspectives on evolution recognize the intricate dynamics of inheritance, encompassing not only genetic factors but also epigenetic, behavioral, and cultural influences and their co-evolutionary dynamics [88;109]. The concept of embodied cognition further underscores the deep interdependence between an organism's brain, body, and environment, emphasizing how actions and sensory experiences actively shape understanding and cognitive processes [166;108].

Autopoiesis describes the self-organizing nature of living systems [101;1] which exist far from equilibrium as thermodynamically open systems that maintain a non-equilibrium steady state (NESS) by exchanging energy and matter with their environment, thereby resisting the natural tendency towards disorder (entropy) as [144; 106;78;47]. The Good Regulator Theorem [15] further emphasizes: "effective regulation requires an internal model."

The Free Energy Principle (FEP) provides a unifying framework for understanding how organisms maintain stability and adapt by *minimizing surprise* [44;1]. Organisms reduce the discrepancy between predictions from their generative models and sensory input, approximated by variational free energy (VFE), a proxy for surprise [42] . To minimize surprise or VFE, they engage in active inference [45;81] – a process of proactively refining their internal models and taking actions that shape future sensory experiences, thereby bringing their predictions in line with reality [123;120]. The Markov blanket, central to the FEP, separates internal states from the external world, enabling localized computations and conditional independence [121;44]. Sensory and active states within the blanket mediate organism-environment interactions [116;54;56].

The FEP's compatibility with scale-free modeling allows the integration of evolutionary concepts across multiple scales, providing insights into the emergence and adaptation of cognitive systems [54]. It extends beyond individual behavior to encompass evolutionary dynamics, proposing a variational formulation of natural selection that views adaptive fitness as a path integral of phenotypic fitness [57]. Interestingly, the **hierarchically mechanistic mind (HMM)** hypothesis describes the human brain as: "an embodied, complex adaptive control system that actively minimizes the variational free-energy (and, implicitly, the entropy) of (far from equilibrium) phenotypic states via self-fulfilling action-perception cycles, which are mediated by recursive interactions between hierarchically organized (functionally differentiated and differentially integrated) neurocognitive processes." These *mechanics* instantiate adaptive priors, and have emerged from selection and self-organization co-acting upon human phenotypes across different timescales [4].

Neuronal Representations

The quest to understand the neural basis of cognition has led to the exploration of various representational units beyond individual neurons, including:

- **Cognits:** Discrete, innate units of semantic representation [36]
- **Engrams:** Physical traces of memory involving specific synaptic patterns [76]

- **Cell Assemblies:** Groups of interconnected neurons representing specific concepts or memories [65]

While insightful, these concepts often lack a unifying framework addressing hierarchical organization, dynamic nature, and evolutionary origins of neural representations. Furthermore, a fundamental challenge in elucidating the neural basis of cognition lies in the inherent opacity of the Markov blanket, that can be referred to as a "black box" problem. While this statistical construct offers a robust framework for modeling the exchange of information between an agent and its environment, it inherently obscures the internal generative model responsible for behavior. As Friston [45;53; 54] highlights, "the Markov blanket renders the internal states of a particle conditionally independent of external states, thus precluding direct observation of the internal generative model."

The **Neuronal Packet Hypothesis (NPH)** [171;113;135] suggests that **"neuronal packets (NPs),"** self-organizing ensembles of neurons, are the *fundamental units of neuronal representation* in the brain [172]. **Superordinate ensembles (SEs)** emerge from the coordinated activity of multiple NPs, enabling the representation of complex concepts. The interactions and signal emissions from NPs allow the system to infer beliefs about external states, updating these beliefs to minimize free energy. This results in a shared generative model that encodes the system's collective beliefs and predictions, facilitating effective responses to stimuli. SEs are hierarchically organized, forming nested layers of representation [135]. Each NP and SE has its own Markov blanket, defining its boundaries and interactions. The Markov blankets of lower-level NPs are nested within those of higher-level SEs, integrating their information via a shared generative model [113]. This structure allows the brain to represent knowledge across multiple scales, from sensory details to abstract categories. Nested SEs interact through reciprocal message passing, where top-down predictions from higher-level SEs modulate lower-level SEs and NPs, while bottom-up sensory evidence or prediction errors update higher-level representations facilitating [82;113;54].

FEP further suggests that the content of neural representations have a role in guiding action and minimizing surprise [134;136]. NPH provides a strong foundation for understanding the neural basis of knowledge representation, and thereby a fertile ground

for further exploration: How do evolutionary pressures shape the emergence and organization of neuronal packets and their hierarchical ensembles? How do individual experiences interact with inherited predispositions to refine and adapt these representations? How do these representational units interact within a network to generate complex cognition and adaptive behavior?

The Thoughtseed Hypothesis

To address the challenges of understanding the internal generative model, often obscured by the inherent opacity of the Markov blanket at the agent level, we propose the "thoughtseed" framework. This framework, building upon the concept of neuronal packets (NPs) and the neuronal packet hypothesis (NPH), integrates insights from the free energy principle (FEP), "Inner Screen" Hypothesis [34; 137] and evolutionary theory.

The thoughtseed hypothesis posits that cognition arises from the dynamic interaction of self-organizing units of embodied knowledge called "thoughtseeds." These thoughtseeds, conceptualized as emergent agents with Markov blankets, are shaped by a hierarchy of evolutionary priors, representing both inherited predispositions and learned experiences. The framework proposes a four-layered hierarchical/heterarchical model of the cognitive agent's internal states, comprising Neuronal Packet Domains (NPDs), Knowledge Domains (KDs), the Thoughtseed Network (TN), and meta-cognition. The thoughtseed framework emphasizes the embodied nature of cognition, where the organism's interactions with its Umwelt [167] and the knowledge encoded in the KDs shape the Thoughtseed Network.

Each level in this model acts as a "nested holographic screen" where information encoded at each level is progressively coarse-grained, moving from fine-grained details at lower levels to more abstract and categorical representations at higher levels [34;137]. The dynamic interplay within this hierarchy facilitates the emergence of thoughtseeds as coherent patterns of activity that guide perception, action, and learning. Crucially, the "Inner Screen" serves as the "locus of conscious experience," [137] where the content of the dominant thoughtseed is projected, resulting in a unified conscious experience.
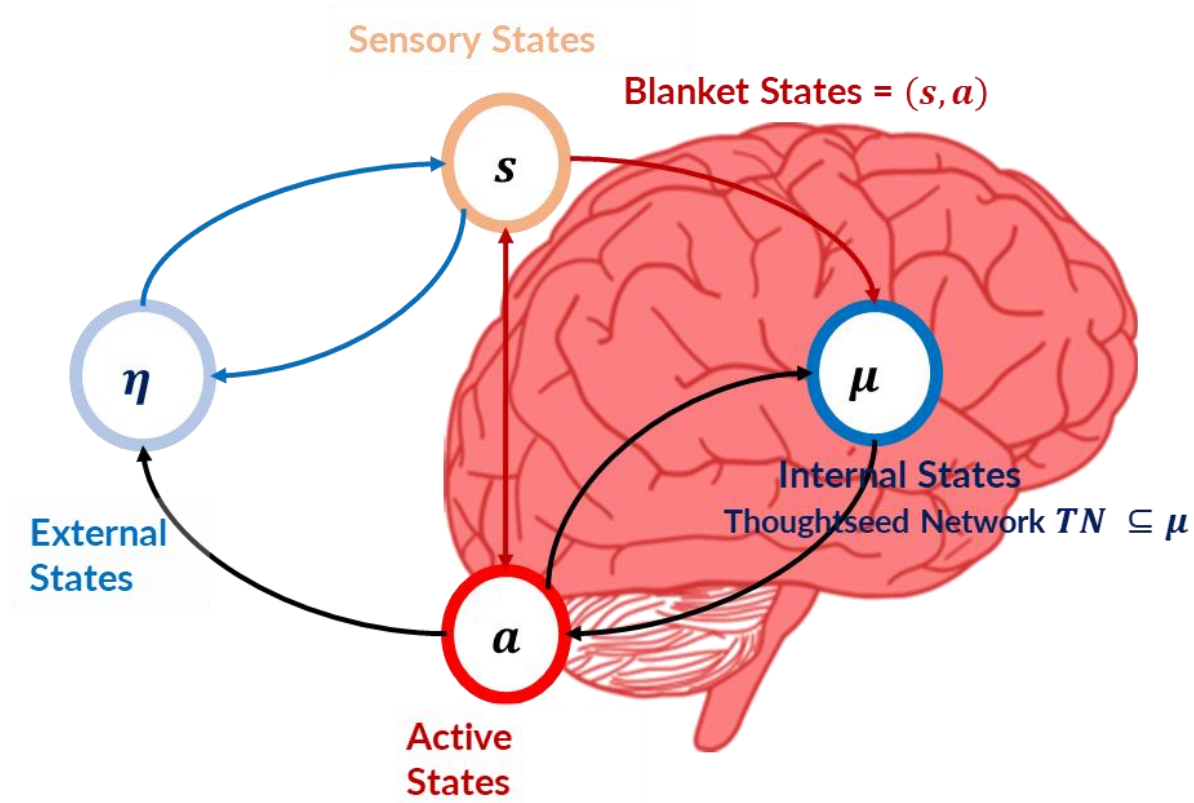
Fig 1. Markov Blanket of an Agent. Adapted from Friston et al., 2023, this diagram illustrates the partitioning of an agent's states into internal states (μ), encompassing the Thoughtseed Network (TN), and external states (η). The Markov blanket, comprising sensory (s) and active states (a), mediates the interaction between the internal and external states. The internal states, housing the TN, can only influence active states, while external states can only influence sensory states. This separation allows the TN to operate with a degree of autonomy, generating predictions and selecting actions based on its internal model of the world.

Methods and Materials: Theoretical Framework

We discuss the layered cognitive framework of thoughtseeds, starting with Evolutionary Priors.

Concept	Symbol	Explanation
Neuronal Packet (NP)	ν	The fundamental unit of neuronal representation, a self-organizing ensemble of neurons that encodes a specific feature or aspect of the world.

Core Attractor	ψ_c	The most probable and stable pattern of neural activity within a manifested NP, embodying its core functionality.
Subordinate Attractor	ψ_{s_i}	Less dominant patterns of neural activity within an NP that may become active under specific conditions or in response to novel stimuli, offering flexibility and adaptability.
Superordinate Ensemble of NPs (SE)	$\mathcal{E}$	A higher-order organization emerging from the coordinated activity of multiple NPs, enabling the representation of more complex and abstract concepts.
Neuronal Packet Domain (NPD)	N	A functional unit within the brain, comprised of interconnected SEs, specialized for specific cognitive processes or tasks.
Knowledge Domain(KD)	K	A large-scale, organized structure within the brain's internal model, representing interconnected networks of concepts, categories, and relationships that constitute a specific area of knowledge or expertise.
Thoughtseed	$\mathcal{T}$	A higher-order construct with agency, emerging from the coordinated activity of SEs across different KDs. It represents a unified and meaningful representation of a concept, idea, or percept and guides perception, action, and decision-making.
Thoughtseed Activation Level	α_i	A measure of a thoughtseed's prominence in the current cognitive landscape, calculated by weighing the probabilities that the brain's state aligns with the thoughtseed's dominant or subordinate attractor states.
Activation Threshold	$\Theta_{activation}$	A global parameter that determines the minimum activation level required for a thoughtseed to be considered part of the active thoughtseed pool.
Active Thoughtseed Pool	$\mathcal{P}_{active}$	The set of thoughtseeds whose activation levels surpass the activation threshold at a given time.
Dominant Thoughtseed	i^*	The thoughtseed within the active thoughtseed pool that has the highest activation level and is primarily shaping the content of consciousness at a given moment.
Thoughtseed Network	TN	The collection of interconnected thoughtseeds within the brain's internal states, hypothesized to encode a generative model of the environment.
Inner Screen	$\mathcal{C}$	The locus of conscious experience, where the content of the dominant thoughtseed is projected.

Table 1. Key Notations of Thoughtseed Framework

Evolutionary Priors

Evolutionary priors, including inherited predispositions and learned experiences, play a pivotal role in shaping an organism's cognitive landscape [86]. These priors influence the emergence, organization, and adaptation of NPDs and KDs, contributing to adaptive behavior. Research suggests that evolutionary priors shape not only broad brain architecture but also specific computations and representations within neural networks [51; 54;124].

The thoughtseed framework posits that NPD formation is guided by evolutionary priors, acting as a "blueprint" that favors the emergence of adaptive neural architectures and functional units [131]. The visual cortex, for instance, exhibits distinct NPDs for color, motion, and form [174; 27], specializations conserved across mammalian species [100]. The specific organization of NPDs is further refined through Hebbian learning. In the context of KDs, evolutionary priors predispose the brain to certain initial configurations and types of information, which are then refined through lifelong learning and experience [150].

These embodied priors significantly influence the types of thoughtseeds that emerge and their interactions within the cognitive system. Thoughtseeds linked to autopoietic and emotional states can effectively model responses essential for evolutionary survival strategies [20;62;105], contributing to the organism's ability to reach and maintain evolutionary stable states (ESS) [149].

Evolutionary priors can be understood within the broader framework of multi-level learning and inheritance systems [75;165]. Even if we are unable to capture the inherent complexities of these processes, still distinguishing two broad categories: **phylogenetic** and **ontogenetic** as shown in Table 2 could be useful [97]. Figure 2 describes the quasi-hierarchical influence of these priors on thoughtseed emergence.

Prior Type	Origin	Influence	Relative Stability	Active Inference Role
Basal (B)	Evolutionary (species-wide)	Fundamental needs, core behaviors	High	Shapes initial state space and action repertoire
Lineage-	Evolutionary	Accumulated	Medium	Guides development and

Specific (L)	(lineage-specific)	knowledge, species-specific behaviors		refines basal priors
Dispositional (D)	Genetic and developmental (individual)	Predispositions towards behaviors and learning styles	Low to Medium	Influences initial policies and policy choices
Learned (λ)	Experiential (individual)	Learned associations, adaptive strategies, updated beliefs	Highly dynamic	Continuously updates beliefs and refines internal model

Table 2: Summary of Evolutionary Priors. Phylogenetic (Basal + Lineage-specific). **Ontogenetic** (Dispositional + Learned). This is a probabilistic mapping, and not deterministic with scope for mutual overlap.
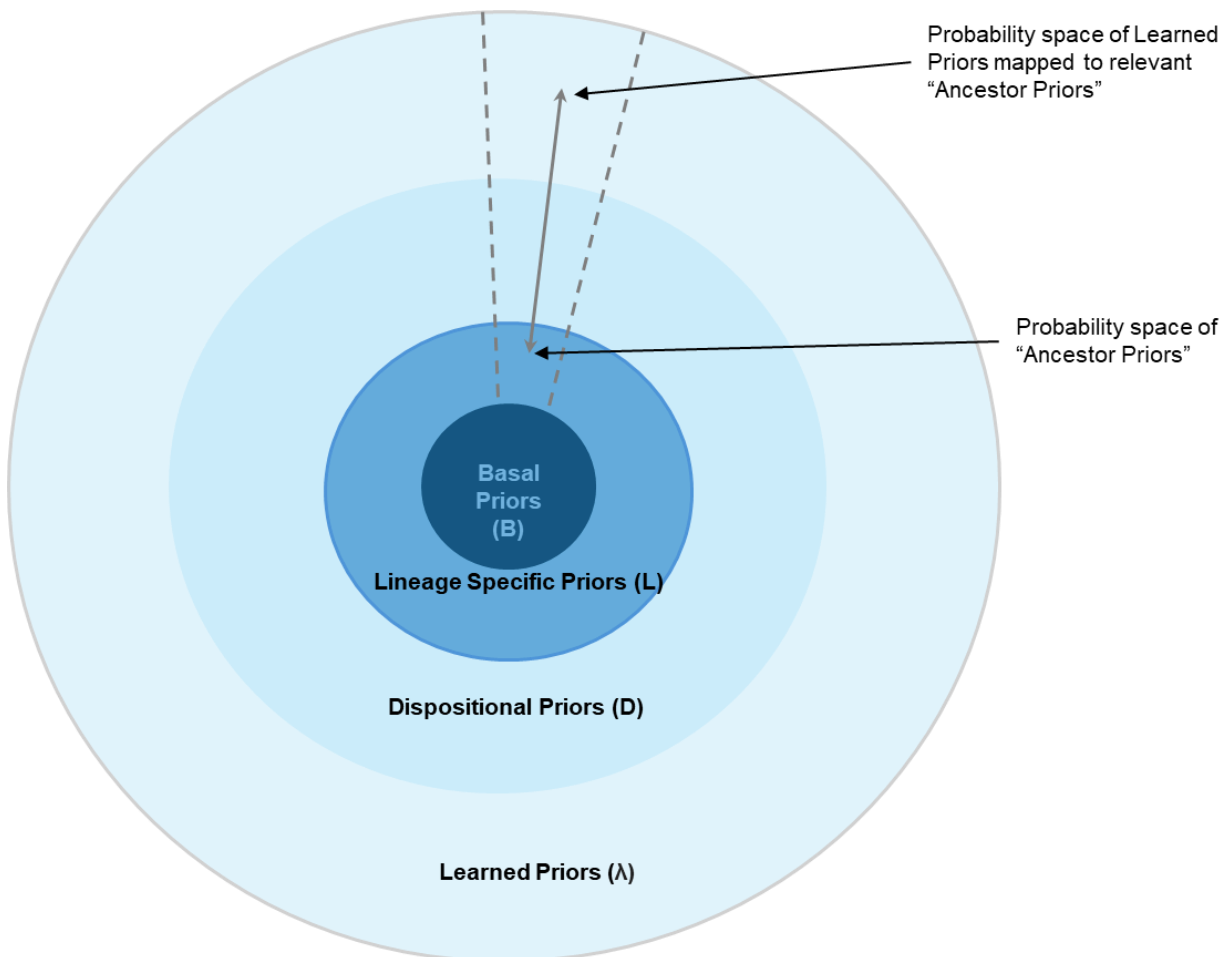
These priors are further discussed in .

Fig 2: Quasi-Hierarchical Influence of Evolutionary Priors on Thoughtseed Emergence. This diagram illustrates the nested hierarchy of evolutionary priors shaping an organism's internal model. **Basal Priors** (core): The most fundamental, species-wide biases inherited from evolutionary history. **Lineage-Specific Priors**: Adaptations specific to a lineage, including *Ancestor Priors* (successful learned strategies passed

down through generations). **Dispositional Priors**: Individual predispositions arising from genetics and development. **Learned Priors** (outermost): Dynamic biases shaped by individual experiences, influencing both Dispositional priors and, over generations, ancestor priors. The arrows indicate the flow of influence, where inner layers constrain outer ones, and learned priors can contribute to the evolution of ancestor priors over time.

Layer 1: Neuronal Packet Domains (NPDs)

The brain's sparse architecture, with few active neurons and low connection density, supports localized functional units for specific cognitive tasks [10;152]. NPDs are hypothesized to emerge through self-organizing interconnected NPs. Their hierarchical organization, via nested Markov blankets, reflects "ascending scales of canonical microcircuits" [68;28], enabling complex computations from simpler units; and efficient information processing and adaptive behavior [113; 54].

Neuronal Packets (NPs): The Fundamental Units of Neuronal Representation

Within each NPD, NPs serve as the fundamental units of neuronal representation [135; 171;172]. An NP could exist in three states:

- **Unmanifested State:** Represents a potential configuration of neural activity with high prior probability under specific conditions, shaped by evolutionary priors. It can be viewed as a sparsely connected neural ensemble with low precision, corresponding to a shallow local minimum in the free energy landscape, indicating low stability and high potential for change.
- **Manifested State:** Emerges from the unmanifested state upon repeated exposure to relevant stimuli, leading to a **phase transition** [172] and the formation of a Markov Blanket stabilized by an **energy barrier** [172]. This Markov blanket structure may enable local computation and autonomy within the NP, while maintaining informational boundaries. It is characterized by increased coherence of neural activity, resulting in a stable state with a **core attractor** representing the most probable and stable pattern of neural activity [47;118]. This state embodies the NP's core functionality with high certainty and can be interpreted as the mode

of the posterior distribution over the NP's internal states, given its Markov blanket [54]. It corresponds to a deeper local minimum in the free energy landscape, reflecting high stability and low surprise [80;33]. The depth of this global minimum could indicate the NP's **binding energy**, reflecting the overall stability and resistance to change of the core representation [172]. **Alternative attractors** represent less dominant patterns of neural activity that may become active under specific conditions or in response to novel stimuli [132]. They can be viewed as shallower local minima in the free energy landscape, separated from the core attractor by **energy barriers** [172]. The existence of alternative attractors, along with the energy barriers that separate them, allows for flexibility and adaptability in the NP's response to changing inputs.

- **Activated (or Spiking) State:** A transient state characterized by heightened neural activity within the manifested NP ensemble, triggered by specific inputs that resonate with the NP's internal model. The NP generates a response that may influence the organism's behavior or cognition. This response can be interpreted as a consequence of the NP's internal dynamics and its attempt to minimize free energy, rather than a deliberate or planned action. The active state can be seen as a temporary shift in the free energy landscape, where the core attractor becomes even more pronounced.

NPs compete for resources by minimizing VFE, ensuring efficient internal models. This competition, viewed as "neural Darwinism," leads to a diverse repertoire of specialized NPs, each representing different aspects of the world [30]. The Markov blanket structure supports a decentralized, modular NP organization, promoting robustness and adaptability.

Superordinate Ensembles (SEs) within NPDs

NPs are hypothesized to dynamically interact to form SEs, within a NPD, representing a higher level of organization and encoding more complex and abstract concepts. The hierarchical organization, where the Markov blankets of lower-level NPs are nested within those of higher-level SEs, allows the brain to represent knowledge across multiple scales,

from sensory details to abstract categories [41;135]. SEs may emerge as stable entities when they accurately predict and explain sensory input [54].

Layer 2: Knowledge Domains (KDs)

KDs can be conceptualized as large-scale, organized structures within the brain's internal model, akin to *knowledge graphs* [70]. They function as nested SEs of "knowledge repositories" that provide the conceptual scaffolding for interpreting sensory information projected from the NPDs.

The dynamic and context-dependent binding process within KDs contributes to the "content of consciousness" on the Inner Screen, conditioned by the active thoughtseed. This process may involve the synchronization of neural activity [148;112], or mechanisms like Binding by Firing Rate Enhancement [140]. This binding process is plausibly facilitated through shared generative models, reciprocal message passing, and attentional selection. Thereby, reflecting "circular causality," where different levels of the hierarchy mutually influence each other through predictions and error corrections [43].

KDs could exhibit both hierarchical and heterarchical structures, enabling flexible and context-dependent knowledge retrieval and utilization for adaptive behavior [122]. The hierarchical aspect reflects the layered organization of knowledge, from concrete to abstract [41;68;137], while the heterarchical aspect captures the interconnectedness and cross-domain interactions [172].

Potential neural correlates of KDs, even though speculative, might include: hippocampus, serving as a central hub for a "memory" KD (integrating information to create episodic and semantic memories) [153;31]; amygdala, associated with an "emotion" KD (linking sensory information with affective responses) [89;126]; prefrontal cortex, acting as a central hub (for multiple KDs related to executive functions and goal-directed behavior) [104;58]. An illustrative example of Visual Object Recognition KD is shown in Supplementary Section 9.4.

Layer 3: Thoughtseeds Network

The thoughtseed framework proposes that thoughtseeds, emerging from the coordinated activity of distributed neural networks, generate the "content of consciousness" on the "Inner Screen" (the locus of conscious experience) [34;137]. The Inner Screen is central to embodied cognition, where thoughtseeds, the emergent agents of cognition, compete for dominance, shaping the content on the Inner Screen—sensory observations, memories, knowledge, and potential actions. The dynamic patterns of neuronal activity and connectivity that constitute an activated thoughtseed are self-sustaining, arising from the coordinated activity of neuronal ensembles across multiple brain regions. Thoughtseeds can be understood as metastable states [22] within the brain's pullback attractor landscape, representing stable patterns of neural activity that minimize free energy and are reinforced through repeated visits, driven by interactions with the environment and internal dynamics [44; 47;49].

The content encompassed by thoughtseeds emerge from the knowledge and beliefs embedded within KDs, reflecting the organism's evolutionary history and individual experiences. This dynamic interplay between thoughtseeds and KDs is pivotal in shaping cognition, behavior, and conscious experience. Thoughtseed activation triggers a cascade of activity within associated KDs, facilitating the retrieval and integration of relevant knowledge and memories. Thoughtseeds can also modulate KDs and shape their encapsulated knowledge, influencing conscious experience and adaptive behavior.

Key Characteristics

- **Thoughtseeds as Sub-Agents:** Thoughtseeds act as autonomous sub-agents within the cognitive system, engaging in active inference to generate predictions, influence actions, and update internal models based on sensory feedback [145]. This agency allows them to explore the environment and develop affordances.
- **Pullback Attractor Dynamics:** When active, thoughtseeds function as pullback attractors [47;64], integrating information from multiple SEs and KDs to form coherent representations. This establishes a **transient Markov blanket**,

maintaining autonomy and computational independence. The pullback attractor nature ensures stability and resistance to perturbations. Each thoughtseed is associated with a **core attractor**, representing its most probable and stable pattern of neural activity, embodying its core functionality or meaning. It may also have **subordinate attractors**, representing less dominant but still accessible patterns of activity that offer flexibility and adaptability in response to varying contexts or stimuli. The ongoing neural activity during the specific thoughtseeds activation are guided by an interplay of bottom-up saliency signals and top-down attentional mechanisms, influenced by meta-cognition and higher-order thoughtseeds.

- **Thoughtseed States:** Thoughtseeds can exist in the following states:
  - **Unmanifested:** The initial state where the thoughtseed exists as a potential configuration within the neural network, but is not yet actively influencing cognition or behavior.
  - **Manifested:** The thoughtseed has emerged and is now part of the cognitive landscape, with the potential to influence thought and action. The sub-states further refine this:
    - **Inactive:** The thoughtseed is present but not currently contributing to conscious experience.
    - **Activated:** The thoughtseed is in the "active thoughtseed pool" and contributes to the content of consciousness.
    - **Dominant:** The thoughtseed has the highest activation level and is primarily shaping conscious experience.
  - **Dissipated:** The thoughtseed has ceased to be active and its influence has diminished, potentially returning to an unmanifested state or leaving remnants that could facilitate future re-activation.

Free Energies, Goals and Policies of a Thoughtseed

The interaction between thoughtseeds and KDs generates **affordances** [61], or potential actions, that **minimize expected free energy (EFE)** [118] . The thoughtseed's generative model, informed by its states and knowledge within KDs, evaluates actions based on

predicted outcomes and prior beliefs, balancing exploration (epistemic affordances) and exploitation (pragmatic affordances) to achieve its goals [50;55].

The **generalized free energy (GFE)** of a thoughtseed extends VFE to incorporate the expected consequences of actions and their impact on future states [118;9] GFE represents the thoughtseed's overall "surprise," considering both immediate VFE and expected future consequences. By minimizing GFE, the thoughtseed selects the optimal policy leading to desirable future states, promoting long-term adaptation and goal achievement.

In the thoughtseed framework, goals and policies emerge from the dynamic interplay between thoughtseeds, KDs, internal states, and perceived affordances [51].

Layer 4: Meta-Cognition & Orchestration of Thoughtseed Dynamics and the Inner Screen

Meta-cognition, the ability to monitor and control one's thoughts, plays a crucial role in shaping the dynamics of the thoughtseed network. This orchestration unfolds on the Inner Screen, a dynamic mental workspace where perception, knowledge, memories, and potential actions converge to form a unified conscious experience [34;137].

Policies, Goals, and Affordances at the Agent Level

The thoughtseed framework extends the concepts of goals, policies, and affordances to the agent level. Agent-level goals represent the desired outcomes for the organism as a whole, while agent-level policies are the overarching strategies employed to achieve these goals. The agent's perception of affordances, both epistemic (opportunities for learning) and pragmatic (opportunities for goal fulfillment), is also shaped by its global goals and the states of its constituent KDs.

The emergence of global goals and policies can be understood through ergodic principles [48], which suggest that long-term behavior can be inferred from the statistical properties of a system's trajectories over time [52]. The brain's pullback attractor landscape, shaped

by knowledge and beliefs within the KDs, exhibits characteristic states, representing stable and recurring patterns of neural activity that are revisited frequently, reflecting the organism's overarching goals and preferred policies. Analyzing these states can provide insights into long-term behavioral tendencies and priorities [10].

Free Energies at the Agent Level - VFE, GFE, EFE

In the thoughtseed framework, three interconnected forms of free energy [118;55] drive adaptive actions and the minimization of surprise:

- **Agent-Level VFE:** The instantaneous surprise experienced by the agent, encompassing the combined VFE of all active thoughtseeds, reflecting discrepancies between predictions and actual observations.

- **Agent Level EFE:** Guides policy selection, quantifying the expected surprise associated with a particular policy. It considers both epistemic (information gain through exploration) and pragmatic (goal fulfillment via exploitation) affordances. The agent, driven to EFE, selects policies balancing exploration and exploitation, promoting adaptive behavior.

- **Agent-Level GFE:** Extends VFE to incorporate the expected consequences of actions and their impact on future states, considering both immediate and anticipated surprise. GFE represents the total free energy over a given time horizon, accounting for potential outcomes and uncertainties resulting from the agent's chosen policy.

The Dynamics of Thoughtseed Selection on the Inner Screen

The emergence of a dominant thoughtseed is a dynamic process of competition and selection among multiple potential thoughtseeds. This competition is influenced by both bottom-up saliency signals and top-down attentional control. The thoughtseed that best explains sensory input, aligns with goals, and minimizes free energy emerges as dominant, guiding behavior.

A thoughtseed's activation level reflects its prominence in the current mental state. It is determined by weighing the likelihood that the brain's current activity aligns with the thoughtseed's dominant or subordinate attractor states. An activation threshold acts as a gatekeeper, ensuring that only the most pertinent thoughtseeds, those with activation levels surpassing this threshold, contribute to the ongoing "content of consciousness," forming the "active thoughtseeds pool." This threshold could be dynamically adjusted based on various factors, including arousal levels, attentional precision weighting, context, and metacognitive goals.

Furthermore, top-down attentional control interacts with bottom-up saliency signals to determine which thoughtseeds become active and compete for dominance.

- **Meta-cognition and Attention:** guided by higher-order thoughtseeds or goals, acts as a selective filter, modulating sensory processing and influencing thoughtseed emergence [26;18;32]. It amplifies the precision of attended sensory signals, thereby activating relevant thoughtseeds and influencing their competition for dominance.
- **Saliency as Prior Expectation:** Saliency, the attention-grabbing quality of a stimulus [72], shapes prior expectations, creating biases towards salient information. This aligns with predictive processing, where prior expectations guide perception and attention[14;115].

Stream of Thoughts on the Inner Screen

The stream of thought can be conceptualized as the dynamic and evolving sequence of the *content* projected onto the Inner Screen, reflecting the "content of consciousness" [34;137]. This content is primarily shaped by the dominant thoughtseed within the *active thoughtseeds pool*, but can also be subtly influenced by other thoughtseeds within the pool vying for prominence. This continuous emergence, competition, and transition of thoughtseeds within the "active thoughtseeds pool" that project content on the Inner Screen mirror the ever-changing landscape of the internal model as it interacts with the environment. It is further influenced by a complex interplay of factors, including the organism's *Umwelt*, attentional mechanisms, saliency, long-term and current goals at

Agent and thoughtseed level, and prior experiences. This process continuously updates the dominant thoughtseed's internal model and actively generates predictions about the world [71;118]. The sequence of thoughtseeds that emerge represents the living system's best attempt to make sense of its environment and guide its behavior towards adaptive outcomes.

The **unitary nature of consciousness**, where a single, coherent experience dominates awareness at any given moment, is reflected in the thoughtseed framework by the *transient dominance* of a single thoughtseed on the Inner Screen. The dominant thoughtseed, acting as a pullback attractor, integrates information from multiple sources and establishes a Markov blanket, creating a temporary boundary that separates its internal processing from the broader cognitive landscape. The thoughtseed's core attractor, representing its most stable and probable state, shapes the content projected onto the Inner Screen, resulting in a unified and coherent conscious experience [3;24;25]. The selection of the dominant thoughtseed is a competitive process, influenced by factors such as sensory input, goals, and prior expectations, ensuring that the content of consciousness is continuously updated and refined to reflect the living system's current needs and priorities [99;128].

Meta-cognition: Influence of Higher-Order Thoughtseeds on Lower-Order Thoughtseeds

Higher-order representations, such as agent-level goals or intentions, exert top-down influence by modulating the precision (or weight) assigned to lower-level predictions. This influences attentional focus and the selection of the dominant thoughtseed, aligning with hierarchical predictive processing models where higher levels bias the interpretation of sensory information [14;71]. The modulation of precision allows the system to prioritize specific predictions, guiding attention and influencing the selection of the dominant thoughtseed that shapes conscious experience [32;115]. This process can be seen as metacognitive control, where higher-order thoughtseeds actively shape information flow and the emergence of conscious content on the Inner Screen [90; 143]. Meta-awareness further refines this process, representing the higher-order thoughtseed's awareness of its

own influence and can be modeled as a probabilistic belief about the accessibility of lower-level thoughtseeds and their associated KDs. It serves as a metacognitive monitoring and control mechanism [102;142].

Results

Illustrative Mathematical Framework

We present an illustrative mathematical framework for the Thoughtseed Framework, providing foundational modeling for the dynamics of entities within active inference and thoughtseed contexts. The specific implementation and complexity of the model may vary based on cognitive processes and brain regions. Further refinement, guided by empirical findings and computational modeling, is essential for understanding the role of the thoughtseed framework in shaping embodied cognition.

Detailed mathematical explanations are provided in Supplementary Section 8.3.
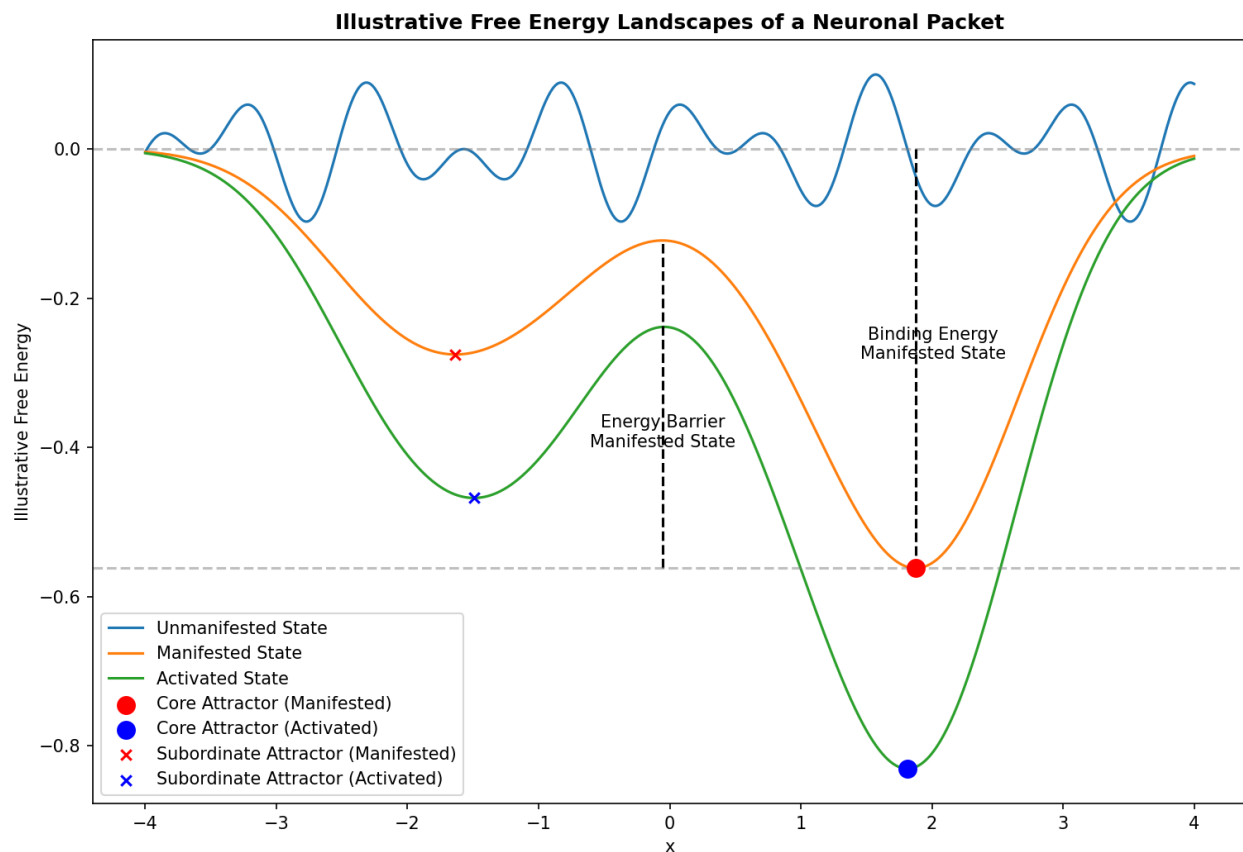
Layer 1: NPDs

Figure 3. Illustrative Free Energy Landscapes of a Neuronal Packet The figure depicts three potential states of a neuronal packet (NP). x-axis represents hypothetical internal states of the NP, while y-axis represents free energy associated with each state. Blue curve depicts an unmanifested **state**, characterized by a relatively flat landscape, indicating low stability and high susceptibility to change. Upon **phase transition** to **manifested state**, the NP is represented by the red curve, featuring a deep local minimum (**core attractor**) and a shallower local minimum (a **subordinate attractor**). The **energy barrier** shows the energy required to transition between the core attractor and the subordinate attractor. The **binding energy**, representing the overall stability of the manifested NP. The green curve illustrates activated state, where the core attractor is further deepened, reflecting heightened neural activity on the NP's core representation.

Neuronal Packets (NPs)

State Representation of an NP, ν denoted by state-vector $\mathbf{x}_\nu$, at time t can be represented as a combination of its attractor states, ψ_c:core attractor, and ψ_{s_i} :subordinate attractors and their corresponding activation levels $\alpha_c(t)$ and $\alpha_{s_i}(t)$.

$$\mathbf{x}_\nu(t) = \{(\psi_c, \alpha_c(t)), (\psi_{s_1}, \alpha_{s_1}(t)), ..., (\psi_{s_n}, \alpha_{s_n}(t))\} \tag{1}$$

Figure 3 shows the potential states of a NP.

Generative Model of a NP describes how it generates sensory observations $\mathbf{s}$, and *optionally* actions $\mathbf{a}$, conditioned on θ_ν (internal model parameters), $type_{NP}$(sensory, active, or internal), prior beliefs u, and the influence of the thoughtseed $\mathcal{T}$.

$$p(s, a, \mathbf{x}_\nu | \theta_\nu, type_{NP}, u, \mathcal{T}) \tag{2}$$

Free Energy Minimization

VFE of an NP can be expressed as:

$$VFE = F(\mathbf{x}_\nu, s, a, \theta_\nu, type_{NP}, u, \mathcal{T}) \tag{3.1}$$

$$VFE = \begin{cases} -\ln p(s|\mathbf{x}_\nu, \theta_\nu, type_{NP}, u, x_\tau) + D_{KL}[q(\mathbf{x}_\nu)||p(\mathbf{x}_\nu|s, u, \mathcal{T})] \\ -\ln p(s, a|\mathbf{x}_\nu, \theta_\nu, type_{NP}, u, x_\tau) + D_{KL}[q(\mathbf{x}_\nu)||p(\mathbf{x}_\nu|s, u, \mathcal{T})] \end{cases} \tag{3.2}$$

Superordinate Ensembles (SEs)

State Representation of an SE $\mathcal{E}$, denoted by $\mathbf{x}_\varepsilon(t)$, at time t can be represented as a vector of set of states of m NPs ν_k using an integration function $\mathcal{F}$:

$$\mathbf{x}_\varepsilon(t) = \mathcal{F}(\{\mathbf{x}_{\nu_k}(t)\}_{k=1}^m) \tag{4}$$

Generative Model of an SE is conditioned on the θ_ε (model parameters) and the active thoughtseed state $\mathcal{T}$, reflecting the potential top-down influence of the thoughtseed on the SE's internal dynamics and its interpretation of sensory information.

$$p(\mathbf{s}, \mathbf{x}_\varepsilon, \{\mathbf{x}_\nu\}_{\nu \in \varepsilon} | \theta_\varepsilon, \mathcal{T})$$

(5)

Free Energy Minimization

VFE of a SE can be expressed as:

$$VFE = F(\mathbf{x}_\varepsilon, \{\mathbf{x}_\nu\}_{\nu \in \varepsilon}, \mathbf{s}, \theta_\varepsilon, \mathcal{T})$$

(6.1)

$$VFE = -\ln p(\mathbf{s} | \mathbf{x}_\varepsilon, \{\mathbf{x}_\nu\}_{\nu \in \varepsilon}, \theta_\varepsilon, \mathcal{T})$$

$$+ D_{KL}[q(\mathbf{x}_\varepsilon, \{\mathbf{x}_\nu\}_{\nu \in \varepsilon}) || p(\mathbf{x}_\varepsilon, \{\mathbf{x}_\nu\}_{\nu \in \varepsilon} | \theta_\varepsilon, \mathcal{T})]$$

(6.2)

Neuronal Packet Domains (NPDs)

State Representation of an NPD N, denoted by state $\mathbf{x}_N$, at time *t* can be represented as a collection of the states of its constituent SEs $\mathbf{x}_{\varepsilon_i}(t)$:

$$\mathbf{x}_N(t) = \{\mathbf{x}_{\varepsilon_1}(t), \mathbf{x}_{\varepsilon_2}(t), ..., \mathbf{x}_{\varepsilon_k}(t)\}$$

(7)

Generative Model: The generative model of an NPD describes how its internal states (the states of its constituent SEs) generate sensory observations $\mathbf{s}$ and actions $\mathbf{a}$, conditioned on θ_N (internal model parameters of NPD) and the active thoughtseed state $\mathcal{T}$. It can be represented as

$$p(\mathbf{s}, \mathbf{a}, \mathbf{x}_N | \theta_N, \mathcal{T})$$

(8)

Free Energy Minimization: The VFE of a NPD can be expressed as:

$$VFE = F(\mathbf{x}_N, \mathbf{s}, \mathbf{a}, \theta_N, \mathcal{T})$$

(9.1)

$$VFE = -\ln p(\mathbf{s}, \mathbf{a}|\mathbf{x}_N, \theta_N, \mathcal{T}) + D_{KL}[q(\mathbf{x}_N)||p(\mathbf{x}_N|\theta_N, \mathcal{T})] \quad \text{(9.2)}$$

Layer 2: Knowledge Domains (KDs)

The illustrative mathematical formalism for KDs captures their role as integrative structures within a Knowledge Graph. An illustration of the dynamic interplay of NPDs and KDs is shown in Figure 4.
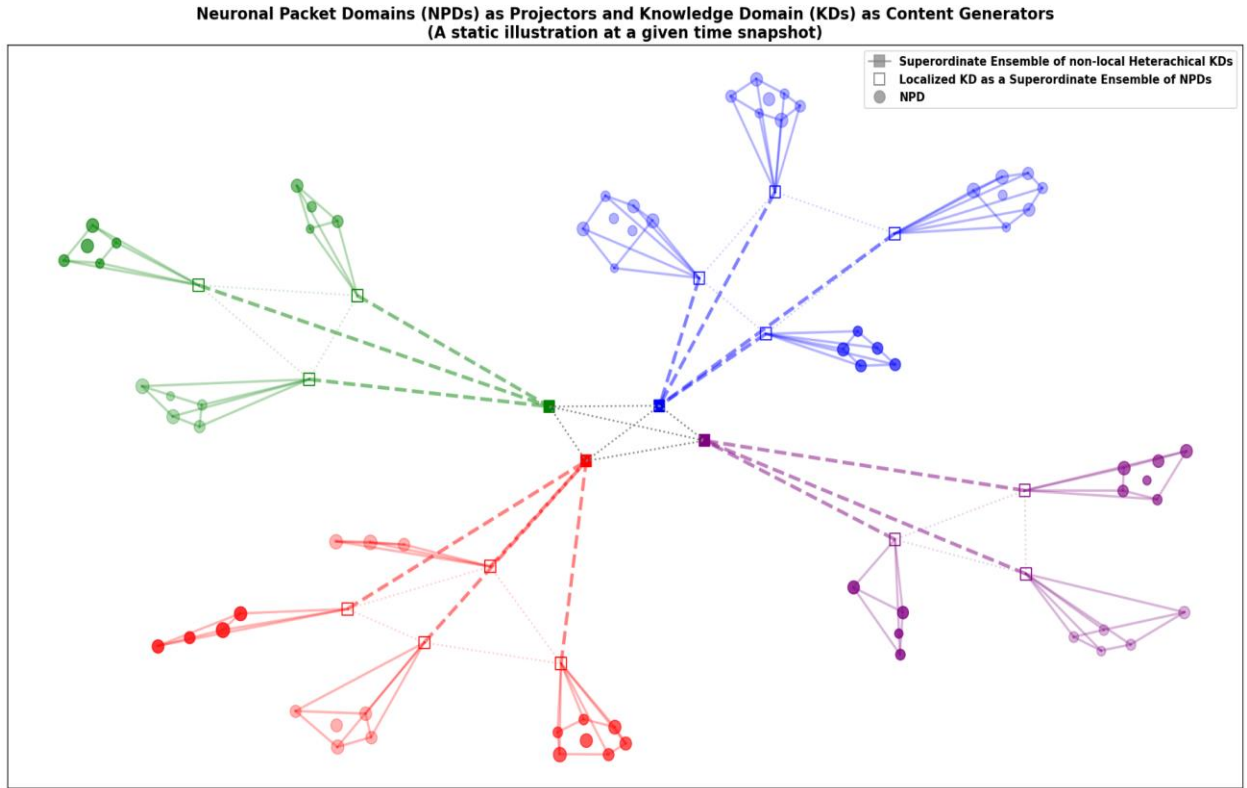

Figure 4. Interplay of Neuronal Packet Domains (NPDs) and **Knowledge Domains (KDs).** This figure illustrates the relationship between NPDs and KDs at a given time snapshot, highlighting their roles as "projectors" and "content generators," respectively, within the thoughtseed framework. **NPDs:** The circles depict NPDs, the functional units of the brain responsible for processing specific types of information. They act as "projectors," providing raw sensory data or generating potential actions. **SE of NPDs:** Unfilled squares symbolize superordinate ensembles (SEs) of NPDs that organize the "projectors" at nested scales. **KDs:** Represented as a filled square. Distinct colors represent different *localized KDs*, suggesting specialization within knowledge domains. KDs integrate and interpret information from multiple SEs of NPDs to generate meaningful and complex representations. **Heterarchical Organization of KDs:** The

presence of a higher-order KDs further suggests the *non-local heterarchical* organization of *KDs* across multiple domains. **Dynamic Interplay:** NPDs provide raw data, and KDs interpret and contextualize this information, contributing to the formation of thoughtseeds and shaping the "content of conscious" experience on the Inner Screen. **Nested holographic screens** can be inferred from the hierarchical organization depicted in the figure, with each level representing a different level of abstraction in the processing and representation of information.

Graph Representation of the Knowledge Graph, as a graph G_K with vertices V_K representing SEs of a specific Knowledge Domain K_i, and edges E_K representing the connections.

$$G_K = (V_K, E_K)$$
(10)

The state of a KD, denoted by $\mathbf{x}_K$, can be represented as a collection of the states of its constituent SEs $\mathbf{x}_{\varepsilon_i}$:

$$\mathbf{x}_K(t) = \left\{ \mathbf{x}_{\varepsilon_1}(t), \mathbf{x}_{\varepsilon_2}(t), ..., \mathbf{x}_{\varepsilon_l}(t) \right\}$$
(11)

Generative Model of a KD describes how its internal states $\{\mathbf{x}_\varepsilon\}_{\varepsilon \in K}$ (states of constituent SEs) aid in generating content $\mathbf{c}$ projected on the "Inner Screen". It encompasses sensory observations, knowledge, memories, and expertise associated within the domain, conditioned on θ_K (internal model parameters), thoughtseed state $\mathcal{T}$, reflecting top-down influence.

$$p\left(\mathbf{c}, \mathbf{x}_K, \{\mathbf{x}_\varepsilon\}_{\varepsilon \in K} \middle| \theta_K, \mathcal{T}\right)$$
(12)

Free Energy Minimization: The primary role of KDs lies in knowledge aggregation and integration, not generating actions. Through free energy minimization, KDs update their internal states (and its constituent SEs) ensuring the content projected onto the Inner Screen aligns with observed data conditioned upon the active thoughtseed. GFE can also be applied here, considering the expected future consequences of knowledge integration and alignment at longer durations.

$$VFE = F(\mathbf{x}_K, \{\mathbf{x}_\varepsilon\}_{\varepsilon \in K}, \mathbf{c}, \theta_K, \mathcal{T}) \tag{13.1}$$

$$VFE = -\ln p(\{\mathbf{x}_\varepsilon\}_{\varepsilon \in K}, \mathbf{c} | \mathbf{x}_K, \theta_K, \mathcal{T}) + D_{KL}[q(\mathbf{x}_K) || p(\mathbf{x}_K | \theta_K, \mathcal{T})] \tag{13.2}$$

Layer 3: Thoughtseed Network

The illustrative mathematical formalism for thoughtseeds within a TN is a complex process and is described as a series of steps.

Mapping Brain's Attractor Landscape to a Thoughtseed's Characteristic States

The brain's pullback attractor landscape, denoted by Ω is hypothesized to be shaped by the knowledge and beliefs encoded within the Knowledge Domains (KDs).

$$\phi_K : \mathbf{x}_K \to \Omega_K \quad \forall K \in \mathcal{K} \tag{14}$$

where the mapping function ϕ_K transforms the state of each Knowledge Domain (KD), represented by $\mathbf{x}_K$, into a corresponding attractor sub-landscape Ω_K.

$$\Omega = \Psi(\{\Omega_K\}_{K \in \mathcal{K}}) \tag{15}$$

where the overall attractor landscape Ω can be represented as the integration of these sub-landscapes Ω_K using a function Ψ.

Let thoughtseed $\mathcal{T}_i$ be represented via characteristic states χ_i.

$$\chi_i \subset \Omega \tag{16.1}$$

$$\chi_i = \chi_i^* \cup \chi_i^s \tag{16.2}$$

where the set of characteristic states χ_i are the regions or subsets within the brain's attractor landscape Ω, but will likely related with its corresponding Ω_K. The characteristic states χ_i can be further classified into a dominant attractor state χ_i^* and a set of subordinate attractor states χ_i^s, with the corresponding mapping functions ξ_i^* and ξ_i^s respectively.

$$\xi_i : \mathbf{x}_{\mathcal{T}_i}(t) \to \chi_i \tag{17.1}$$

$$\xi_i^* : \mathbf{x}_{\mathcal{T}_i}(t) \to \chi_i^* \tag{17.2}$$

$$\xi_i^s : \mathbf{x}_{\mathcal{T}_i}(t) \to \chi_i^s \tag{17.3}$$

Mapping of Thoughtseed Goals, Policies and Affordances

Thoughtseed Goals are described as a mapping g_i from the characteristic states $\chi_i(t)$ of a thoughtseed $\mathcal{T}_i$ and the states of the relevant KDs $\{\mathbf{x}_K(t)\}_{K \in \mathcal{K}_i}$ to its corresponding set of goals $\mathbf{G}_i(t)$.

$$g_i : \chi_i(t) \times \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}_i} \to \mathbf{G}_i(t) \tag{18}$$

Thoughtseed Policies: Policies $\mathbf{\Pi}_i$ for a thoughtseed $\mathcal{T}_i$ at time t, be described as a mapping h_i from the states of the relevant KDs $\{\mathbf{x}_K(t)\}_{K \in \mathcal{K}_i}$, to its corresponding set of goals $\mathbf{G}_i(t)$, and its affordances $\mathbf{A}_{epistemic}$ and $\mathbf{A}_{pragmatic}$.

$$h_i : \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}_i} \times \mathbf{G}_i(t) \times \mathbf{A}_{epistemic} \times \mathbf{A}_{pragmatic} \to \mathbf{\Pi}_i(t) \tag{19}$$

Thoughtseed Affordances: Epistemic affordances and Pragmatic affordances can be described using mapping functions k_e and k_p, respectively.

$$k_e : \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}_i} \times \mathbf{s} \to \mathbf{A}_{epistemic} \tag{20.1}$$

$$k_p : \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}_i} \times \mathbf{s} \times \mathbf{G}_i(t) \to \mathbf{A}_{pragmatic} \tag{20.2}$$

Generative Model of a Thoughtseed

Thoughtseed can generate content $\mathbf{c}$ projected onto the Inner Screen, sensory observations $\mathbf{s}$, actions $\mathbf{a}$, attentional modulation γ and meta-awareness state m. This generative model is conditioned on characteristic states χ_i, relevant KDs $\{\mathbf{x}_K\}_{K \in \mathcal{K}}$, their projected content $\{\mathbf{c}_K\}_{K \in \mathcal{K}}$, model parameters $\theta_{\mathcal{T}}$, prior beliefs $\mathbf{u}$, input saliency σ, available affordances $\mathbf{A}$, and selected policy $\mathbf{\Pi}_i$.

$$p(\mathbf{c}, \mathbf{s}, \mathbf{a}, \gamma, m | \chi_i, \{\mathbf{x}_K\}_{K \in \mathcal{K}}, \{\mathbf{c}_K\}_{K \in \mathcal{K}}, \theta_{\mathcal{T}}, \mathbf{u}, \sigma, \mathbf{A}, \mathbf{\Pi}_i)$$

(21)

VFE of thoughtseed can be described as:

$$VFE = -\ln p(\mathbf{c}, \mathbf{s}, \mathbf{a} | \chi_i, \{\mathbf{x}_K\}_{K \in \mathcal{K}}, \{\mathbf{c}_K\}_{K \in \mathcal{K}}, \theta_{\mathcal{T}}, \mathbf{u}, \sigma, \mathbf{A}, \mathbf{\Pi}_i)$$

(22)

$$+ D_{KL}[q(\chi_i, \gamma, m) \| p(\chi_i, \gamma, m | \mathbf{s}, \{\mathbf{x}_K\}_{K \in \mathcal{K}}, \{\mathbf{c}_K\}_{K \in \mathcal{K}}, \mathbf{u}, \sigma, \mathbf{A})]$$

Free Energies of a Thoughtseed

$$GFE_i(t) = VFE_i(t) + \mathbb{E}_{q(\mathbf{s}, \mathbf{a} | \mathbf{\Pi}_i(t))}[EFE_i(t+1)]$$

(23)

where:

- $GFE_i(t)$: The GFE of $\mathcal{T}_i$ at time t .
- $VFE_i(t)$: The VFE of $\mathcal{T}_i$, as defined in Equation (22).
- $\mathbb{E}_{q(\mathbf{s}, \mathbf{a} | \mathbf{\Pi}_i(t))}[EFE_i(t+1)]$: The EFE of future states, averaged over the predicted sensory states $\mathbf{s}$ and actions $\mathbf{a}$ under the policy $\mathbf{\Pi}_i(t)$ selected by the thoughtseed $\mathcal{T}_i$ at time t.

EFE of a thoughtseed, quantifying the anticipated "surprise" or uncertainty associated with a particular policy, both its epistemic and pragmatic affordances, can be described as:

$$EFE_i(\mathbf{\Pi}_i, t) = EFE_i^{epistemic}(\mathbf{\Pi}_i, t) + EFE_i^{pragmatic}(\mathbf{\Pi}_i, t)$$

(24.1)

$$EFE_i^{epistemic}(\mathbf{\Pi}_i, t) = \mathbb{E}_{q(\mathbf{s}, \mathbf{a} | \mathbf{\Pi}_i)}[-\ln p(\mathbf{c}, \mathbf{s}, \mathbf{a} | \chi_i, \{\mathbf{x}_K\}_{K \in \mathcal{K}}, \{\mathbf{c}_K\}_{K \in \mathcal{K}}, \theta_{\mathcal{T}}, \mathbf{u}, \sigma, \mathbf{A}_{\mathbf{epistemic}})]$$

(24.2)

$$EFE_i^{pragmatic}(\mathbf{\Pi}_i, t) = D_{KL}[q(\mathbf{s}, \mathbf{a} | \mathbf{\Pi}_i) \| p(\mathbf{s}, \mathbf{a} | \chi_i, \{\mathbf{x}_K\}_{K \in \mathcal{K}}, \{\mathbf{c}_K\}_{K \in \mathcal{K}}, \theta_{\mathcal{T}}, \mathbf{u}, \sigma, \mathbf{A}_{\mathbf{pragmatic}})]$$

(24.3)

Activation Level of a Thoughtseed

The activation level $\alpha_i(t)$ of a thoughtseed's internal states $\mathbf{x}_{\mathcal{T}_i}$ is hypothesized to be determined by the weighted sum of probabilities that the brain's state resides within its dominant or subordinate attractor states. Using mapping functions ξ_i^* or ξ_i^s respectively, reflecting the thoughtseed's overall prominence in the current cognitive landscape.

$$\alpha_i(t) = w^* \cdot p(\xi_i^*(\mathbf{x}_{\mathcal{T}_i}(t)), t) + \sum_{s \in \chi_i^s} w_s \cdot p(\xi_i^s(\mathbf{x}_{\mathcal{T}_i}(t)), t)$$

(25)

Activation Threshold Parameter and the Active Thoughtseeds Pool

An **activation threshold** $\Theta_{activation}(t)$ is hypothesized to serve as a gatekeeper for the Inner Screen, preventing the 'noise' of weakly activated thoughtseeds, ensuring that only the most pertinent and impactful thoughtseeds contribute to the ongoing stream of thought. The set of thoughtseeds whose activation levels $\alpha_i(t)$ surpass this threshold $\Theta_{activation}(t)$ constitutes the "**active thoughtseeds pool**," denoted by $\mathcal{P}_{active}(t)$ at time *t*.

$$\mathcal{P}_{\mathbf{active}}(t) = \{\mathbf{x}_{\mathcal{T}_i}(t) | \alpha_i(t) > \Theta_{activation}(t)\}$$

(26)

An illustrative example of a "dog" thoughtseed is discussed in Supplementary Section 9.5.

Layer 4: Metacognition: The Orchestration of Thoughtseed Dynamics and the Inner Screen

The illustrative mathematical formalism showcasing dominant thoughtseed dynamics and the unitary conscious experience on the Inner Screen is a complex process, and involves a series of steps.

Formulation of Goals, Policies and Affordances at Agent Level

Global Goals: The agent's goals $\mathbf{G}_{agent}(t)$ could be represented as a function $g_{agent}(t)$ of the characteristic states χ_i^* within the brain's pullback attractor landscape that are frequently revisited above a frequency threshold θ_{freq} and exhibit a certain degree of stability or persistence S_i.

$$g_{agent} : \{(\chi_i^*, S_i) | f_i > \theta_{freq}\} \rightarrow \mathbf{G}_{agent}(t) \tag{27}$$

Global Policies: Once the agent's goals are established, selection of appropriate policies $\mathbf{\Pi}_{agent}(t)$ to achieve these goals are guided by the knowledge and beliefs stored within the KDs, as well as current context and perceived affordances.

$$h_{agent} : \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}} \times \mathbf{G}_{agent}(t) \rightarrow \mathbf{\Pi}_{agent}(t) \tag{28}$$

The functions $g_{agent}(t)$ and h_{agent} encapsulate the complex processes involved in deriving goals and policies from the interplay of characteristic states, knowledge domains, and the organism's internal dynamics.

Global Affordances: The global affordances at an agent level $\mathbf{A}_{epistemic-agent}(t)$ and $\mathbf{A}_{pragmatic-agent}(t)$ are described via mapping functions $k_{e-agent}$ and $k_{p-agent}$ respectively.

$$k_{e-agent} : \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}} \times \mathbf{s} \rightarrow \mathbf{A}_{epistemic-agent}(t) \tag{29.1}$$

$$k_{p-agent} : \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}} \times \mathbf{s} \times \mathbf{G}_{agent}(t) \rightarrow \mathbf{A}_{pragmatic-agent}(t) \tag{29.2}$$

Free Energies at Agent-Level

Agent-level VFE: It is calculated as sum of VFEs $VFE_i(t)$ of active thoughtseeds in $\mathcal{P}_{\mathbf{active}}(t)$ plus other factors contributing to the agent's overall surprise at that instant.

$$VFE_{agent}(t) = \sum_{i \in \mathcal{P}_{\mathbf{active}}(t)} VFE_i(t) + \dots \tag{30}$$

Agent-level GFE: It is calculated as as sum of VFEs $VFE_i(t)$ and expected EFE $\mathbb{E}_{q(\mathbf{s},\mathbf{a}|\Pi_{agent}(t))}[EFE_{agent}(t+1)]$ under the agent's current policy $\mathbf{\Pi}_{agent}(t)$,

capturing *anticipated surprise* or uncertainty associated with predicted sensory observations $\mathbf{s}$ and potential actions $\mathbf{a}$.

$$GFE_{agent}(t) = VFE_{agent}(t) + \mathbb{E}_{q(\mathbf{s},\mathbf{a}|\Pi_{agent}(t))}[EFE_{agent}(t+1)] \quad \text{(31)}$$

Agent-level EFE: It is calculated as the sum of EFEs EFE_i of active thoughtseeds in $\mathcal{P}_{\mathbf{active}}(t)$ weighted by their respective activation levels $\alpha_i(t)$. It allows the agent to evaluate potential outcomes of different policies, considering both their epistemic and pragmatic values, and select the policy $\Pi_{agent}(t)$ to minimize overall surprise.

$$EFE_{agent}(\mathbf{\Pi}_{agent}, t) = \sum_{i \in \mathcal{P}_{\mathbf{active}}(t)} \alpha_i(t) \cdot EFE_i(\mathbf{\Pi}_i, t) \quad \text{(32.1)}$$

$$EFE_{agent}(\mathbf{\Pi}_{agent}, t) = \sum_{i \in \mathcal{P}_{active}(t)} \alpha_i(t) \cdot EFE_i^{epistemic}(\mathbf{\Pi}_i, t) + \sum_{i \in \mathcal{P}_{active}(t)} \alpha_i(t) \cdot EFE_i^{pragmatic}(\mathbf{\Pi}_i, t)$$

$$\text{(32.2)}$$

where $EFE_i^{epistemic}$ and $EFE_i^{pragmatic}$ are the epistemic and pragmatic components of the thoughtseed's EFE, respectively.

Selection of the Dominant Thoughtseed via EFE Minimization

It is chosen via a competitive process, where the thoughtseed with lowest cumulative EFE in the active thoughtseeds pool $\mathcal{P}_{active}$ over an arbitrary time period Δt is selected as the dominant thoughtseed, denoted by i^*.

$$i^*(t - \Delta t, t) = \underset{i \in \mathcal{P}_{\mathbf{active}}(t-\Delta t, t)}{\operatorname{argmin}} \sum_{t'=t-\Delta t}^{t} EFE_i(\mathbf{\Pi}_i, t') \quad \text{(33)}$$

We additionally hypothesize that EFE minimization is suitable for quick, reactive decisions where the primary concern is minimizing immediate prediction errors. Whereas, for long-term or deliberative planning GFE minimization could be a better candidate for dominant thoughtseed selection.

Unitary Nature of the Content on the "Inner Screen"

The content $\mathbf{c}_i^*$ on the Inner Screen $\mathcal{C}(t)$ at any moment is shaped a single, dominant thoughtseed reflecting the unitary nature of conscious experience.

$$\mathcal{C}(t) = \mathbf{c}_i^*(t - \Delta t, t)$$

(34)

Influence of higher-order thoughtseed on lower-order thoughtseeds

Higher-order thoughtseeds could adjust the attentional precision parameter γ, and meta-awareness parameter m influencing lower-level thoughtseeds. The influence of higher-order thoughtseed on lower-level thoughtseeds can be described as:

$$p(\gamma, \{\mathbf{x}_\tau\}_{\tau \in \mathcal{T}}, m, \mathbf{\Pi}, \mathbf{G} | \theta_m, \mathbf{u}, \{i \in \mathcal{P}_{\mathbf{active}}\})$$

(35)

Illustrative Cognitive Architecture

Formal Definition of a Thoughtseed

A **thoughtseed** is a *higher-order **transient Markov-blanketed*** construct with ***agency***, that ***generates*** the **"content of consciousness"** projected on the "Inner Screen." It emerges from the coordinated activity of superordinate ensembles (SEs) across different knowledge domains (KDs), enabling active exploration of the environment and the development of affordances—possibilities for action—that are associated with specific patterns of neuronal activity.

4-Layered Architecture of Internal States

The thoughtseed framework proposes a hierarchical organization of cognitive processes, spanning four interconnected layers as shown in Fig 5. One can imagine these nested levels as akin to "nested holographic screens," each encoding information about the world at increasing levels of abstraction [34;137].

1. **Neuronal Packet Domains (NPDs):** The foundational level, comprising interconnected groups of neurons that process sensory information and generate potential actions. NPDs are categorized into sensory, active, and internal state domains, contributing to higher-order representations.

2. **Knowledge Domains (KDs):** Large-scale heterarchical structures representing interconnected networks of concepts, categories, and relationships. KDs serve as repositories of knowledge, memories, and expertise, providing the conceptual scaffolding for interpreting sensory information and guiding behavior.

3. **Thoughtseeds Network:** Thoughtseeds emerge as dynamic entities from the coordinated activity of superordinate ensembles (SEs) across multiple KDs. They represent coherent patterns of neural activity associated with specific concepts, percepts, or potential actions, acting as sub-agents within the cognitive system.

4. **Meta-Cognition, Dominant Thoughtseed and Higher-Order Thoughtseeds:** This level encompasses higher-order cognitive processes like attentional control, self-awareness, and goal-directed behavior. Higher-order thoughtseeds, representing overarching goals or intentions, modulate lower-level thoughtseed activity and influence attentional focus, contributing to the formation of global policies that guide overall behavior.

The Inner Screen as the Locus of Conscious Experience

The "Inner Screen" serves as the locus of conscious experience, or the metaphorical stage where the dynamic "contents of consciousness" converge [34;137]. It's an active mental workspace where perception, knowledge, memories, and potential actions come together to form a cohesive and meaningful representation of the world. This representation is crucial for adaptive behavior and conscious awareness.

Thoughtseeds: The Actors on the Inner Screen

While NPDs act as specialized functional units, processing specific information, KDs interpret and contextualize this raw data, generating the content that shapes conscious

experience. Thus, NPDs can be seen as the "projectors" of sensory information, while KDs act as the "content generators" that provide meaning and context.

Thoughtseeds, the emergent agents of cognition, compete for dominance on the Inner Screen, shaping the content of conscious experience. They are **not** the literal "contents of consciousness" but rather the underlying processes that organize and generate those contents. Thoughtseeds can be likened to "actors" on the stage of the Inner Screen, actively creating and orchestrating the "scenes, dialogues, and actions" derived from Knowledge Domains (KDs) that together form the tapestry of conscious experience.

Higher-level goals and policies, situated at the agent level, can also be considered as a form of 'high-level' thoughtseed, influencing the selection and activation of lower-level thoughtseeds providing top-down guidance and modulating the flow of information within the thoughtseed network.

The Dynamics of Thoughtseed Selection

Thoughtseeds, acting as autonomous sub-agents, actively generate predictions about the world, including the content that could be projected onto the Inner Screen. They influence action selection to test those predictions and update their internal models based on sensory feedback. The thoughtseed that most effectively minimizes cumulative EFE of the agent, emerges as the dominant one, capturing the spotlight on the Inner Screen and shaping the agent's conscious experience and behavior.

The emergence of a dominant thoughtseed is a dynamic process influenced by bottom-up saliency, top-down attentional control, and the thoughtseed's own internal dynamics, represented as a pullback attractor within the brain's state space. The dominant thoughtseed establishes a Markov blanket, creating a temporary boundary that separates its internal processing from the broader cognitive landscape. This allows the thoughtseed to maintain computational and informational autonomy while actively engaging with the environment. The thoughtseed's core attractor, representing its most stable and probable state, shapes the content projected onto the Inner Screen.

The Thought Lifecycle and the Unitary Nature of Conscious Experience

Thoughtseeds, as dynamic entities within the cognitive landscape, exhibit a lifecycle characterized by transitions between *inactive, transitioning,* and *activated states.* The activated state, including when a thoughtseed is dominant, is metastable, allowing for flexible responses to environmental or internal changes. Thoughtseeds enter the activated state when their activation levels surpass a *global activation threshold*, and they join the *"active thoughtseeds pool."* The dominant thoughtseed is then selected from this active pool through a competitive process, primarily determined by the thoughtseed that minimizes cumulative expected free energy (EFE).

At any given moment, a single thoughtseed is **transiently** dominate the "stream of thoughts," representing the "content of consciousness." This unitary nature suggests that thoughtseeds act as *discrete* and *coherent Markov-blanketed entities*, even though they may emerge from the complex interaction of multiple SEs of KDs. The dominant thoughtseed, is *metastable*, and susceptible to change in response to new information, shifting goals, or fluctuations in internal dynamics.

The transient dominance of a single thoughtseed within an arbitrarily small time period contributes to the unitary nature of conscious experience, where a single, coherent experience dominates awareness. This ensures that the "content of consciousness" is continuously updated and refined to reflect the living system's current needs and priorities. It allows for flexible and adaptive behavior, enabling the organism to navigate its world effectively and make informed decisions in the face of uncertainty and change.
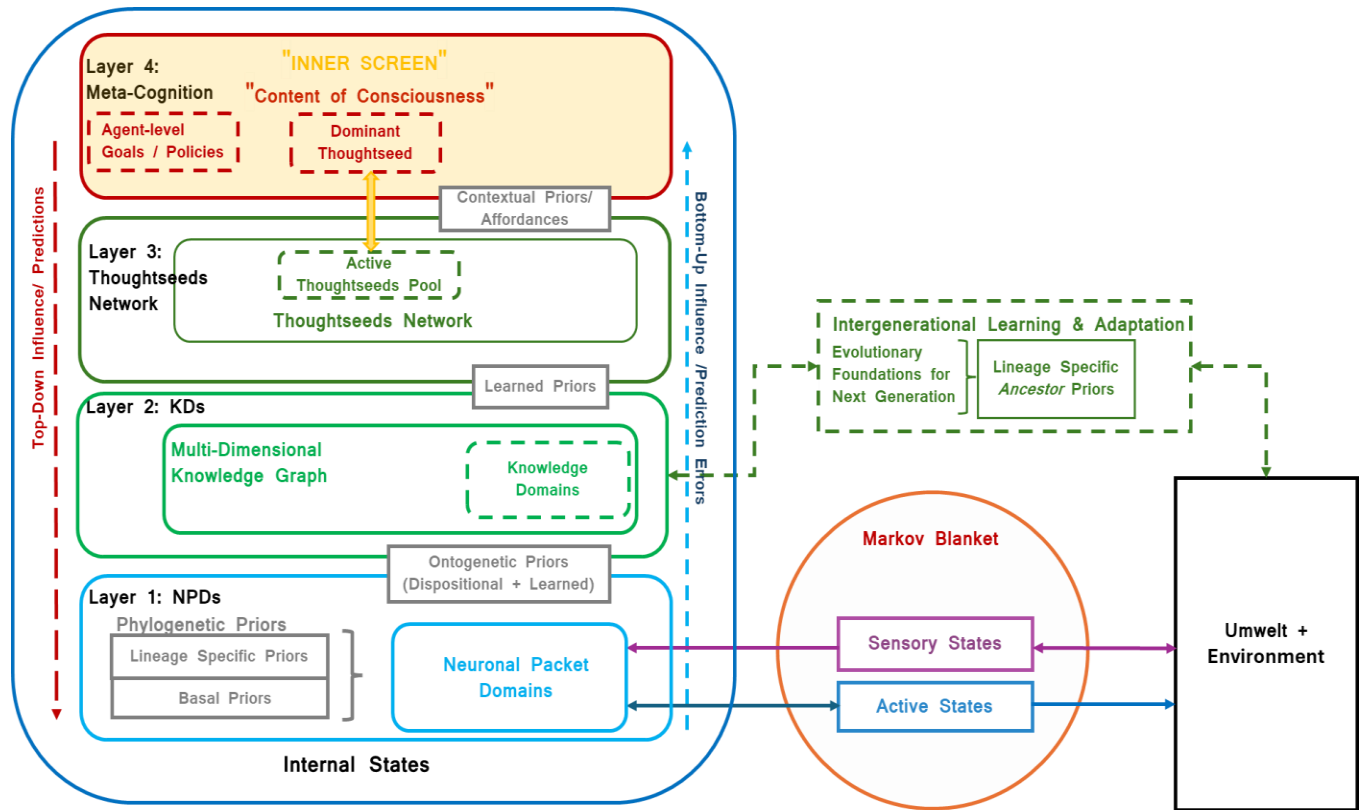
Fig 5. Illustrative Thoughtseed Framework. This diagram provides a high-level conceptual illustration of the **internal states** within the **four-layered Thoughtseed Framework**. It highlights the hierarchical organization of cognitive processes, from the foundational level of **Neuronal Packet Domains (NPDs)** and **Knowledge Domains(KDs)**, to the emergent layers of **Thoughtseed Network, Meta-cognition** and the **Inner Screen**. Each layer can be understood as a "nested holographic screen" encoding information about the world at increasing levels of abstraction: Dashed red lines represent top-down influences and predictions; and Dashed blue lines represent prediction errors and bottom-up influences. The role of priors and intergenerational learning in shaping the system's dynamics is currently speculative and requires further investigation. The interaction with the **Umwelt** (the organism's subjective world) and the broader environment occurs via the **Markov blanket** at the **Agent** level. The dominant thoughtseed, shapes the "content of consciousness," by actively generating predictions and guiding behavior.

5. Discussion

Towards a Computational Theory of Embodied Cognition

The thoughtseed framework, integrating concepts of neuronal packets, the Inner Screen, and the Free Energy Principle, offers a promising approach for a general theory of

embodied cognition [1;34;125;135;137;170;171;172]. It posits that thoughtseeds, as higher-order, agentic constructs emerging from the coordinated activity across knowledge domains (KDs), represent coherent and dynamic patterns of neural activity that embody concepts or percepts. These thoughtseeds are shaped by both evolutionary priors and the organism's ongoing experiences, enabling active exploration of the environment and the development of affordances. The content of these dynamic thoughtseeds is hypothesized to be projected onto the Inner Screen, shaping the organism's conscious experience and guiding its perception, action, and decision-making. The framework emphasizes the deep interconnectedness between an organism's internal states, its actions, and the environment it interacts with. It suggests that cognition is not merely an internal process but is deeply rooted in the organism's bodily interactions with the world, aligning with the embodied/enactive view of cognition [160].

This framework aligns with the computational theory of mind (CTM) [129;35;138], proposing that mental computations can be understood as emergent processes arising from the interactions of thoughtseeds. Conceived as the fundamental units of thought within the framework, thoughtseeds could be seen as the computational building blocks that the brain employs to implement complex cognitive functions such as decision-making, planning, and problem-solving. This suggests that the thoughtseed framework could provide a mechanistic explanation for how the brain realizes the computations that underlie these higher-order cognitive processes.

Bridging the Gap Between Neuroscience and Contemplative Traditions

The thoughtseed framework aims to offer new perspectives into understanding the nature of meditation and its advanced stages, which are often described in various traditions that can seem disparate in their approaches and techniques [94;160;39]. The framework proposes that these traditional approaches represent distinct trajectories along specific knowledge domains, potentially crystallizing the internal generative model in different ways, yet ultimately converging towards similar goals of achieving equanimity and tranquility of the mind. The systematic investigation of various meditative stages across

different traditions, including Buddhist and Yogic/Advaitic traditions, could be instrumental in generating testable predictions and offering a mechanistic understanding of these transformative experiences [8;156;17].

The framework's hierarchical organization, with deeper layers representing more abstract concepts, may provide a neural basis for the progressive refinement of attention and awareness in meditation. Specifically, the stabilization of core attractors within thoughtseeds could be seen as analogous to the deepening of concentration and focus in meditation, while the emergence of higher-order thoughtseeds might correspond to the development of more refined and integrated states of awareness. Investigating thoughtseed dynamics could potentially provide a hierarchical and progressive map of these states across different contemplative traditions, elucidating the mechanisms underlying these transformative experiences.

Future goals include mechanistically describing and simulating the fundamental processes underlying Krishnamurti's teachings, particularly his emphasis on non-doing and the dissolution of the observer-observed duality ("Thinker is the thought," "Observer is the Observed") [83] and the attentive sleepful state [84]. Further goals include, developing simulations in *altered states of consciousness* [157] like lucid dreaming [29]; and advanced meditative states like *jhanas* in Buddhist literature [2], *samādhis* in Yogic literature [163; 155], *Yoga Nidra* [133;79], and the "Clear Light of Sleep" [168].

Creating a Framework for Computational Phenomenology

The thoughtseed framework has the potential to serve as a reference for developing a formal framework in "computational phenomenology" [142] for *asking the right set of questions* across a diverse group of people, including advanced meditators and neuroscientists. This framework could facilitate a new approach in investigating the neural correlates and computational mechanisms underlying advanced meditative states and contemplative practices. By bridging the gap between subjective experiences and objective measurements, this approach could lead to a deeper understanding of the transformative potential of meditation and its implications for mental well-being and cognitive enhancement.

Limitations and Future Research Directions

Investigating the neural mechanisms underlying thoughtseed formation and dynamics requires a multi-scale approach, from the initial assembly of co-activated neurons to the emergence of stable Markov blankets embodying distinct attractor states. [172;91;68;135]. The development of biologically plausible computational models [6;119], like spiking neural networks, incorporating synaptic plasticity and homeostatic mechanisms, could plausibly simulate the self-organization of thoughtseeds and the evolution of attractor dynamics, potentially revealing neural signatures like spatiotemporal activity patterns or connectivity motifs[12;40;151].

However, empirically identifying thoughtseeds presents significant challenges. The inherent metastability of thoughtseed dynamics, characterized by rapid state transitions, makes isolating the core attractor representing the core knowledge or behavior difficult [132;162;22]. The hierarchical nature of thoughtseeds, with nested Markov blankets, further complicates this process [53;82]. Additionally, the inherent opacity of the Markov blanket [45;53; 54] at the agent level obscures the internal generative model responsible for behavior, making direct observation of thoughtseed dynamics challenging. A thorough understanding of the specific cognitive domain is crucial to guide experimental design and analysis, avoiding misinterpretations due to the complexities of mapping distributed neural activity to specific cognitive processes [21;141].

The framework's applicability can be explored through simulations of cognitive development, such as object permanence - the understanding that objects persist even when out of sight, is a cornerstone of cognitive development [127]. The emergence of thoughtseed-like mechanisms in infants during this developmental stage could offer valuable insights into the formation and dynamics of stable mental representations. Additionally, investigating the neural underpinnings of numerical cognition in toddlers, a cognitive domain closely linked to the development of abstract concepts and symbolic representation [107], could further elucidate the role of thoughtseeds in higher-order cognitive processes. These initial simulations would serve as a crucial bridge between the theoretical constructs of the thoughtseed framework and empirical observations,

providing a mechanistic demonstration of how the framework can be applied to understand a wider range of cognitive abilities.

6. Conclusion

The thoughtseed framework provides a novel, biologically-grounded model for understanding the organization and emergence of embodied cognition. By integrating insights from diverse fields, including evolutionary biology, dynamical systems theory, active inference, and contemplative practices, it offers a unified account of cognitive phenomena, from basic physiological regulation to higher-order thought processes.

While this is an initial step, the framework offers a promising direction for future research. Detailed simulations of specific cognitive phenomena and the identification of thoughtseed-like patterns in neuroimaging studies are crucial next steps. Further investigation into the role of evolutionary priors and their influence on thoughtseed mechanisms is also needed, which can aid in designing effective phenomenological experiments. The thoughtseed framework has the potential to provide a unified account of cognitive phenomena, from basic physiological regulation to higher-order thought processes, and may bridge the gap between neuroscience and contemplative traditions, paving the way for the development of more adaptive and intelligent systems.

7. Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

8. Author Contributions

Conceptualization and Mathematical Formalism, Original Drafting: Prakash; Editing: All authors

9. Supplementary Section

9.1. Evolutionary Priors: Detailed Descriptions and Learning Dynamics

Phylogenetic Priors

Phylogenetic priors represent inherited knowledge and biases from an organism's evolutionary lineage [66], shaping its internal model of the world [96]. Research suggests these priors influence both broad brain architecture and specific computations within neural networks [54;124]. Although the precise mechanisms underlying their encoding and expression remain an active area of study, they likely involve a complex interplay of genetic, epigenetic, and developmental processes.

Within the thoughtseed framework, phylogenetic priors play a critical role in shaping the formation and organization of neuronal packet domains (NPDs). These NPDs, in turn, constrain the development of knowledge domains (KDs) and the emergent thoughtseed network. Phylogenetic priors bias the system toward representations and behaviors that have historically proven adaptive. The strength and specificity of these priors can vary depending on the cognitive domain and the evolutionary pressures that have shaped it.

- **Basal Priors (B):** Fundamental biological needs and core behaviors conserved across broad groups (e.g., autopoietic processes, threat responses) [159;89]. They are thought to influence both unconscious and conscious processes, including instincts, unconscious biases, and autonomic functions crucial for survival [114;130;73]. These priors exhibit slow evolutionary change, forming the basis for survival and essential physiological functions.
- **Lineage-Specific Priors (L):** Adaptations and strategies successful for a specific lineage within its niche, influenced by genetic, epigenetic, and cultural inheritance [164]. These may evolve more rapidly in response to environmental changes, exemplified by bird songs [98], and primate tool use [169]. Lineage-specific priors

can guide the development of complex behaviors that are adapted to a particular ecological context.

Ontogenetic Priors

Ontogenetic priors encapsulate an individual's unique contribution to knowledge, reflecting their specific experiences and learning throughout their lifetime [75;95]. These priors are more flexible and adaptable compared to phylogenetic priors. Within the thoughtseed framework, ontogenetic priors significantly influence the development and refinement of knowledge domains (KDs).

- **Dispositional Priors (D):** Unique to each organism, these priors arise from genetic variation and developmental experiences, representing predispositions towards certain behaviors or learning styles. For instance, some individuals may exhibit a dispositional prior towards risk-taking, while others may be more cautious.

- **Learned Priors (λ):** These dynamic priors evolve throughout an individual's life, shaped by their interactions with the environment. They encapsulate "lessons learned" [154;147], encoding successful strategies and avoiding negative outcomes. A child learning to ride a bicycle exemplifies how learned priors can be acquired and refined over time. These priors represent the most flexible and adaptable aspect of the thoughtseed framework, reflecting the ongoing plasticity of the neural substrate.

The interplay between phylogenetic and ontogenetic priors within the thoughtseed framework is dynamic. Phylogenetic priors guide the acquisition of ontogenetic priors, continuously updating beliefs based on environmental interactions. From synaptic refinement to KD reorganization and new thoughtseed emergence, the internal model adapts, allowing increasingly accurate predictions and adaptive behavior. Ontogenetic priors provide flexibility in novel situations, while phylogenetic priors preserve essential knowledge. KD refinement contributes to goal, policy, and affordance development, shaping the organism's internal model.

The internal model, guiding perception, action, and learning, is fundamentally shaped by the knowledge and beliefs encoded within the KDs, a connection that can be mathematically represented by conditioning generative models on the states of relevant KDs.

Illustrative Mathematical Model for Evolutionary Priors

The complex interplay of evolutionary priors shapes an organism's internal model. We can capture this using a probabilistic approach that accounts for both the quasi-hierarchical structure and dynamic nature of these priors. A Bayesian factor graph representation could effectively capture these intricate interactions and dependencies.

- **Basal Priors (B):** Modeled as a vector of probability distributions, $B = [P(B_1), P(B_2), ..., P(B_n)]$, where each $P(B_i)$ represents the probability distribution over possible states for the *i*-th basal prior. These priors reflect fundamental needs and core behaviors, plausibly encoded in the genome and exhibiting slow evolutionary change.

- **Lineage-Specific Priors (L):** Represented as a vector of conditional probability distributions, $L = [P(L_1|B), P(L_2|B), ..., P(L_m|B)]$, conditioned on the basal priors. These priors encompass adaptations and behavioral strategies successful for a specific lineage.

- **Dispositional Priors (D):** Modeled as a vector of conditional probability distributions, $D = [P(D_1|L), P(D_2|L), ..., P(D_k|L)]$, conditioned on both basal and lineage-specific priors. These priors are unique to each organism, arising from genetic variation and developmental experiences, representing predispositions towards certain behaviors or learning styles.

- **Learned Priors (λ):** Represented as a vector of probability distributions, $\lambda = [P(\lambda_1|D, E_1), P(\lambda_2|D, E_2), ..., P(\lambda_j|D, E_k)]$ where E_k represents individual experiences. These dynamic priors evolve throughout an organism's lifetime, shaped by their interactions with the environment, encapsulating "lessons learned."

Evolutionary priors shape the organism's internal states, including the thoughtseed network. Each thoughtseed, as a component of μ, is influenced by this multi-layered prior structure, guiding adaptive behaviors that minimize surprise. To model the complex interplay of evolutionary priors within this generative model, we can utilize modeling tools like Forney Factor Graphs or Bayesian Graphs (FFGs) [38;92]. FFGs provide a flexible representation of the hierarchical and potentially non-linear relationships between priors (B, L, D, λ) and their influence on the thoughtseed network. This graphical representation visualizes the intricate relationships between priors and facilitates efficient message-passing algorithms for inference and learning.

9.2 Free Energy Principle "Thing" formulation

Within the Free Energy Principle (FEP), the states of a "thing" [56] are partitioned into sensory (s), active (a), internal (μ), and external (η) states.

Crucially, the internal states house the *Thoughtseed Network (TN)*, a collection of interconnected thoughtseeds as shown in Fig 1.

The Markov Blanket formalizes the separation between a "thing's" internal and external states as being conditionally independent given the blanket states.

$$\mu \perp \eta | s, a \Leftrightarrow p(\eta, \mu | s, a) = p(\eta | s, a) * p(\mu | s, a)$$ (S.1)

$$P(\mu, s, a) = P(\mu | s, a) * P(s | a) * P(a | \mu)$$ (S.2)

where (1) posits that once the blanket states (sensory and active) are known, knowing the external states provides no additional information about the internal states (including Thoughtseed Network), and vice versa. Thus, (2) describes the joint probability of particular states (μ, a, s) independently of external states (η). Thereby, *Thoughtseed Network TN ⊆ Internal states*, can be described only using particular states.

As a "thing," an organism strives to minimize surprise—unexpected sensory input that contradicts its internal model of the world. This surprise is mathematically quantified as variational free energy (VFE). In this context, surprise refers to the unexpectedness of

sensory input, given the "thing's" prior beliefs. Formally, it is defined as the negative logarithm of the probability of the observation:

$$S = ln(P(s))$$
(S.3)

VFE is calculated using variational inference techniques and represents the discrepancy between the "thing's" expected and actual sensory states, given its current beliefs and actions:

$$VFE = E_q[ln(q(\mu)) - ln(p(\mu, s))]$$
(S.4)

where:

- E_q is the expectation under the variational distribution $q(\mu)$
- $q(\mu)$ is the variational approximation of the posterior distribution over the thoughtseed's internal states μ
- $p(\mu, s)$ is the true joint probability distribution of the internal states and sensory input.

Crucially, VFE provides an upper bound on surprisal, meaning that:

$$VFE \geq S$$
(S.5)

By minimizing VFE, the agent indirectly minimizes surprise, aligning its internal model with the external world [46]. This is the driving force behind active inference, where agents actively seek out information and engage in actions that are expected to reduce their prediction errors. This process of surprise minimization guides the formation, activation, and adaptation of thoughtseeds within the Thoughtseed Network (TN), ensuring that the organism's internal model remains a reliable and efficient guide for behavior.

9.3 Detailed Explanations of Equations

$$\mathbf{x}_\nu(t) = \{(\psi_c, \alpha_c(t)), (\psi_{s_1}, \alpha_{s_1}(t)), ..., (\psi_{s_n}, \alpha_{s_n}(t))$$
(1)

where:

- ψ_c: Core attractor state.

- $\alpha_c(t)$: Activation level of the core attractor at time t
- ψ_{s_i}: i-th subordinate attractor state
- α_{s_i}: Activation level of the i-th subordinate attractor at time t

$$p(s, a, \mathbf{x}_\nu | \theta_\nu, type_{NP}, u, \mathcal{T})$$ (2)

where:

- s: The sensory observations generated by the NP
- a: The actions generated by the NP
- $\mathbf{x}_\nu$: The state vector of the NP
- θ_ν: The parameters of the NP's generative model. These parameters determine how the NP interprets sensory input and generates predictions about the world.
- $type_{NP}$: a is generated only if NP is of type 'active'
- u: Prior beliefs of NP
- $\mathcal{T}$: State of the active thoughtseed. It can also be represented by $\mathbf{x}_\mathcal{T}$.

$$VFE = F(\mathbf{x}_\nu, s, a, \theta_\nu, type_{NP}, u, \mathcal{T})$$ (3.1)

$$VFE = \begin{cases} -\ln p(s|\mathbf{x}_\nu, \theta_\nu, type_{NP}, u, x_\tau) + D_{KL}[q(\mathbf{x}_\nu)||p(\mathbf{x}_\nu|s, u, \mathcal{T})] \\ -\ln p(s, a|\mathbf{x}_\nu, \theta_\nu, type_{NP}, u, x_\tau) + D_{KL}[q(\mathbf{x}_\nu)||p(\mathbf{x}_\nu|s, u, \mathcal{T})] \end{cases}$$ (3.2)

where:

- The first term represents the **surprise** or **negative log-likelihood** of the sensory input, given the NP's internal states, model parameters, type, prior beliefs, and the influence of the thoughtseed. Minimizing VFE drives perception and learning by adjusting internal states and parameters to better predict and explain sensory observations. In the case of $type_{NP}$ is "active" (the lower part of Equation 3.2), minimizing VFE influences action selection, by selecting actions that minimize surprise and lead to expected sensory outcomes.

- The second term is the *KL divergence* between the *approximate posterior belief* $q(\mathbf{x}_\nu)$ and *prior belief* $p(\mathbf{x}_\nu|s, u, \mathcal{T})$ of the NP, also conditioned on the thoughtseed's state. Minimizing this discrepancy improves the accuracy of the generative model's predictions.

$$\mathbf{x}_\varepsilon(t) = \mathcal{F}(\{\mathbf{x}_{\nu_k}(t)\}_{k=1}^{m})$$ (4)

where:

- $\mathbf{x}_\varepsilon(t)$: The state of the SE $\mathcal{E}$ at time t, represented as a vector.
- $\mathcal{F}$: A function that integrates or combines the states of the constituent NPs.
- $\{\mathbf{x}_{\nu_k}(t)\}_{k=1}^{m}$: The set of states of all NPs ν_k that belong to the SE $\mathcal{E}$ at time t.
- *m*: The number of NPs within the SE.

$$p(\mathbf{s}, \mathbf{x}_\varepsilon, \{\mathbf{x}_\nu\}_{\nu \in \varepsilon}|\theta_\varepsilon, \mathcal{T})$$ (5)

where:

- $\mathbf{s}$: The sensory input received by the SE.
- $\{\mathbf{x}_\nu\}_{\nu \in \varepsilon}$: The set of states of all NPs ($\mathcal{V}$) that belong to the SE $\mathcal{E}$.
- θ_ε: The parameters of the generative model for the SE.

$$VFE = F(\mathbf{x}_\varepsilon, \{\mathbf{x}_\nu\}_{\nu \in \varepsilon}, \mathbf{s}, \theta_\varepsilon, \mathcal{T})$$ (6.1)

$$VFE = -\ln p(\mathbf{s}|\mathbf{x}_\varepsilon, \{\mathbf{x}_\nu\}_{\nu \in \varepsilon}, \theta_\varepsilon, \mathcal{T})$$
$$+ D_{KL}[q(\mathbf{x}_\varepsilon, \{\mathbf{x}_\nu\}_{\nu \in \varepsilon})||p(\mathbf{x}_\varepsilon, \{\mathbf{x}_\nu\}_{\nu \in \varepsilon}|\theta_\varepsilon, \mathcal{T})]$$ (6.2)

where the right side has two terms:

- The first term represents the surprise or negative log-likelihood of the sensory input $\mathbf{s}$, given the SE's internal state $\mathbf{X}_{\varepsilon}$, the states of its constituent NPs $\mathbf{X}_{\nu}$ the SE's model parameters θ_{ε}, and the influence of the active thoughtseed $\mathcal{T}$.
- The second term, is the Kullback-Leibler (KL) divergence between:

    ○ The approximate posterior distribution $q\left(\mathbf{x}_{\varepsilon}, \{\mathbf{x}_{\nu}\}_{\nu \in \varepsilon}\right)$ representing the SE's beliefs about its own state and the states of its constituent NPs.

    ○ The prior distribution $p\left(\mathbf{x}_{\varepsilon}, \{\mathbf{x}_{\nu}\}_{\nu \in \varepsilon} \mid \theta_{\varepsilon}, \mathcal{T}\right)$, which encapsulates its prior expectations, also conditioned on the thoughtseed state $\mathcal{T}$.

Thus, SE aims to minimize VFE by updating its internal states and the inferred states of its NPs to better explain the sensory input (reducing surprise) while staying close to its prior expectations (minimizing the KL-divergence).

$$\mathbf{x}_N(t) = \left\{\mathbf{x}_{\varepsilon_1}(t), \mathbf{x}_{\varepsilon_2}(t), ..., \mathbf{x}_{\varepsilon_k}(t)\right\} \tag{7}$$

where:

- $\mathbf{x}_{\varepsilon_i}(t)$ represents the state of the i-th SE within the NPD at time t.
- k is the number of SEs within the NPD.

$$p(\mathbf{s}, \mathbf{a}, \mathbf{x}_N \mid \theta_N, \mathcal{T}) \tag{8}$$

where:

- $\mathbf{s}$: The sensory input received by the NPD.
- $\mathbf{a}$: The actions generated by the NPD.
- $\mathbf{X}_N$: The internal states of the NPD (the collection of SE states).
- θ_N: The parameters of the generative model for the NPD, which capture its prior beliefs about the causal relationships between its internal states, sensory inputs, and actions.

- $\mathcal{T}$: The state of the active thoughtseed, which can exert top-down influence on the NPD's generative model.

$$VFE = F(\mathbf{x}_N, \mathbf{s}, \mathbf{a}, \theta_N, \mathcal{T})$$ (9.1)

$$VFE = -\ln p(\mathbf{s}, \mathbf{a}|\mathbf{x}_N, \theta_N, \mathcal{T}) + D_{KL}[q(\mathbf{x}_N)||p(\mathbf{x}_N|\theta_N, \mathcal{T})]$$ (9.2)

where:

- The first term represents the *surprise* or *negative log-likelihood* of the sensory input $\mathbf{s}$ and actions $\mathbf{a}$, given the NPD's internal states $\mathbf{x}_N$ and model parameters θ_N, and modulated by the thoughtseed's state $\mathcal{T}$.

- The second term is the *KL divergence* between the approximate posterior distribution $q(\mathbf{x}_N)$ over the NPD's internal states and its prior distribution $p(\mathbf{x}_N|\theta_N, \mathcal{T})$, which is also conditioned on the thoughtseed's state $\mathcal{T}$.

$$G_K = (V_K, E_K)$$ (10)

where:

- V_K is the set of vertices representing the Superordinate Ensembles (SEs) within the KD.
- E_K is the set of edges representing the connections between these SEs.
- The strength or weight of these connections can be captured in a weight matrix.

$$\mathbf{x}_K(t) = \{\mathbf{x}_{\varepsilon_1}(t), \mathbf{x}_{\varepsilon_2}(t), ..., \mathbf{x}_{\varepsilon_l}(t)\}$$ (11)

where:

- $\mathbf{x}_{\varepsilon_i}(t)$ represents the state of the i-th SE within the KD at time *t*.
- *l* is the number of SEs within the KD.

$$p(\mathbf{c}, \mathbf{x}_K, \{\mathbf{x}_\varepsilon\}_{\varepsilon \in K} | \theta_K, \mathcal{T})$$ (12)

where:

- $\mathbf{c}$: The content projected onto the Inner Screen, representing the knowledge, memories, and expertise associated with the KD, K
- $\{\mathbf{x}_\varepsilon\}_{\varepsilon \in K}$: The set of states of all SEs ($\mathcal{E}$) that belong to the KD K
- θ_K: The parameters of the generative model for the KD

$$VFE = F(\mathbf{x}_K, \{\mathbf{x}_\varepsilon\}_{\varepsilon \in K}, \mathbf{c}, \theta_K, \mathcal{T})$$ (13.1)

$$VFE = -\ln p(\{\mathbf{x}_\varepsilon\}_{\varepsilon \in K}, \mathbf{c} | \mathbf{x}_K, \theta_K, \mathcal{T}) + D_{KL}[q(\mathbf{x}_K) || p(\mathbf{x}_K | \theta_K, \mathcal{T})]$$ (13.2)

where:

- The first term represents the surprise or negative log-likelihood of the states of the constituent SEs $\{\mathbf{x}_\varepsilon\}_{\varepsilon \in K}$ and the content $\mathbf{c}$ on the Inner Screen, given the KD's internal states $\mathbf{x}_K$, model parameters θ_K, and the influence of active thoughtseed $\mathcal{T}$.
- The second term is the KL divergence between the approximate posterior distribution $q(\mathbf{x}_K)$ over the KD's internal states and its prior distribution $p(\mathbf{x}_K | \theta_K, \mathcal{T})$, conditioned on the thoughtseed state $\mathcal{T}$.

$$\phi_K : \mathbf{x}_K \to \Omega_K \quad \forall K \in \mathcal{K}$$ (14)

where the mapping function ϕ_K transforms the state of each Knowledge Domain (KD), represented by $\mathbf{x}_K$, into a corresponding attractor sub-landscape Ω_K.

$$\Omega = \Psi(\{\Omega_K\}_{K \in \mathcal{K}})$$ (15)

where the overall attractor landscape Ω can be represented as the integration of these sub-landscapes Ω_K using a function Ψ.

$$\chi_i \subset \Omega \tag{16.1}$$

$$\chi_i = \chi_i^* \cup \chi_i^s \tag{16.2}$$

where the set of characteristic states χ_i are the regions or subsets within the brain's attractor landscape Ω. The characteristic states χ_i can be further classified into a dominant attractor state χ_i^* and a set of subordinate attractor states χ_i^s

$$\xi_i : \mathbf{x}_{\mathcal{T}_i}(t) \rightarrow \chi_i \tag{17.1}$$

$$\xi_i^* : \mathbf{x}_{\mathcal{T}_i}(t) \rightarrow \chi_i^* \tag{17.2}$$

$$\xi_i^s : \mathbf{x}_{\mathcal{T}_i}(t) \rightarrow \chi_i^s \tag{17.3}$$

where:

- ξ_i: Maps the thoughtseed's $\mathcal{T}_i$ internal state vector $\mathbf{x}_{\mathcal{T}_i}$ at time t to a specific characteristic states χ_i.
- ξ_i^* Maps the thoughtseed's internal state to its dominant attractor state χ_i^*.
- ξ_i^s: Maps the thoughtseed's internal state to one of its subordinate attractor states within the set χ_i^s

$$g_i : \chi_i(t) \times \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}_i} \rightarrow \mathbf{G}_i(t) \tag{18}$$

where:

- g_i: The mapping function specific to thoughtseed $\mathcal{T}_i$
- $\chi_i(t)$: The characteristic state of thoughtseed $\mathcal{T}_i$ at time t
- $\{\mathbf{x}_K(t)\}_{K \in \mathcal{K}_i}$: The set of states of all KDs relevant to thoughtseed $\mathcal{T}_i$ at time t
- $\mathbf{G}_i(t)$: The set of goals associated with thoughtseed $\mathcal{T}_i$ at time t

$$h_i : \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}_i} \times \mathbf{G}_i(t) \times \mathbf{A}_{epistemic} \times \mathbf{A}_{pragmatic} \rightarrow \mathbf{\Pi}_i(t) \tag{19}$$

where:

- h_i: The mapping function specific to thoughtseed $\mathcal{T}_i$

- $\mathbf{A}_{epistemic}$: Represents the epistemic affordances

- $\mathbf{A}_{pragmatic}$: Represents the pragmatic affordances

$$k_e : \{\mathbf{x}_K(t)\}_{K\in\mathcal{K}_i} \times \mathbf{s} \to \mathbf{A}_{epistemic} \tag{20.1}$$
$$k_p : \{\mathbf{x}_K(t)\}_{K\in\mathcal{K}_i} \times \mathbf{s} \times \mathbf{G}_i(t) \to \mathbf{A}_{pragmatic} \tag{20.2}$$

where:

- k_e: The mapping function for the epistemic affordances
- k_p: The mapping function for the pragmatic affordances
- $\mathbf{s}$: The sensory inputs

$$p(\mathbf{c}, \mathbf{s}, \mathbf{a}, \gamma, m | \chi_i, \{\mathbf{x}_K\}_{K\in\mathcal{K}}, \{\mathbf{c}_K\}_{K\in\mathcal{K}}, \theta_{\mathcal{T}}, \mathbf{u}, \sigma, \mathbf{A}, \mathbf{\Pi}_i) \tag{21}$$

where:

- $\mathbf{c}$: The content projected onto the Inner Screen, representing the percept, concept, or idea associated with the thoughtseed
- $\mathbf{s}$: The sensory input received by the agent, potentially integrated across multiple modalities
- $\mathbf{a}$: The actions selected by the thoughtseed, influencing the environment and subsequent sensory input
- γ: Attentional precision parameter, modulating the influence of sensory information
- m: Meta-awareness state, reflecting the thoughtseed's self-awareness
- χ_i: The characteristic states associated with the thoughtseed $\mathcal{T}_i$
- $\{\mathbf{x}_K\}_{K\in\mathcal{K}}$: The set of states of all KDs (K) that the thoughtseed interacts with
- $\{\mathbf{c}_K\}_{K\in\mathcal{K}}$: The set of contents projected onto the Inner Screen from all the KDs that the thoughtseed interacts with
- $\theta_{\mathcal{T}}$: The parameters of the generative model for the thoughtseed
- $\mathbf{u}$: Prior beliefs or expectations

- σ: Saliency of the sensory input
- $\mathbf{A}$: Set of possible actions afforded by the Umwelt/environment
- $\mathbf{\Pi}_i$: Selected Policy

$$VFE = -\ln p(\mathbf{c}, \mathbf{s}, \mathbf{a}|\chi_i, \{\mathbf{x}_K\}_{K\in\mathcal{K}}, \{\mathbf{c}_K\}_{K\in\mathcal{K}}, \theta_{\mathcal{T}}, \mathbf{u}, \sigma, \mathbf{A}, \mathbf{\Pi}_i) \quad (22)$$

$$+D_{KL}[q(\chi_i, \gamma, m)||p(\chi_i, \gamma, m|\mathbf{s}, \{\mathbf{x}_K\}_{K\in\mathcal{K}}, \{\mathbf{c}_K\}_{K\in\mathcal{K}}, \mathbf{u}, \sigma, \mathbf{A})]$$

where:

- First term represents *surprise* or *negative log-likelihood* of the potentially observed variables ($\mathbf{c}$, $\mathbf{s}$, and $\mathbf{a}$) *given* the characteristic states χ_i, relevant KDs $(\mathbf{x}_K, \mathbf{c}_K)$, model parameters $\theta_{\mathcal{T}}$, prior beliefs $\mathbf{u}$, saliency σ, affordances $\mathbf{A}$ and selected policy $\mathbf{\Pi}_i$.

- Second term is the *KL divergence* between the *approximate posterior distribution* $q(\chi_i, \gamma, m)$ over χ_i, γ, and m and its *prior distribution* $p(\chi_i, \gamma, m|\mathbf{s}, \{\mathbf{x}_K\}_{K\in\mathcal{K}}, \{\mathbf{c}_K\}_{K\in\mathcal{K}}, \mathbf{u}, \sigma, \mathbf{A})$, conditioned on $\mathbf{s}$, states of relevant KDs $(\mathbf{x}_K, \mathbf{c}_K)$, $\mathbf{u}$, σ and $\mathbf{A}$.

$$GFE_i(t) = VFE_i(t) + \mathbb{E}_{q(\mathbf{s},\mathbf{a}|\mathbf{\Pi}_i(t))}[EFE_i(t+1)] \quad (23)$$

where:

- $GFE_i(t)$: The generalized free energy of thoughtseed $\mathcal{T}_i$ at time t.
- $VFE_i(t)$: The variational free energy of thoughtseed at time , as defined in Equation (23).
- $\mathbb{E}_{q(\mathbf{s},\mathbf{a}|\mathbf{\Pi}_i(t))}[EFE_i(t+1)]$: The expected free energy of future states, averaged over the predicted sensory states ($\mathbf{S}$) and actions ($\mathbf{a}$) under the policy $\mathbf{\Pi}_i(t)$ selected by the thoughtseed $\mathcal{T}_i$ at time t.

$$EFE_i(\mathbf{\Pi}_i, t) = EFE_i^{epistemic}(\mathbf{\Pi}_i, t) + EFE_i^{pragmatic}(\mathbf{\Pi}_i, t) \quad (24.1)$$

$$EFE_i^{epistemic}(\mathbf{\Pi}_i, t) = \mathbb{E}_{q(\mathbf{s},\mathbf{a}|\mathbf{\Pi}_i)}[-\ln p(\mathbf{c}, \mathbf{s}, \mathbf{a}|\chi_i, \{\mathbf{x}_K\}_{K \in \mathcal{K}}, \{\mathbf{c}_K\}_{K \in \mathcal{K}}, \theta_{\mathcal{T}}, \mathbf{u}, \sigma, \mathbf{A}_{\mathbf{epistemic}})]$$

(24.2)

$$EFE_i^{pragmatic}(\mathbf{\Pi}_i, t) = D_{KL}[q(\mathbf{s}, \mathbf{a}|\mathbf{\Pi}_i)||p(\mathbf{s}, \mathbf{a}|\chi_i, \{\mathbf{x}_K\}_{K \in \mathcal{K}}, \{\mathbf{c}_K\}_{K \in \mathcal{K}}, \theta_{\mathcal{T}}, \mathbf{u}, \sigma, \mathbf{A}_{\mathbf{pragmatic}})]$$

(24.3)

where:

- The EFE_i Equation 24.1 shows the decomposition of expected free energy into its epistemic and pragmatic components.
- The $EFE_i^{epistemic}$ Equation 24.2 quantifies the expected surprise or information gain associated with a policy, reflecting the thoughtseed's drive to explore and reduce uncertainty.
- The $EFE_i^{pragmatic}$ Equation 24.3 measures the divergence between the predicted and prior outcomes under a policy, reflecting the thoughtseed's preference for actions that lead to desired goals.

$$\alpha_i(t) = w^* \cdot p(\xi_i^*(\mathbf{x}_{\mathcal{T}_i}(t)), t) + \sum_{s \in \chi_i^s} w_s \cdot p(\xi_i^s(\mathbf{x}_{\mathcal{T}_i}(t)), t)$$

(25)

where:

- w^*: Weight associated with the dominant attractor state.
- w^s: Weight associated with the s-th subordinate attractor state.
- $p(\xi_i^*(\mathbf{x}_{\mathcal{T}_i}(t)), t)$: Probability of the brain's state being in the dominant attractor state of thoughtseed at time , as mapped by the function ξ_i^*.
- $p(\xi_i^s(\mathbf{x}_{\mathcal{T}_i}(t)), t)$: Probability of the brain's state being in the s-th subordinate attractor state of thoughtseed at time , as mapped by the function ξ_i^s.

$$\mathcal{P}_{\mathbf{active}}(t) = \{\mathbf{x}_{\mathcal{T}_i}(t)|\alpha_i(t) > \Theta_{activation}(t)\}$$

(26)

This equation states that the active pool consists of the internal states $\mathbf{x}_{\mathcal{T}_i}$ of thoughtseeds whose activation levels $\alpha_i(t)$ exceed the activation threshold $\Theta_{activation}(t)$.

$$g_{agent} : \{(\chi_i^*, S_i) | f_i > \theta_{freq}\} \to \mathbf{G}_{agent}(t) \tag{27}$$

where:

- $\mathbf{G}_{agent}(t)$: The set of global goals of the agent at time t.

- $g_{agent}(t)$: A function that maps the set of frequently revisited and stable characteristic states to the agent's global goals.

- S_i Represents the stability measure (e.g., dwell time) of the characteristic state χ_i^* above a certain threshold.

- f_i: The frequency of visitation of the characteristic state

- θ_{freq}: The frequency threshold for determining frequently revisited states

$$h_{agent} : \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}} \times \mathbf{G}_{agent}(t) \to \mathbf{\Pi}_{agent}(t) \tag{28}$$

where:

- $\mathbf{\Pi}_{agent}(t)$: The set of global policies or action plans of the agent at time t

- h_{agent}: A function that maps the states of all KDs and the agent's global goals to its overall policy.

- $\{\mathbf{x}_K(t)\}_{K \in \mathcal{K}}$: The set of states of all KDs at time t

$$k_{e-agent} : \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}} \times \mathbf{s} \to \mathbf{A}_{epistemic-agent}(t) \tag{29.1}$$

$$k_{p-agent} : \{\mathbf{x}_K(t)\}_{K \in \mathcal{K}} \times \mathbf{s} \times \mathbf{G}_{agent}(t) \to \mathbf{A}_{pragmatic-agent}(t) \tag{29.2}$$

where:

- $\mathbf{A}_{epistemic-agent}(t)$: Set of global epistemic affordances (opportunities for information gain) for the agent at time t.

- $\mathbf{A}_{pragmatic-agent}(t)$: Set of global pragmatic affordances (opportunities for goal achievement) for the agent at time t.

- $k_{e-agent}$ and $k_{p-agent}$: The mapping functions that determine global epistemic and pragmatic affordances, respectively.

- $\{\mathbf{x}_K(t)\}_{K\in\mathcal{K}}$: The set of states of all KDs at time t.

- $\mathbf{S}$: The current sensory input.

- $\mathbf{G}_{agent}(t)$: The set of global goals of the agent at time t.

$$VFE_{agent}(t) = \sum_{i\in\mathcal{P}_{\mathbf{active}}(t)} VFE_i(t) + \ldots$$

(30)

where:

- VFE_{agent}: The mapping function that calculates the agent-level variational free energy at a given time t.

- $\sum_{i\in\mathcal{P}_{\mathbf{active}}(t)} VFE_i(t)$: The sum of the variational free energies of all active thoughtseeds at time t.

- … : Represents potential additional terms that capture other factors contributing to the agent's overall surprise or uncertainty.

$$GFE_{agent}(t) = VFE_{agent}(t) + \mathbb{E}_{q(\mathbf{s},\mathbf{a}|\Pi_{agent}(t))}[EFE_{agent}(t+1)]$$

(31)

where:

- GFE_{agent}: The mapping function that calculates the agent-level generalized free energy at a given time t.

- $\mathbb{E}_{q(\mathbf{s},\mathbf{a}|\Pi_{agent}(t))}[EFE_{agent}(t+1)]$: The expected free energy of future states, averaged over the predicted sensory states $\mathbf{s}$ and actions $\mathbf{a}$ under the agent's policy $\Pi_{agent}(t)$ at time t.

$$EFE_{agent}(\mathbf{\Pi}_{agent}, t) = \sum_{i\in\mathcal{P}_{\mathbf{active}}(t)} \alpha_i(t) \cdot EFE_i(\mathbf{\Pi}_i, t)$$

(32.1)

where:

- $EFE_{agent}(\mathbf{\Pi}_{agent}, t)$: Expected free energy of the agent at time t under the policy $EFE_{agent}(\mathbf{\Pi}_{agent}, t)$.

- $\displaystyle\sum_{i\in\mathcal{P}_{\text{active}}(t)} \alpha_i(t) \cdot EFE_i(\mathbf{\Pi}_i, t)$: The sum of the EFEs of all thoughtseeds in the active thoughtseeds pool, weighted by their respective activation levels $\alpha_i(t)$. EFE equations of a thoughtseed are defined in Equation 24.1, 24.2, 24.3 and activation level of a thoughtseed is defined in Equation 25.

$$EFE_{agent}(\mathbf{\Pi}_{agent}, t) = \sum_{i\in\mathcal{P}_{active}(t)} \alpha_i(t)\cdot EFE_i^{epistemic}(\mathbf{\Pi}_i, t) + \sum_{i\in\mathcal{P}_{active}(t)} \alpha_i(t)\cdot EFE_i^{pragmatic}(\mathbf{\Pi}_i, t)$$

$$(32.2)$$

where $EFE_i^{epistemic}$ and $EFE_i^{pragmatic}$ are the epistemic and pragmatic components of the thoughtseed's EFE, respectively.

$$i^*(t - \Delta t, t) = \underset{i\in\mathcal{P}_{\text{active}}(t-\Delta t, t)}{\operatorname{argmin}} \sum_{t'=t-\Delta t}^{t} EFE_i(\mathbf{\Pi}_i, t')$$

$$(33)$$

- i^*: Index of the dominant thoughtseed within the time window $[t - \Delta t, t]$

- $\displaystyle\underset{i\in\mathcal{P}_{\text{active}}(t-\Delta t, t)}{\operatorname{argmin}} \sum_{t'=t-\Delta t}^{t} EFE_i(\mathbf{\Pi}_i, t')$: Calculates the cumulative Expected Free Energy (EFE) for each thoughtseed i within the set of active thoughtseed pool $\mathcal{P}_{active}$ over the specified time window Δt, considering the thoughtseed's policy $\mathbf{\Pi}_i$. The dominant thoughtseed is then selected as the one with the lowest cumulative EFE, reflecting the thoughtseed that is expected to lead to the least surprise or uncertainty over that period.

$$\mathcal{C}(t) = \mathbf{c}_i^*(t - \Delta t, t)$$

$$(34)$$

where:

- $\mathcal{C}(t)$: The content on the Inner Screen at time t

- $\mathbf{c}_i^*$: Content associated with the dominant thoughtseed $\mathcal{T}_i^*$ within the time window Δt. The identification of the dominant thoughtseed index i is described in Equation 33.

$$p(\gamma, \{\mathbf{x}_\tau\}_{\tau \in \mathcal{T}}, m, \mathbf{\Pi}, \mathbf{G} | \theta_m, \mathbf{u}, \{i \in \mathcal{P}_{\text{active}}\}) \qquad (35)$$

where:

- γ: Attentional Precision Parameter, which modulates the precision of the predictions of lower-level thoughtseeds.

- $\{\mathbf{x}_\tau\}_{\tau \in \mathcal{T}}$: The set of internal states of all lower-level thoughtseeds that are competing for attention.

- m: The meta-awareness state of the higher-order thoughtseed, reflecting its awareness of its own influence on the lower-level thoughtseeds and its role in shaping attentional focus.

- $\mathbf{\Pi}$: The set of policies or action plans considered by the higher-order thoughtseed.

- $\mathbf{G}$: The goals or desired outcomes that the higher-order thoughtseed aims to achieve.

- θ_m: The parameters of the generative model for the higher-order thoughtseed.

- $\mathbf{u}$: Prior beliefs or expectations, shaped by evolutionary and learned priors.

- $\{i \in \mathcal{P}_{\text{active}}\}$: Thoughtseeds within the set of active thoughtseed pool $\mathcal{P}_{active}$

9.4 Illustrative Example: The Visual Object Recognition KD

The visual object recognition KD serves as an illustration of the complex nature of KDs and their relationship with NPDs. It not only encompasses the hierarchical processing [173;139] of visual features from basic sensory inputs to complex object representations but also potentially integrates information from other modalities, such as auditory and tactile information, as well as semantic and emotional associations stored in memory.

This multi-modal integration allows for a rich and nuanced understanding of objects, going beyond their mere visual appearance. For instance, recognizing a dog could involve not only processing its visual features but also associating it with its characteristic sounds (barking), tactile sensations (fur), and emotional connotations (loyalty, companionship).

- **NPDs as Functional Units:** Within the visual object recognition KD, multiple NPDs are hypothesized to operate as specialized processing units, each contributing to the construction of a holistic visual experience. The primary visual cortex (V1) could act as an early visual NPD, extracting basic features like edges and orientations from the retinal input. The ventral visual stream, encompassing areas like V4 and the lateral occipital complex (LOC), could function as a shape-processing NPD, integrating these features into representations of shapes and objects [63]. The inferotemporal cortex (IT) could act as a higher-order object recognition NPD, further combining shape information with other features to recognize complex objects like faces (FFA) or tools (EBA) [77]. The NPDs may employ a sparse representation of visual information, leading to efficient and robust encoding of visual features [110;67].

- **KD as Integrative Framework and the Inner Screen:** The visual object recognition KD itself serves as the integrative framework that binds the outputs from these diverse NPDs into a coherent and meaningful percept. It acts as the interpretive and contextual processing layer, organizing and making sense of the "images" projected by the NPDs onto the Inner Screen. The specific content on the Inner Screen is dynamically shaped by the interplay between the activated KDs, the organism's goals and expectations, and the saliency of environmental stimuli.

  This "binding process" that leads to the emergence of a coherent percept on the Inner Screen could involve the synchronization of neural activity across different NPDs [148;112], the formation of superordinate ensembles that encompass multiple NPs, and the modulation of attentional mechanisms to prioritize relevant information. The formation of these ensembles or assemblies might be facilitated by the enhancement of firing rates in specific neurons, as suggested by Roelfsema

[140], leading to the binding of information across different NPDs. The dynamic and context-dependent nature of this process allows for flexible and adaptive perception, enabling the organism to respond effectively to the ever-changing demands of its environment.

- **Nested Markov Blankets and Holographic Representation:** Each NP and SE within the visual object recognition KD is hypothesized to possess its own Markov blanket, creating a nested hierarchy of representations. This nested structure, reminiscent of "nested holographic screens," allows for the integration and exchange of information across multiple levels of abstraction. The information encoded at each level is progressively coarse-grained, moving from fine-grained sensory details at lower levels to more abstract and categorical representations at higher levels [34].

9.5 Illustrative Example: "Dog" Thoughtseed

The "dog" thoughtseed serves as an illustrative example of how thoughtseeds emerge and function within thoughtseed framework. It represents an individual's integrated knowledge, beliefs, and associated behaviors related to dogs [103] encapsulating a complex and multifaceted concept within the brain's internal model.

A Generative Model Perspective

Scenario: You are walking in a park, surrounded by various stimuli, including trees, birds chirping, and other people. Suddenly, you spot a Golden Retriever playing fetch.

Level 1: Neuronal Packet Domains (NPDs) - The Projectors

- **Sensory NPDs:**
  - **Visual NPD:** The sight of the Golden Retriever could trigger the activation of various NPs within the visual NPD, potentially encoding features like its golden fur, wagging tail, and playful movements. These visual features might then be projected onto the Inner Screen. The activation of these NPs could also lead to the formation of superordinate ensembles (SEs)

representing more complex visual features, such as the dog's overall shape or its facial expressions.

- ○ **Auditory NPD:** The sound of the dog barking could activate NPs within the auditory NPD, potentially projecting the auditory percept onto the Inner Screen.
- ○ **Other Sensory NPDs:** If you were to pet the dog or smell it, the somatosensory and olfactory NPDs would likely also become active, potentially projecting their respective sensory information onto the Inner Screen.

- **Active State NPDs:**
  - ○ **Motor NPDs:** The sight of the dog playing fetch might trigger potential actions within the motor NPDs, such as reaching out to pet the dog or throwing a ball. These potential actions could be represented on the Inner Screen as imagined movements or motor plans.

- **Internal State NPDs:**
  - ○ These NPDs are hypothesized to process and integrate information from the sensory and active NPDs, contributing to the formation of higher-order representations within relevant KDs. For example, they might recognize the combination of visual and auditory features as indicative of a "dog" and activate the corresponding SEs within the "animal" KD.

Level 2: Knowledge Domains (KDs) - The Content Generators

- **"Animal" KD:** This KD is thought to contain SEs representing various animals, their characteristics, and behaviors. The activation of the "dog" SE within this KD could provide rich conceptual content about dogs, their typical actions (e.g., barking, playing), and their relationship to humans (e.g., as pets). This information might be projected onto the Inner Screen, contributing to the meaningful interpretation of the sensory input.
- **Other Relevant KDs:**

- ○ **"Pet" KD:** This KD could provide additional context about the dog's social role and potential interactions, influencing the observer's expectations and possible actions.
  - ○ **"Emotion" KD:** This KD might contribute to the affective experience associated with the dog, generating feelings of joy, excitement, or perhaps even fear, depending on the observer's past experiences.
  - ○ **"Memory" KD:** This KD might activate memories of past encounters with dogs, further enriching the representation of the current experience and influencing predictions about the dog's behavior.

Level 3: Thoughtseed - The Pullback Attractor and Agent

- **"Dog" Thoughtseed ($X_{\mathcal{T}}$):** The "dog" thoughtseed is hypothesized to emerge as a dominant attractor state within the brain's dynamic landscape, integrating perceptual, conceptual, and emotional information from activated KDs and NPDs. It represents a unified and meaningful concept of "dog," encompassing its various attributes, behaviors, and associated experiences. The thoughtseed's internal states, corresponding to characteristic states within the attractor landscape, dynamically evolve as the organism interacts with the dog and its environment.
- **Agency and the Inner Screen:** The "dog" thoughtseed, as an active agent, influences the content and dynamics of the Inner Screen. It generates predictions about the dog's behavior, such as expecting it to continue playing fetch or approach the observer. These predictions are compared to the actual sensory input, potentially leading to updates in the thoughtseed's internal model and the refinement of its representation on the Inner Screen. The thoughtseed also actively explores and evaluates affordances, selecting actions that align with its goals and minimize free energy.
- **Attentional Modulation:** The thoughtseed could also exert top-down control over attention, focusing the spotlight on specific aspects of the dog or its environment. For example, it might prioritize the visual processing of the dog's movements or the auditory processing of its barks, depending on the observer's goals and interests.

- **Competition and Selection:** The "dog" thoughtseed competes with other potential thoughtseeds for dominance on the Inner Screen. Its selection as the dominant thoughtseed is determined by its Expected Free Energy (EFE), which considers both its epistemic and pragmatic affordances. The thoughtseed's saliency, relevance to the current context, and alignment with the organism's goals also influence its chances of being selected. The emergence of the "dog" thoughtseed as the dominant one reflects its ability to provide the most accurate and parsimonious explanation for the current sensory input and the organism's goals, thus minimizing free energy.
- **Action Selection:** The active "dog" thoughtseed might guide the selection of actions towards the dog. These actions, such as petting the dog or calling its name, are aimed at fulfilling the thoughtseed's predictions and minimizing free energy.

Level 4: Meta-Cognitive Level

- **Higher-Order Thoughtseeds:** Higher-order thoughtseeds, representing goals or intentions (e.g., "I want to play with the dog"), can further modulate the activity of the "dog" thoughtseed and influence its policy selection. They can exert top-down control over attentional mechanisms, influencing the focus of the "spotlight" on the "Inner Screen" and the salience of different aspects of the "dog" thoughtseed. This allows for flexible and adaptive behavior, as the organism can prioritize certain aspects of the "dog" concept based on its current goals and intentions.
- **Attentional Control:** These higher-order thoughtseeds can also exert top-down control over attentional mechanisms, influencing the focus of the spotlight on the Inner Screen and the salience of different aspects of the "dog" thoughtseed.

Stream of Thoughts:

The continuous emergence, competition, and transition of thoughtseeds create the dynamic stream of thoughts. As the observer interacts with the dog and the environment, different thoughtseeds may become activated, reflecting the changing focus of attention and the ongoing process of active inference. For example, the initial "dog" thoughtseed might transition to a "play" thoughtseed as the observer engages in a game of fetch, or

to a "fear" thoughtseed if the dog suddenly barks aggressively. This dynamic interplay of thoughtseeds on the Inner Screen reflects the organism's continuous attempts to make sense of its environment, predict future events, and select actions that minimize surprise and achieve its goals.

Illustrative Generative Model for the "Dog" Thoughtseed

To illustrate the hierarchical generative model of the "dog" thoughtseed, we present the following equations as a simplified example, incorporating the concepts of NPDs, KDs, and the Inner Screen. Note that this formalization is not exhaustive and serves to demonstrate the key principles of the framework.

Level 1: Neuronal Packet Domains (NPDs) - The Projectors

Sensory NPDs
Visual NPD: Generative Model

$$p\left(\mathbf{s}_{visual}, \{\mathbf{x}_\nu\}_{\nu \in N_{visual}} \middle| \theta_{visual}, \mathcal{T}\right) \tag{S.6}$$

where:

- $\mathbf{s}_{visual}$: Visual sensory input related to the dog
- $\{\mathbf{x}_\nu\}_{\nu \in N_{visual}}$: Internal states of NPs within the visual NPD
- θ_{visual}: Parameters of the generative model for the visual NPD encapsulate its prior beliefs about how visual features are generated and organized
- $\mathcal{T}$: The state of the active thoughtseed ("dog" in this case)

Auditory NPD: Generative Model

$$p\left(\mathbf{s}_{auditory}, \{\mathbf{x}_\nu\}_{\nu \in N_{auditory}} \middle| \theta_{auditory}, \mathcal{T}\right) \tag{S.7}$$

where:

- $\mathbf{s}_{auditory}$: Auditory sensory input related to the dog
- $\{\mathbf{x}_\nu\}_{\nu \in N_{auditory}}$: Internal states of NPs within the auditory NPD
- $\theta_{auditory}$: Parameters of the generative model for the auditory NPD

- $\mathcal{T}$: The state of the active thoughtseed ("dog" in this case)
- γ : The attentional precision parameter

Other Sensory NPDs:

Similar generative models can be defined for other sensory modalities (e.g., olfactory, somatosensory) if relevant to the "dog" concept.

Active State NPDs

Motor NPDs:Generative Model

$$p(\mathbf{a}_{motor}, \{\mathbf{x}_\nu\}_{\nu \in N_{motor}} | \theta_{motor}, \mathcal{T})$$
(S.8)

where:

- $\mathbf{a}_{motor}$: Potential actions towards the dog
- $\{\mathbf{x}_\nu\}_{\nu \in N_{motor}}$: Internal states of NPs within the motor NPD
- θ_{motor}: Parameters of the generative model for the motor NPD
- $\mathcal{T}$: The state of the active thoughtseed

Internal State NPDs: Generative Model

$$p(\{\mathbf{x}_K\}_{K \in \mathcal{K}}, \{\mathbf{x}_\nu\}_{\nu \in N_{internal}} | \theta_{internal}, \mathcal{T})$$
(S.9)

where:

- $\{\mathbf{x}_K\}_{K \in \mathcal{K}}$: States of the relevant KDs (e.g., "animal," "pet," "emotion") that contextually interact with the KD associated with the NPD.
- $\{\mathbf{x}_\nu\}_{\nu \in N_{internal}}$: Internal states of NPs within the internal state NPD
- $\theta_{internal}$: Parameters of the generative model for the internal state NPD
- $\mathcal{T}$: The state of the active thoughtseed

Level 2: Knowledge Domains (KDs) - The Content Generators

"Animal" KD: Generative Model

$$p(\mathbf{x}_{animal}, \{\mathbf{x}_\varepsilon\}_{\varepsilon \in K_{animal}}, \{\mathbf{x}_N\}_{N \in \mathcal{N}_{animal}} | \theta_{animal}, \mathcal{T})$$ (S.10)

where:

- $\mathbf{x}_{animal}$: Internal states of the "animal" KD
- $\{\mathbf{x}_\varepsilon\}_{\varepsilon \in K_{animal}}$: States of the SEs within the "animal" KD, including the "dog" SE
- $\{\mathbf{x}_N\}_{N \in \mathcal{N}_{animal}}$: States of the NPDs that contribute to the "animal" KD
- θ_{animal}: Parameters of the generative model for the "animal" KD
- $\mathcal{T}$: The state of the active thoughtseed

Other Relevant KDs:

- Similar generative models can be defined for other relevant KDs (e.g., "pet," "emotion," "memory").

Level 3: Thoughtseed - The Pullback Attractor and Agent

"Dog" Thoughtseed: Generative Model

$$p(\mathbf{c}, \mathbf{s}, \mathbf{a}, \mathbf{x}_\mathcal{T}, \gamma, m | \{\mathbf{x}_K\}_{K \in \mathcal{K}}, \{\mathbf{c}_K\}_{K \in \mathcal{K}}, \theta_\mathcal{T}, \mathbf{u}, \sigma, \mathbf{A})$$ (S.11)

where:

- $\mathbf{c}$: The content projected onto the Inner Screen, representing the conscious experience of the "dog" concept. This could include visual imagery, auditory sensations, emotional feelings, and associated memories related to dogs.
- $\mathbf{s}$: Sensory inputs from various modalities (visual, auditory, etc.) that are relevant to the "dog" concept.
- $\mathbf{a}$: Actions taken by the organism in response to the "dog" thoughtseed, such as petting the dog, calling its name, or running away.
- $\mathbf{x}_\mathcal{T}$: The internal states of the "dog" thoughtseed, representing its current beliefs, expectations, and predictions about the dog and its environment.
- γ: Attentional Precision Parameter, higher values imply greater confidence or certainty in the sensory representations.

- m: Meta-Awareness Parameter indicating the ability to explicitly notice or the content of the "dog" thoughtseed is opaque on the Inner Screen.
- $\{\mathbf{x}_K\}_{K\in\mathcal{K}}$: States of the relevant KDs (e.g., "animal," "pet," "emotion") that the "dog" thoughtseed interacts contextually.
- $\{\mathbf{c}_K\}_{K\in\mathcal{K}}$: The set of contents projected onto the Inner Screen from all the KDs that the "Dog" thoughtseed interacts with
- $\theta_{\mathcal{T}}$: The parameters of the generative model for the "dog" thoughtseed
- $\mathbf{u}$: Prior beliefs or expectations about the world, shaped by both evolutionary factors (e.g., innate predispositions towards dogs) and learned experiences.
- σ: Saliency of the sensory input
- $\mathbf{A}$: Set of possible actions afforded by the Umwelt/environment

Level 4: Meta-Cognitive Level - Attentional Control & the Inner Screen

Generative Model (for a higher-order thoughtseed influencing attention):

$$p(\gamma, \{\mathbf{x}_\tau\}_{\tau\in\mathcal{T}}, m, \mathbf{\Pi}, \mathbf{G}|\theta_m, \mathbf{u}) \tag{S.12}$$

where:

- γ: Attentional Precision Parameter, which modulates the precision of the predictions of lower-level thoughtseeds.
- $\{\mathbf{x}_\tau\}_{\tau\in\mathcal{T}}$: The set of internal states of all lower-level thoughtseeds that are competing for attention.
- m:The meta-awareness state of the higher-order thoughtseed, reflecting its awareness of its own influence on the lower-level thoughtseeds and its role in shaping attentional focus.
- $\mathbf{\Pi}$: The set of policies or action plans considered by the higher-order thoughtseed.
- $\mathbf{G}$: The goals or desired outcomes that the higher-order thoughtseed aims to achieve.

- θ_m: The parameters of the generative model for the higher-order thoughtseed.
- $\mathbf{u}$:Prior beliefs or expectations, shaped by evolutionary and learned priors.

This illustrative example demonstrates how the thoughtseed framework can be applied to explain the emergence of complex concepts and their role in shaping embodied cognition. By integrating information from multiple NPDs and KDs, thoughtseeds may provide a unified and meaningful representation of the world, enabling organisms to interact with their environment in an adaptive and goal-directed manner.

10. Acknowledgments

11. References

1. Allen, M., Friston, K.J. (2018). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. Synthese 195, 2459–2482.
2. Analayo. (2012). Satipaṭṭhāna: The Direct Path to Realization. Windhorse Publications.
3. Baars, B. J. (1988). A cognitive theory of consciousness. Cambridge University Press.
4. Badcock, P. B., Friston, K. J., Ramstead, M. J., Ploeger, A., & Hohwy, J. (2019). The hierarchically mechanistic mind: an evolutionary systems theory of the human brain, cognition, and behavior. Cognitive, Affective, & Behavioral Neuroscience, 19, 1319-1351.
5. Baldwin, J. M. (1896). A new factor in evolution. The American Naturalist, 30(354), 441-451.
6. Boscaglia, M., Gastaldi, C., Gerstner, W., & Quian Quiroga, R. (2023). A dynamic attractor network model of memory formation, reinforcement and forgetting. PLOS Computational Biology, 19(12), e1011727.
7. Boyd, R., & Richerson, P. J. (1988). Culture and the evolutionary process. University of Chicago Press.
8. Brefczynski-Lewis, J. A., Lutz, A., Schaefer, H. S., Levinson, D. B., & Davidson, R. J. (2007). Neural correlates of attentional expertise in long-term meditation practitioners. Proceedings of the National Academy of Sciences, 104(27), 11483-11488.
9. Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. Journal of Mathematical Psychology, 81, 55–79.

10. Brunton, S. L., & Kutz, J. N. (2019). Data-driven science and engineering: Machine learning, dynamical systems, and control. Cambridge University Press.

11. Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10(3), 186-198.

12. Buzsáki, G. (2006). Rhythms of the brain. Oxford University Press.

13. Chalmers, D. J. (1995). Facing up to the problem of consciousness. Journal of consciousness studies, 2(3), 200-219.

14. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and brain sciences, 36(3), 181-204.

15. Conant, R. C., & Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. International Journal of Systems Science, 1(2), 89-97.

16. Constant, A., Ramstead, M. J., Veissière, S., Campbell, J. O., & Friston, K. J. (2021). A variational approach to niche construction. Journal of The Royal Society Interface, 18(177), 20200785.

17. Cooper, A. C., Ventura, B., & Northoff, G. (2022). Beyond the veil of duality-topographic reorganization model of meditation. Neuroscience of consciousness, 2022(1), niac013. https://doi.org/10.1093/nc/niac013

18. Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience, 3(3), 201-215.

19. Darwin, C. (1859). On the Origin of Species. John Murray.

20. Darwin, C. (1872). The Expression of the Emotions in Man and Animals. London: John Murray.

21. Decharms, R. C., & Zador, A. (2000). Neural representation and the cortical code. Annual review of neuroscience, 23(1), 613-647.

22. Deco, G., & Kringelbach, M. L. (2016). Metastability and coherence in brain dynamics. Trends in cognitive sciences, 20(4), 251-257.

23. Deco, G., Kringelbach, M. L., Jirsa, V. K., & Ritter, P. (2017). The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. Scientific reports, 7(1), 1-17.

24. Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition, 79(1-2), 1-37.

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

26. Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. Annual Review of Neuroscience, 18(1), 193-222.

27. DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? Neuron, 73(3), 415-434.

28. Douglas, R. J., & Martin, K. A. (1991). A functional microcircuit for cat visual cortex. *The Journal of physiology, 440*(1), 735-769.

29. Dresler, M., Wehrle, R., Spoormaker, V. I., Koch, S. P., Holsboer, F., Steiger, A., ... & Czisch, M. (2012). Neural correlates of dream lucidity obtained from contrasting lucid versus non-lucid REM sleep: a combined EEG/fMRI case study. Sleep, 35(7), 1017-1020.

30. Edelman, G. M. (1987). Neural Darwinism: The theory of neuronal group selection. Basic Books.

31. Eichenbaum, H. (2017). The hippocampus and its role in navigation and memory: A twenty-five-year perspective. Neuron, 95(4), 644-661.

32. Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. Frontiers in human neuroscience, 4, 215.

33. Fields, C., Friston, K., Glazebrook, J. F., & Levin, M. (2016). Eigenforms, interfaces and holographic encoding; Toward an evolutionary account of objects and spacetime. Physics of life reviews, 18, 1-21.

34. Fields, C., Glazebrook, J. F., & Levin, M. (2022). Minimal physicalism as a scale-free substrate for cognition and consciousness. Neuroscience of Consciousness, 2022(1), niac013.

35. Fodor, J. A. (1975). The language of thought. Harvard University Press.

36. Fodor, J. A. (1983). The modularity of mind: An essay on faculty psychology. MIT press.

37. Foglia, L., & Wilson, R. A. (2013). Embodied cognition. Wiley Interdisciplinary Reviews: Cognitive Science, 4(3), 319-325.

38. Forney, G. D. (2001). Codes on graphs: normal realizations. IEEE Transactions on Information Theory, 47(2), 520-548.

39. Fox, K. C., Nijeboer, S., Dixon, M. L., Floman, J. L., Ellamil, M., Rumak, S. P., ... & Christoff, K. (2014). Is meditation associated with altered brain structure? A systematic review and meta-analysis of morphometric neuroimaging in meditation practitioners. Neuroscience & Biobehavioral Reviews, 43, 48-73.

40. Fries, P. (2015). Rhythms for cognition: Communication through coherence. Neuron, 88(1), 220-235.

41. Friston, K. (2008). Hierarchical models in the brain. PLoS computational biology, 4(11), e1000211.

42. Friston, K. (2009). The free-energy principle: A rough guide to the brain? Trends in Cognitive Sciences, 13(7), 293-301.

43. Friston, K. J., & Kiebel, S. J. (2009). Predictive coding under the free-energy principle. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1521), 1211-1221.

44. Friston, K. J. (2010). The free energy principle: A unified approach to biological systems. Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1560), 2257-2268.

45. Friston, K., Kilner, J., & Frith, C. (2010). A free energy formulation of active inference. Frontiers in Computational Neuroscience, 4, 168.

46. Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. Biological cybernetics, 104(1-2), 137-160.

47. Friston, K. J., & Ao, P. (2012). Free energy, value, and attractors. Computational mathematics and modeling, 23(4), 538-551.

48. Friston, K. (2013). Life as we know it. Journal of The Royal Society Interface, 10(86), 20130475.

49. Friston, K., Sengupta, B., & Auletta, G. (2014). Cognitive dynamics: From attractors to active inference. *Proceedings of the IEEE*, *102*(4), 427-445.

50. Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. Cognitive neuroscience, 6(4), 187-214.

51. Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. Neural Computation, 29(1), 1-49.

52. Friston, K. J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. Neuroscience & Biobehavioral Reviews, 90, 486-501.

53. Friston K.J. (2020). A Free Energy Principle for a Particular Physics. arXiv preprint arXiv:1906.10184.

54. Friston, K. J., Fagerholm, E. D., Zarghami, T. S., Parr, T., Hipólito, I., Magrou, L., & Razi, A. (2021). Parcels and particles: Markov blankets in the brain. Network Neuroscience, 5(1), 211-251.

55. Friston, K. (2022). Affordance and active inference. In Affordances in Everyday Life: A Multidisciplinary Collection of Essays (pp. 211-219). Cham: Springer International Publishing.

56. Friston, K., Da Costa, L., Sakthivadivel, D. A. R., Heins, C., Pavliotis, G. A., Ramstead, M., & Parr, T. (2023). Path integrals, particular kinds, and strange things. Physics Letters Review, 1372, 08016.

57. Friston, K., Friedman, D. A., Parr, T., & Ramstead, M. J. (2023). A Variational Synthesis of Evolutionary and Developmental Dynamics. Entropy 2023, 25, 964.

58. Fuster, J. M. (2015). The prefrontal cortex (Fifth edition.). Academic Press.

59. Futuyma, D. J. (2017). Evolutionary biology. Trends in Cognitive Sciences, 21(7), 500-501.

60. George, D., & Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. PLoS computational biology, 5(10), e1000532.

61. Gibson, J. J. (1979). The ecological approach to visual perception. Houghton Mifflin.

62. Ginsburg, S., & Jablonka, E. (2019). The evolution of the sensitive soul: Learning and the origins of consciousness. MIT Press.

63. Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. Vision research, 41(10-11), 1409-1422.

64. Haken, H. (1983). Synergetics: An introduction. Non-equilibrium phase transitions and self-organization in physics, chemistry, and biology (3rd rev. enl. ed.). Springer-Verlag.

65. Hebb, D.O. (1949). The Organization of Behavior. New York: Wiley & Sons.

66. Hennig, W. (1966). Phylogenetic systematics. University of Illinois Press.

67. Herzog, M. H., & Clarke, A. M. (2014). Why vision is sparse. *Trends in Cognitive Sciences*, *18*(5), 243-249.

68. Hipólito, I., Ramstead, M. J. D., Convertino, L., Bhat, A., Friston, K., & Parr, T. (2021). Markov blankets in the brain. Neuroscience and Biobehavioral Reviews, 125, 88–97.

69. Hobson, J. A., & Friston, K. J. (2012). Waking and dreaming consciousness: Neurobiological and functional considerations. Progress in neurobiology, 98, 82-98.

70. Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. D., Gutierrez, C., ... & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys (Csur)*, *54*(4), 1-37.

71. Hohwy, J. (2013). The predictive mind. Oxford University Press.

72. Itti, L., & Koch, C. (2001). Computational modelling of visual attention. Nature Reviews Neuroscience, 2(3), 194-203.

73. Jänig, W. (2006). The integrative action of the autonomic nervous system: Neurobiology of homeostasis. Cambridge University Press.

74. Jablonka, E., & Lamb, M. J. (2002). The changing concept of epigenetics. Annals of the New York Academy of Sciences, 981(1), 82-96.

75. Jablonka, E., & Lamb, M. J. (2005). Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life. MIT Press.

76. Josselyn, S. A., & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. Science, 367(6473), eaaw4325.

77. Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of neuroscience, 17(11), 4302-4311.

78. Kauffman, S. A. (1993). The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press.

79. Kavi, P. C. (2023). Conscious entry into sleep: Yoga nidra and accessing subtler states of consciousness. In Progress in Brain Research (Vol. 280, pp. 43-60). Elsevier.

80. Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2009). A hierarchy of time-scales and the brain. PLoS computational biology, 5(11), e1000509.

81. Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. Cognitive Processing, 8(3), 159-166.

82. Kirchhoff, M., Parr, N., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. Journal of the Royal Society Interface, 15(138), 20170792.

83. Krishnamurti J. (1956). The First and Last Freedom. HarperOne.

84. Krishnamurthi, J., 1974. An Attentive Sleepful State. J. Krishnamurti. YouTube. Available at

85. https://www.youtube.com/watch?v¼wOW1b3DnU6g.

86. Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. Trends in Neurosciences, 27(12), 712-719.

87. Lacal, I., & Ventura, R. (2018). Epigenetic Inheritance: Concepts, Mechanisms and Perspectives. Frontiers in Molecular Neuroscience.

88. Laland, K. N., Odling-Smee, F. J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. Behavioral and brain sciences, 23(1), 131-146.

89. LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. Proceedings of the National Academy of Sciences, 114(10), E2016-E2025.

90. Limanowski, J., & Friston, K. (2018). Active inference, curiosity and insight. Neural Computation, 30(10), 2633-2683.

91. Litwin-Kumar, A., & Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. Nature neuroscience, 15(11), 1498-1505.

92. Loeliger, H. A., Dauwels, J., Hu, J., Korl, S., Ping, L., & Kschischang, F. R. (2007). The factor graph approach to model-based signal processing. Proceedings of the IEEE, 95(6), 1295-1322.

93. Lumsden, C. J., & Wilson, E. O. (1981). Genes, mind, and culture: The coevolutionary process. Harvard University Press.

94. Lutz, A., Jha, A. P., Dunne, J. D., & Saron, C. D. (2015). Investigating the phenomenological and neurocognitive components of mindfulness: A conceptual and empirical review. American Psychologist, 70(7), 582.

95. Lux, V., Gredebäck, G., Non, A. L., & Krüger, M. (2022). Editorial: The next step in developmental embodiment research: Integrating concepts and methods. Frontiers in Systems Neuroscience, 16, 871449.

96. MacLean, E. L., Matthews, L. J., Hare, B. A., Nunn, C. L., Anderson, R. C., Aureli, F., ... & Wobber, V. (2012). How does cognition evolve? Phylogenetic comparative psychology. Animal cognition, 15(2), 223-238.

97. Mameli, M. (2017). The inheritance of features. The British Journal for the Philosophy of Science, 68(2), 365-399.

98. Marler, P., & Tamura, M. (1964). Culturally transmitted patterns of vocal behavior in sparrows. Science, 146(3650), 1483-1486.

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

100. Markov, N. T., Ercsey-Ravasz, M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., ... & Kennedy, H. (2014). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. Cerebral cortex, 24(1), 17-36.

101. Maturana, H. R., & Varela, F. J. (1980). Autopoiesis and cognition: The realization of the living. D. Reidel Publishing Company.

102. Metzinger, T. (2003). Phenomenal transparency and cognitive self-reference. Phenomenology and the Cognitive Sciences, 2(4), 353-393.

103. Miklósi, Á. (2014). Dog behaviour, evolution, and cognition. Oxford University Press.

104. Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. Annual Review of Neuroscience, 24(1), 167-202.

105. Nesse, R. M., & Ellsworth, P. C. (2009). Evolution, emotions, and emotional disorders. American Psychologist, 64(2), 129.

106. Nicolis, G., & Prigogine, I. (1977). Self-organization in nonequilibrium systems. New York, NY: Wiley.

107. Nieder, A. (2017). The neuronal code for number. Nature Reviews Neuroscience, 18(5), 283-296.

108. Noë, A. (2004). Action in perception. MIT press.

109. Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). Niche construction: The neglected process in evolution. Princeton University Press.

110. Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, *14*(4), 481-487.

111. Ororbia, A., & Friston, K. (2023). Mortal computation: A foundation for biomimetic intelligence. arXiv preprint arXiv:2311.09589.

112. Palacios, E. R., Isomura, T., Parr, T., & Friston, K. (2019). The emergence of synchrony in networks of mutually inferring neurons. *Scientific reports*, *9*(1), 6412.

113. Palacios, E.R., Razi, A., Parr, T., Kirchhoff, M., Friston, K., 2020. On Markov blankets and hierarchical self-organisation. J. Theor. Biol. 486, 110089.

114. Panksepp, J., & Biven, L. (2012). The archaeology of mind: Neuroevolutionary origins of human emotions. Norton & Company.

115.	Parr, T., & Friston, K. J. (2017). Working memory, attention, and salience in active inference. Scientific reports, 7(1), 1-14.

116.	Parr, T., Friston, K., & Da Costa, L. (2018). Markov blankets, information geometry and stochastic thermodynamics. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2115), 20170212.

117.	Parr, T., & Friston, K. J. (2018). Working memory as active inference. Biological cybernetics, 112(4), 417-443.

118.	Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. Biological Cybernetics, 113(5), 495-513.

119.	Parr, T., Da Costa, L., Heins, C., Ramstead, M. J. D., & Friston, K. J. (2021). Memory and Markov Blankets. Entropy, 23(9), 1105.

120.	Parr, T., Pezzulo, G., & Friston, K. J. (2022). Active inference: the free energy principle in mind, brain, and behavior. MIT Press.

121.	Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann Publishers Inc.

122.	Pessoa, L. (2014). Understanding brain networks and brain organization. Physics of life reviews, 11(3), 400-435.

123.	Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. Progress in Neurobiology, 134, 17-35.

124.	Pezzulo, G., Parr, T., & Friston, K. (2022). The evolution of brain architectures for predictive coding and active inference. Philosophical Transactions of the Royal Society B: Biological Sciences, 377(1840), 20200531.

125.	Pezzulo, G., D'Amato, L., Mannella, F., Priorelli, M., Van de Maele, T., Stoianov, I. P., & Friston, K. (2024). Neural representation in active inference: Using generative models to interact with—and understand—the lived world. Annals of the New York Academy of Sciences, 1534(1), 45-68.

126.	Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. Neuron, 48(2), 175-187.

127.	Piaget, J., & Cook, M. T. (1954). The development of object concept.

128.	Pinto, Y., Neville, D. A., Otten, M., Corballis, P. M., Lamme, V. A., de Haan, E. H., ... & Fabri, M. (2017). Split brain: divided perception but undivided consciousness. Brain, 140(5), 1231-1237.

129.	Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), Art, Mind, and Religion (pp. 37-48). University of Pittsburgh Press.

130.   Porges, S. W. (2001). The polyvagal theory: phylogenetic substrates of a social nervous system. International Journal of Psychophysiology, 42(2), 123-146.

131.   Quartz, S. R., & Sejnowski, T. J. (1997). The neural basis of cognitive development: A constructivist manifesto. Behavioral and Brain Sciences, 20(4), 537-556.

132.   Rabinovich, M. I., Huerta, R., & Laurent, G. (2008). Transient dynamics for neural processing. Science, 321(5885), 48-50.

133.   Rama, S., 2007. Om the Eternal Witness: Secrets of the Mandukya Upanishad. Himalayan Institute Hospital Trust, Doiwala, U.A.

134.   Ramstead, M. J. D., Kirchhoff, M. D., & Friston, K. J. (2020). A tale of two densities: Active inference is enactive inference. Adaptive Behavior, 28(4), 225–239.

135.   Ramstead, M. J. D., Hesp, C., Tschantz, A., Smith, R., Constant, A., & Friston, K. (2021). Neural and phenotypic representation under the free-energy principle. Neuroscience and biobehavioral reviews, 120, 109–122.

136.   Ramstead, M. J., Friston, K. J., & Hipólito, I. (2023). Is the free-energy principle a formal theory of semantics? From variational density dynamics to neural and phenotypic representations. Entropy, 25(1), 142.

137.   Ramstead, M. J., Albarracin, A. M., Kiefer, A., Klein, B., Fields, C., Friston, K., & Safron, A. (2023). The inner screen model of consciousness: applying the free energy principle directly to the study of conscious experience. Neuroscience of Consciousness, 2023(1), niac013.

138.   Rescorla, M. (2015). The computational theory of mind. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Fall 2015 Edition). Stanford University.

139.   Riesenhuber, M., & Poggio, T. (2000). Hierarchical models of object recognition in cortex. Nature neuroscience, 3(11), 1199-1204.

140.   Roelfsema, P. R. (2023). Solving the binding problem: Assemblies form when neurons enhance their firing rate—they don't need to oscillate or synchronize. *Neuron*, *111*(7), 1003-1019.

141.   Rué-Queralt, J., Stevner, A., Tagliazucchi, E., Laufs, H., Kringelbach, M. L., Deco, G., & Atasoy, S. (2021). Decoding brain states on the intrinsic manifold of human brain dynamics across wakefulness and sleep. Communications Biology, 4(1), 854.

142.   Sandved-Smith, L., Hesp, C., Mattout, J., Friston, K., Lutz, A., & Ramstead, M. J. D. (2021). Towards a computational phenomenology of mental action: modelling meta-awareness and attentional control with deep parametric active inference. Neuroscience of Consciousness, 2021(1), niab018.

143.   Safron, A. (2020). The radically embodied conscious cybernetic Bayesian brain: Towards explaining the emergence of agency and conscious experience. Entropy, 22(7), 752.

144.   Schrödinger, E. (1944). What is life? The physical aspect of the living cell. Cambridge University Press

145.   Seth, A. K. (2015). The cybernetic Bayesian brain: from interoceptive inference to sensorimotor contingencies. In T. K. Metzinger & J. M. Windt (Eds.), Open MIND (pp. 1-24). Frankfurt am Main: MIND Group.

146.   Seth, A. K., & Bayne, T. (2022). Theories of consciousness. Nature Reviews Neuroscience, 23(7), 439-452.

147.   Siegel, S. J., & Wonderly, D. M. (2018). Social learning and cultural evolution. Annual Review of Psychology, 69, 441-466.

148.   Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations?. *Neuron*, *24*(1), 49-65.

149.   Smith, J. M. (1982). Evolution and the Theory of Games. In Did Darwin get it right? Essays on games, sex and evolution (pp. 202-215). Boston, MA: Springer US.

150.   Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. Developmental science, 10(1), 89-96.

151.   Sporns, O. (2011). Networks of the Brain. MIT Press.

152.   Sporns, O., & Betzel, R. F. (2016). Modular brain networks. Annual Review of Neuroscience, 39, 63-84.

153.   Squire, L. R., Wixted, J. T., & Clark, R. E. (2015). Recognition memory and the medial temporal lobe: a new perspective. Nature Reviews Neuroscience, 16(11), 672-683.

154.   Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

155.   Taimni, I. K. (1961). The science of yoga: The yoga sutras of Patanjali. The Theosophical Publishing House.

156.   Tang, Y. Y., Hölzel, B. K., & Posner, M. I. (2015). The neuroscience of mindfulness meditation. Nature Reviews Neuroscience, 16(4), 213-225.

157.   Tart, C. T. (1972). Altered states of consciousness.

158.   Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. Science, 331(6022), 1279-1285. https://doi.org/10.1126/science.1192788

159.   Tinbergen, N. (1951). The study of instinct. Oxford University Press.

160.   Thompson, E. (2007). Mind in life: Biology, phenomenology, and the sciences of mind. Harvard University Press.

161.   Thompson, E. (2015). Dreamless sleep, the embodied mind, and consciousness (Vol. 37, No. T). Frankfurt am Main: MIND Group.

162.   Tognoli, E., & Kelso, J. S. (2014). The metastable brain. Neuron, 81(1), 35-48.

163.   Tripathi, V., & Bharadwaj, P. (2021). Neuroscience of the yogic theory of consciousness. Neuroscience of Consciousness, 2021(2), niab030.

164.   Uller, T., & Laland, K. N. (2019). Evolutionary developmental biology: developmental constraints and evolutionary trajectories. Philosophical Transactions of the Royal Society B: Biological Sciences, 374(1779), 20180182.

165.   Vanchurin, V., Wolf, Y. I., Katsnelson, M. I., & Koonin, E. V. (2022). Toward a theory of evolution as multilevel learning. Proceedings of the National Academy of Sciences, 119(20), e2120042119.

166.   Varela, F. J., Thompson, E., & Rosch, E. (1991). The embodied mind: Cognitive science and human experience. MIT press.

167.   von Uexküll, J. (1934). A stroll through the worlds of animals and men: A picture book of invisible worlds. In Instinctive behavior: The development of a modern concept. (pp. 5-80). Springer, New York, NY.

168.   Wangyal, T.R., 1998. The Tibetan Yogas of Dream and Sleep. Snow Lion, Ithaca, NY.

169.   Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., ... & Boesch, C. (1999). Cultures in chimpanzees. Nature, 399(6737), 682-685.

170.   Yufik, Y. M. (2013). Understanding, consciousness and thermodynamics of cognition. Chaos, Solitons & Fractals, 55, 44–59. doi:10.1016/j.chaos.2013.04.010

171.   Yufik, Y. M., & Friston, K. (2016). Life and Understanding: The Origins of "Understanding" in Self-Organizing Nervous Systems. Frontiers in systems neuroscience, 10, 98.

172.   Yufik Y. M. (2019). The Understanding Capacity and Information Dynamics in the Human Brain. Entropy (Basel, Switzerland), 21(3), 308.

173.   Zamora-López, G., Zhou, C. S., & Kurths, J. (2011). Exploring brain function from anatomical connectivity. Frontiers in Neuroscience, 5, 83.

174.   Zeki, S. (1993). A vision of the brain. Blackwell scientific publications.