Relevance Activation Theory (RAT): A Cue-Indexed Model of Gradient-Based Cognition

Flyxion

June 20, 2025

Abstract

We propose Relevance Activation Theory (RAT), a computational and neurocognitive framework modeling cognition through cue-activated relevance gradients rather than static representations. RAT formalizes cognition across neurocognitive systems, AI agents, and abstract cognitive structures, describing behavior, memory, creativity, and trauma as flows along or reshaping of scalar relevance fields indexed by environmental or internal cues. We demonstrate how RAT unifies hippocampal navigation, AI policy optimization, and abstract cognitive dynamics, offering testable predictions for neuroscience and scalable architectures for artificial intelligence.

1 Introduction

This section introduces Relevance Activation Theory (RAT), a novel framework that redefines cognition as dynamic navigation through cue-triggered relevance fields, challenging traditional representational models. We outline the limitations of static cognitive maps, introduce the rat metaphor for cue-based navigation, and describe the paper's structure across neurocognitive, AI, and abstract cognitive domains.

Traditional cognitive models rely on static, map-like representations [1], assuming the brain stores explicit knowledge structures. However, such models struggle to capture the dynamic, context-sensitive nature of cognition [2]. We propose Relevance Activation Theory (RAT), inspired by the metaphor of a rat navigating affordance spaces via cues, not rigid maps. RAT posits that cognition emerges from gradient flows over relevance fields triggered by cues, eschewing symbolic representations for continuous, adaptive dynamics.

Representational theories overemphasize fixed structures, neglecting how context reshapes cognitive priorities [3]. RAT reframes cognition as navigation through a high-dimensional affordance space, where cues activate scalar relevance fields guiding behavior, memory, and creativity. This paper formalizes RAT across three domains: neurocognitive systems, AI agents, and abstract cognitive geometries, with implications for neuroscience, AI, and clinical psychology.

Section 2 defines RAT's neurocognitive formulation, Section 3 outlines its AI implementation, Section 4 explores its abstract geometry, and Section 5 discusses implications and future work.

2 Neurocognitive Formulation

This section formalizes RAT's neurocognitive model, drawing on hippocampal place cell dynamics and Hebbian learning. We define relevance fields as scalar functions over perceptual or motor spaces, describe cue-driven activation, and model behavior as gradient-based navigation, linking to synaptic reinforcement and place field approximations.

2.1 Relevance Fields

RAT models cognition as navigation over a scalar relevance field $R : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X}$ is a perceptual or motor space. Relevance quantifies the behavioral utility of a state $x \in \mathcal{X}$ given a cue.

2.2 Cue-Driven Activation

Cues trigger localized activation via Gaussian bumps:

$$A_c(x) = \phi(\|x - x_c\|) \cdot w_c, \tag{1}$$

where ϕ is a radial basis function, x_c is the cue's center, and w_c is its weight [4].

2.3 Action as Gradient Flow

Behavior follows gradient ascent on the relevance field:

$$\frac{dx}{dt} = f(\nabla R(x), \theta), \tag{2}$$

where f is a dynamics function parameterized by θ, modeling motor or cognitive transitions [5].

2.4 Synaptic Reinforcement via Trails

Relevance trails reinforce connections via Hebbian learning:

$$\Delta w_{ij} = \eta A_i A_j, \tag{3}$$

where η is the learning rate and A_i, A_j are activations of connected nodes [6].

2.5 Place Field Approximation

Relevance fields approximate hippocampal place fields:

$$R(x) = \sum_i \alpha_i e^{-\|x-\mu_i\|^2/2\sigma^2}, \tag{4}$$

where α_i scales the i-th bump centered at μ_i with width σ [7].

3 AI Implementation

This section translates RAT into artificial intelligence, defining how agents can navigate environments or semantic spaces using cue-driven relevance gradients. We introduce embeddings for cues and states, neural networks for relevance prediction, softmax policies for action selection, and dynamic affordance graphs for learning.

3.1 Embedding Cue-Relevance

Cues and states are embedded in a shared space:

$$c, x \in \mathbb{R}^d, \quad \kappa(c, x) = \langle c, x \rangle, \tag{5}$$

where κ measures cue-state alignment [8].

3.2 Learned Relevance Network

A neural network $R_\theta(x) \in \mathbb{R}$ predicts relevance, trained via supervised or cue-driven objectives.

3.3 Policy Based on Gradient Maximization

Behavior follows a softmax policy:

$$\pi(a|x) \propto \exp(\beta R(x + a)), \tag{6}$$

where β controls exploration [9].

3.4 Affordance Graphs

Edges update via:

$$w_{ij}^{t+1} = (1 - \alpha) w_{ij}^t + \alpha \cdot \text{affordance}_{ij}, \tag{7}$$

modeling dynamic affordance learning.

4 Cognitive Theory and Abstract Geometry

This section abstracts RAT into a general cognitive theory using topological and geometric tools. We model relevance as context-sensitive fiber bundles, affordances as sheaves, attention as vector fields, and processes like trauma and creativity as dynamic field manipulations.

4.1 Relevance Fiber Bundles

Context-sensitive relevance is modeled as a fiber bundle $\pi : \mathcal{R} \to \mathcal{C}$, where $\mathcal{C}$ is the cue space [10].

4.2 Cue Sheaf

Affordances form a sheaf $\mathcal{F} : \text{CueCat}^{\text{op}} \to \text{Set}$, encoding local-to-global cue consistency [11].

4.3 Attention Vector Fields

Attention is a vector field $\vec{A} : \mathcal{C} \to T\mathcal{C}$, directing cognitive focus along relevance gradients.

4.4 Rewriting Trauma Fields

Trauma fields are reshaped via coactivation:

$$R_{c^*}^{\text{new}} = \gamma R_{\tilde{c}} + (1-\gamma) R_{c^*}^{\text{old}}, \tag{8}$$

where γ blends new and old fields [12].

4.5 Creative Geodesics

Creativity follows low-energy paths:

$$\gamma(t) \subset \mathcal{C} \text{ minimizing } \int |\nabla R(\gamma(t))| dt, \tag{9}$$

optimizing semantic exploration [13].

5 Discussion

This section synthesizes RAT's contributions, connecting it to neuroscience, AI, and clinical applications. We compare RAT to predictive coding, highlight its implications for therapy and creativity, and propose future research directions, including real-time implementations and empirical validation.

RAT bridges hippocampal place cell dynamics [7] with predictive coding [14] and AI policy learning [15]. Unlike predictive representations, RAT emphasizes dynamic, cue-driven navigation, offering a unified account of behavior, memory, and creativity. In AI, RAT enables agents to "feel" affordance spaces, while in therapy, it suggests trauma can be reshaped by cue reweighting. Future work includes real-time RAT agents, fMRI validation, and cross-species modeling.

References

[1] Anderson, J. R. (1990). *The Adaptive Character of Thought*. Erlbaum.

[2] Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–660.

[3] Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204.

[4] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[5] Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.

[6] Hebb, D. O. (1949). *The Organization of Behavior*. Wiley.

[7] O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. *Brain Research*, 34(1), 171–175.

[8] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv:1301.3781*.

[9] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.

[10] Baez, J., & Stay, M. (2010). Physics, topology, logic and computation: a Rosetta Stone. In *New Structures for Physics* (pp. 95–172). Springer.

[11] Mac Lane, S., & Moerdijk, I. (1992). *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Springer.

[12] van der Kolk, B. A. (2014). *The Body Keeps the Score: Brain, Mind, and Body in the Healing of Trauma*. Viking.

[13] Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. MIT Press.

[14] Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex. *Nature Neuroscience*, 2(1), 79–87.

[15] Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.