

Article

Not peer-reviewed version

A Dual-Model Architecture for Cognition

[Jaba Tkemaladze](#) *

Posted Date: 12 January 2026

doi: 10.20944/preprints202601.0842.v1

Keywords: cognitive architecture; variational inference; model conflict; active inference; quantum cognition; phase transition



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Dual-Model Architecture for Cognition

Jaba Tkemaladze

Longevity Clinic, Inc, Georgia; djabbat@gmail.com

Abstract

This paper introduces Ze, a novel theoretical framework for cognitive architecture based on the concurrent operation of two distinct generative models of the environment: a causal (forward) model M_A and a counterfactual (inverse) model M_B . The core dynamics of Ze arise from the minimization of two separate variational free energies, F_A and F_B , and the management of the conflict between them, $\Delta F = |F_A - F_B|$. This conflict regulates a phase transition between an interference regime, where model outputs are constructively fused, and a localization regime, which resolves the conflict through a discrete projection. We formally establish a deep structural isomorphism with quantum measurement, particularly the double-slit experiment, without invoking quantum physics in the substrate. Ze is proposed as a complete, falsifiable theory that reinterprets cognitive “collapse” as an optimization-driven transition, generates novel experimental predictions, and integrates perception, action, and representational learning into a unified architecture. This preprint provides the full mathematical elaboration of the framework.

Keywords: cognitive architecture; variational inference; model conflict; active inference; quantum cognition; phase transition

1. Introduction

The quest for a formal, computationally grounded theory of cognition and consciousness remains a central challenge in neuroscience and artificial intelligence. Prevailing paradigms, such as predictive processing and the free energy principle, posit that the brain functions as a hierarchical generative model minimizing surprise or variational free energy (Friston, 2010). While powerful, these frameworks often describe a single, unified model of perception. The Ze formalism proposes a fundamental extension: intelligent systems require the maintenance of two distinct, asymmetric generative models of the same environment.

This duality is motivated by the need to explain not just passive perception, but also counterfactual reasoning, narrative understanding, and the sudden reorganization of belief. We introduce a causal (forward) model, M_A , which predicts sensory data based on temporal dynamics, and a counterfactual (inverse) model, M_B , which infers explanations, goals, and latent causes. Their independent operation and subsequent interaction form the core of the Ze architecture.

The theory is built rigorously on established variational mechanics, making no assumptions about quantum brain processes. However, it reveals a profound formal analogy with quantum measurement. This analogy is not merely metaphorical but indicates that the mathematical principles governing the resolution of uncertainty may be universal across different physical and computational substrates. Ze offers a new perspective on classic cognitive phenomena—perceptual bistability, insight problem-solving, and the sleep-wake cycle—and yields specific, testable predictions that distinguish it from standard models.

2. Mathematical Formalism

2.1. Core Variables and Basic Structure

Consider an agent receiving a stream of observations $o_{1:T}$. Ze posits two separate generative models:

M_A : Causal (forward) model

M_B : Counterfactual (inverse) model.

Each model maintains its own latent states, s^A_t and s^B_t , and its own approximate posterior distribution over them:

$$q_A(s^A_t) \approx P(s^A_t | o_{1:t})$$

$$q_B(s^B_t) \approx P(s^B_t | o_{1:t}).$$

These posteriors are updated via variational inference, but they are not required to be symmetric in time, structure, or granularity (Botvinick & Toussaint, 2012).

2.2. Two Variational Free Energies

Each model minimizes its own variational free energy functional, an upper bound on surprise (Friston, 2010):

$$F_A(o, q_A) = \mathbb{E}_{q_A(s^A)} [\ln q_A(s^A) - \ln p(o, s^A | M_A)]$$

$$F_B(o, q_B) = \mathbb{E}_{q_B(s^B)} [\ln q_B(s^B) - \ln p(o, s^B | M_B)].$$

The generative models $p(o, s^X | M_X)$ can have radically different factorizations. M_A typically follows a forward-temporal dynamic, while M_B may be structured to infer causes backward-in-time or from high-level abstractions (Solway & Botvinick, 2012). This structural asymmetry is foundational.

2.3. Model Conflict: The Core Quantity

The central dynamic variable in Ze is the model conflict or interpretation divergence:

$$\Delta F = F_A(o, q_A) - F_B(o, q_B).$$

This is not a sensory signal but a metacognitive, structural variable that regulates the system's operational regime (Fleming & Daw, 2017).

A small ΔF indicates consensus, permitting constructive interference.

A large ΔF signals dissonance, triggering diagnostic localization.

2.4. Interference as Posterior Compatibility

Interference is formalized as the constructive blending of the models' posteriors. We measure their compatibility using the Jensen-Shannon divergence (JSD), a symmetric, bounded metric:

$$J = D_{JS}(q_A(s) || q_B(s)).$$

Here, s represents a common representational subspace.

When $J \approx 0$, posteriors are compatible, and interference—a form of optimal fusion—is possible (Ernst & Banks, 2002).

When $J \gg 0$, interference is suppressed to prevent representational corruption.

2.5. Localization as a Phase Transition

Localization is triggered when the conflict exceeds a threshold θ , initiating a phase transition:

If $\Delta F > \theta \Rightarrow$ Localization is triggered.

The process involves a projection onto a resolved state $\hat{s}$ that best reconciles the models:

$$q(s) \rightarrow q(s | \hat{s}), \text{ where } \hat{s} = \operatorname{argmin}_s [\alpha F_A(s) + (1-\alpha) F_B(s)].$$

The weighting parameter α reflects the relative confidence of each model. This projection is the cognitive analogue of a quantum measurement collapse, but it arises from optimization, not an external postulate (Bruza et al., 2015).

2.6. Active Actions: Ze is Not a Passive Observer

Ze extends to active behavior. Each model can propose actions through its own policy:

$$a_t \sim \pi_A(a_t | s^A_t, \Omega_t) \text{ or } a_t \sim \pi_B(a_t | s^B_t, \Omega_t).$$

The system selects the policy expected to minimize the total future free energy (Friston et al., 2017):

$$\pi^{\wedge} = \operatorname{argmin}_{\pi \in \{\pi_A, \pi_B\}} \mathbb{E}_{q(\cdot|\pi)} [F_A(o_{\tau}, q_A) + F_B(o_{\tau}, q_B)].$$
 Action is thus a tool for active learning and conflict resolution.

2.7. Which-Path Information and Dimensionality Growth

Learning occurs through “which-path” information: the discovery of a hidden contextual variable e . This expands the agent’s state space:

$$s \rightarrow (s, e).$$

This expansion increases model misspecification, causing both ΔF and $\mathcal{I}$ to rise sharply ($\Delta F \uparrow, \mathcal{I} \uparrow$), making localization inevitable and forcing structural model revision (Gershman & Niv, 2010).

2.8. The Quantum Eraser as a Cognitive Operator

Conversely, Ze includes a simplification operator, the quantum eraser $\mathcal{E}$. It acts by decorrelating a contextual variable:

$$\mathcal{E}: p(e | s) \rightarrow \text{const}.$$

It does not erase past data but reduces the effective environmental dimensionality, lowers conflict ($\Delta F \downarrow$), and can restore the interference regime ($\Delta F < \theta$) when detailed distinctions become irrelevant or costly (Tononi & Cirelli, 2014).

2.9. Sleep and Wakefulness as Parameter Regimes

The sleep-wake cycle is modeled as a shift in a path fixation parameter λ that weights a specificity cost in the free energy:

$$F_A^{\wedge \lambda} = F_A + \lambda \cdot \text{path}(s).$$

Wakefulness ($\lambda \gg 1$): High λ enforces a precise, committed model for real-time action, suppressing $\mathcal{E}$.
Sleep ($\lambda \rightarrow 0$): Low λ removes the specificity penalty, enabling broad exploration, the operation of $\mathcal{E}$, and memory consolidation (Lewis & Durrant, 2011).

3. Formal Correspondence with Quantum Mechanics

The mathematical structure of Ze exhibits a strict isomorphism with the quantum double-slit experiment.

Table 1. Formal Correspondence between the Double-Slit Experiment and the Ze Architecture.

Physical Concept (Double-Slit)	Ze Architecture	Cognitive	Formal Mechanism	Expression /
Wave (Superposition)	Function	Model Compatibility	Low ΔF , high compatibility	posterior ($\mathcal{I} \approx 0$) allows blended states.
Interference Pattern	Constructive Interference Regime		Fused percepts and predictions from q_A and q_B when $\mathcal{I} \approx 0$.	

Which-Path Information	State Space Expansion		Acquisition of e expands $s \rightarrow (s, e)$, increasing conflict $\Delta F \uparrow$.
Collapse (Measurement)	Localization Transition	Phase	$\Delta F > \theta$ triggers projection to a resolved state $\hat{s}$.
Quantum Eraser	Erasure Operator $\mathcal{E}$		$\mathcal{E}$ decorrelates e from s , reducing ΔF , restoring $\mathcal{I} \approx 0$.
Decoherence (Environment)	Self-Induced Decoherence via Action		Commitment to a policy π_A or π_B provides continuous sensory feedback, stabilizing a “classical” percept.

This correspondence demonstrates that the principles of superposition, interference, measurement, and erasure are not unique to quantum physics but describe general dynamics of systems that manage uncertainty across competing internal models (Busemeyer & Bruza, 2012). In Ze, “collapse” is not a postulate but an emergent optimization-driven phase transition.

4. Experimental Predictions and Falsifiability

Ze is a strict theory, not a metaphor, because it is built on standard variational calculus, makes no quantum physical claims, proposes a novel architectural hypothesis, and yields falsifiable predictions.

4.1. Key Testable Predictions

Table 2. Key Experimental Predictions of the Ze Theory.

Level	Prediction	Expected Empirical Signature
Neurophysiological	Distinct neural populations for q_A and q_B ; conflict ΔF encoded in neuromodulatory systems.	Dissociable fMRI/EEG networks; ΔF correlates with pupil dilation (LC-NE activity) (Aston-Jones & Cohen, 2005).

Behavioral	Active policy alternation accelerates conflict resolution.	In volatile tasks, subjects using two distinct exploratory actions identify change-points faster than passive observers.
Cognitive	The erasure operator $\mathcal{E}$ is more active during sleep/low demand.	Sleep enhances generalization over specific details; post-sleep tasks show reduced interference from irrelevant contextual cues (Tomov et al., 2021).
Computational	A dual-model Ze agent outperforms a single-model agent in environments with hidden contexts.	Higher survival rate/score in RL environments requiring discovery of latent “which-path” variables e.

4.2. Minimal Critical Prediction

A critical, distinguishing prediction is that active alternation between policies π_A and π_B leads to faster localization ($\Delta F > \theta \rightarrow \hat{s}$) than passive observation. This contrasts with passive decoherence models, where resolution speed depends only on information rate. An experimental paradigm using a volatile decision task can directly test this (Findling et al., 2023).

5. Discussion and Conclusions

The Ze formalism integrates perception, action, and learning into a unified, mathematically precise architecture. Its core innovation is the mandatory duality of generative models, whose conflict and reconciliation dynamics govern cognition.

5.1. Implications for Cognitive Science

Ze provides a principled explanation for phenomena like perceptual bistability (spontaneous localization), insight (sudden conflict resolution), and the function of sleep (periodic erasure and integration). It frames cognition as a constant negotiation between a causal “what is” model and a counterfactual “what could be” model.

5.2. A Foundational Shift

The theory reinterprets the “hard problem” of cognitive collapse—how a definite percept emerges from ambiguous data—as a soft, optimization-based phase transition. This demystifies the process and aligns it with broader physical principles of self-organization (Tschacher & Haken, 2007).

5.3. Future Directions

Future work must focus on formal computational implementations of the Ze agent and designing the critical experiments outlined in Table 2. Furthermore, the isomorphism with quantum

mechanics invites cross-disciplinary dialogue, suggesting that quantum information theory may offer powerful tools for modeling high-level cognitive processes without requiring a quantum brain.

In conclusion, Ze is proposed as a complete, falsifiable framework that advances our formal understanding of intelligent systems by unifying variational inference, active learning, and structural model revision under the governance of dual-model dynamics.

Figure 1. Schematic of the Ze Architecture. A block diagram showing the two generative models, M_A and M_B , receiving observations o_t and maintaining posteriors q_A and q_B . The diagram illustrates the calculation of F_A , F_B , ΔF , and the feedback of the conflict signal to regulate the interference/localization switch and policy selection (π_A , π_B). The erasure operator $\mathcal{E}$ and the path fixation parameter λ are also shown as modulatory inputs.

Figure 2. Dynamical Regimes and the Phase Transition. A bifurcation diagram plotting a system stability measure against the model conflict ΔF . It shows a stable “Interference” attractor for $\Delta F < \theta$, which loses stability at the threshold θ . For $\Delta F > \theta$, two stable “Localized” attractors appear, corresponding to resolutions favoring M_A or M_B . Arrows indicate the system’s trajectory during a localization event.

References

- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403–450.
- Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, 16(10), 485–488.
- Bruza, P. D., Wang, Z., & Busemeyer, J. R. (2015). Quantum cognition: a new theoretical approach to psychology. *Trends in Cognitive Sciences*, 19(7), 383–393.
- Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge University Press.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433.
- Findling, C., Chopin, N., & Koechlin, E. (2023). Imprecise neural computations as a source of adaptive behavioural variability. *Nature Communications*, 14, 3686.
- Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, 124(1), 91–114.
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, 29(1), 1–49.
- Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology*, 20(2), 251–256.
- Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences*, 15(8), 343–351.
- Solway, A., & Botvinick, M. M. (2012). Goal-directed decision making as probabilistic inference: A computational framework and potential neural correlates. *Psychological Review*, 119(1), 120–154.
- Tomov, M. S., Yagati, S., Kumar, A., Yang, W., & Gershman, S. J. (2021). Discovery of hierarchical representations for efficient planning. *PLoS Computational Biology*, 16(4), e1007594.
- Tononi, G., & Cirelli, C. (2014). Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration. *Neuron*, 81(1), 12–34.
- Tschacher, W., & Haken, H. (2007). Intentionality in non-equilibrium systems? The functional aspects of self-organized pattern formation. *New Ideas in Psychology*, 25(1), 1–15.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.