Formal Toy Model Specification

Consciousness as a Maintained Selective Field-Order

March 24, 2026

Abstract

This document gives a self-contained formal specification for a toy model of consciousness understood as a maintained selective field-order rather than as a mere container of contents or a neural implementation claim. The model treats consciousness at time t as a structured field of disclosure-tokens with mode, salience, articulation, valence, persistence, and orientation-pull, linked by a dynamic relational matrix and organized by a subject-relative uptake operator. The model is designed to formalize foregrounding, background retention, residue, closure, fragmentation, propagated framing, and correction in a compact way. It is intentionally minimal and non-neural. Its purpose is to render the object tractable enough for conceptual analysis, simulation, and design exploration.

1. Purpose and Status

This is a toy model. It is not a completed theory of consciousness, not a neural theory, and not a claim to have solved the hard problem. Its function is narrower: to provide a compact formal rendering of consciousness as a maintained selective order in which things become available in varying degrees of salience, articulation, and uptake.

The model aims to make the following structurally visible:

- consciousness as a maintained field rather than a flat list of contents;
- non-identity of presence, salience, articulation, and uptake;
- mode-differentiation among perceptual, affective, cognitive, and reflexive disclosure;
- the existence of foreground, background, and residue;
- closure, fragmentation, and correction as field-level dynamics.

2. Model Identity

The model represents consciousness at time t as a dynamic field

$$\mathcal{C}_t = (C_t, W_t, U_t),$$

where:

- C_t is the set of active disclosure-tokens;
- W_t is the relational structure among those tokens;
- U_t is a subject-relative uptake operator over the field.

The core idea is that conscious life is not exhausted by explicit contents alone. What matters is the organization of what is present, salient, weakly carried, strongly articulated, backgrounded, pressing, fixated, or corrected.

3. Time and Domain

Time is discrete:

$$t \in \{0, 1, 2, \dots\}.$$

At each time t, there is a finite set of disclosure-tokens

$$C_t = \{d_1^t, d_2^t, \dots, d_{n_t}^t\}.$$

Each token represents a local disclosure-event or disclosure-candidate. Tokens may arise from perceptual input, bodily/affective pressure, memory activation, imagination, thought, or reflexive awareness.

4. Disclosure-Token State Space

Each token d_i^t carries a state vector

$$d_i^t = (m_i,\ s_i^t,\ a_i^t,\ v_i^t,\ p_i^t,\ q_i^t,\ e_i^t),$$

with the following components:

Symbol	Meaning
m_i	mode type, with $m_i \in \{\text{perc}, \text{aff}, \text{cog}, \text{refl}\}$
$s_i^t \in [0,1]$	salience: how strongly the token stands out in the field
$a_i^t \in [0,1]$	articulation: how structured, explicit, or thinkable the token is
$v_i^t \in [-1,1]$	valence / affective pressure
$p_i^t \in [0,1]$	persistence: tendency to remain active or reactivate
$q_i^t \in [0,1]$	orientation-pull: tendency to shape action, attention, or interpretation
$e_i^t \in [0,1]$	exogenous input strength for the current step

These dimensions are not identical. A token may be salient but poorly articulated, articulated but low in salience, affectively loaded but weakly taken up, or highly persistent while rarely foregrounded.

5. Relational Structure

The field includes a weighted relational matrix

$$W_t = (W_{ij}^t) \in [0,1]^{n_t \times n_t}.$$

Here W_{ij}^t measures the strength of relation between d_i^t and d_j^t. It may represent, depending on the case:

- perceptual grouping,
- conceptual association,
- memory linkage,
- affective resonance,
- framing dependence,

- self-referential relation.

The model therefore treats consciousness not as an unordered set, but as a structured field with local clustering, support, propagation, and possible fragmentation.

6. Subject-Position and Uptake

The subject-position is not modeled as one more token in the field. Instead it is represented as an uptake operator

$$U_t : C_t \to [0, 1],$$

so that

$$u_i^t = U_t(d_i^t)$$

measures the degree to which token d_i^t is actively taken up at time t.

This distinguishes:

- being present in the field,
- being salient in the field,
- being actively taken up.

These should not be collapsed.

6.1. Uptake Score

Define an intermediate uptake score

$$r_i^t = \omega_s s_i^t + \omega_q q_i^t + \omega_v |v_i^t| + \omega_w \sum_{j=1}^{n_t} W_{ij}^t s_j^t + \omega_g g_i^t,$$

where $g_i^t \in [0, 1]$ is task- or goal-relevance and the ω-coefficients are nonnegative parameters.

Then define normalized uptake by a softmax:

$$u_i^t = \frac{\exp(\tau r_i^t)}{\sum_{k=1}^{n_t} \exp(\tau r_k^t)},$$

where $\tau > 0$ is an inverse-temperature parameter controlling selectivity.

7. Token Generation

At each time step, the field is refreshed by survival of prior tokens and arrival of new ones:

$$C_{t+1}^{\text{raw}} = C_t^{\text{survive}} \cup P_t \cup B_t \cup M_t \cup I_t \cup R_t^{\text{meta}},$$

where:

- P_t: perceptual inputs,
- B_t: bodily or affective inputs,
- M_t: memory activations,
- I_t: imaginative or cognitive constructions,
- R_t^{meta}: reflexive higher-order tokens.

Thus consciousness is modeled as a renewed field, not a static repository.

8. Update Dynamics

8.1. Salience Update

Salience evolves according to prior salience, affective loading, relational support, exogenous input, and uptake:

$$s_i^{t+1} = \sigma\Big(\alpha_s s_i^t + \beta_v |v_i^t| + \beta_w \sum_{j=1}^{n_t} W_{ij}^t s_j^t + \beta_e e_i^t + \beta_u u_i^t - \delta_s\Big).$$

Here:

- α_s controls carry-over of prior salience,
- β_v controls affective amplification,
- β_w controls field-support from linked tokens,
- β_e controls exogenous input gain,
- β_u controls reinforcement through uptake,
- δ_s is a salience decay term,
- σ is a squashing function into $[0, 1]$.

A simple choice is the logistic map

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$

8.2. Articulation Update

Articulation increases when tokens are worked on cognitively, taken up, or reflexively stabilized:

$$a_i^{t+1} = \text{clip}\left(a_i^t + \eta_c \chi_i^t + \eta_r \rho_i^t + \eta_u u_i^t - \delta_a(1 - u_i^t)\right),$$

where:

- $\chi_i^t \in [0, 1]$ is cognitive elaboration applied to token i,
- $\rho_i^t \in [0, 1]$ is reflexive uptake of token i,
- η_c, η_r, η_u are gains,
- δ_a is articulative decay,
- $\text{clip}(x)$ clips values to $[0, 1]$.

8.3. Persistence Update

Persistence measures survival tendency across steps:

$$p_i^{t+1} = \sigma\Big(\alpha_p p_i^t + \beta_p u_i^t + \gamma_p \sum_{j=1}^{n_t} W_{ij}^t u_j^t - \delta_p\Big),$$

where $\alpha_p, \beta_p, \gamma_p, \delta_p \geq 0$.

High persistence is useful for modeling recurring thoughts, mood-laden themes, unresolved concerns, and background pressures that do not immediately disappear.

8.4. Relational Update

Relation strengths update through reinforcement, co-uptake, and cognitive restructuring:

$$W_{ij}^{t+1} = \text{clip}\left((1 - \lambda_W) W_{ij}^t + \eta_W u_i^t u_j^t + \eta_\Phi \Phi_{ij}^t\right),$$

where:

- λ_W is relation decay,
- η_W is co-uptake reinforcement,
- Φ_{ij}^t is a thought-driven or reflexive reframing term.

9. Mode-Specific Roles

Modes do not require separate ontological compartments, but they do influence updates differently.

9.1. Perceptual Tokens

Perceptual tokens are typically characterized by:

$$m_i = \text{perc},$$

with relatively high initial e_i^t, moderate or high s_i^t, and initially lower a_i^t.

9.2. Affective Tokens

Affective tokens tend to act as field reweightings rather than only articulated contents:

$$m_i = \text{aff}.$$

They may globally or locally amplify salience. One simple local form is

$$s_i^{t+1} \leftarrow s_i^{t+1}(1 + \lambda_A A_i^t),$$

where $A_i^t \in [0, 1]$ is affective amplification.

9.3. Cognitive Tokens

Cognitive tokens increase articulation and can restructure the field:

$$m_i = \text{cog}.$$

They tend to contribute positively to χ_i^t and Φ_{ij}^t.

9.4. Reflexive Tokens

Reflexive tokens represent higher-order uptake:

$$m_i = \text{refl}.$$

They may be generated by an operator

$$d_k^{t+1} = \mathfrak{R}(d_i^t),$$

producing tokens such as:

- "I am anxious,"
- "this thought keeps returning,"
- "something is wrong but unclear."

10. Foreground, Background, and Residue

10.1. Foreground and Background

Define a foreground threshold θ_F and background threshold θ_B with $0 \leq \theta_B < \theta_F \leq 1$.

Then:

$$F_t = \{d_i^t \in C_t \mid s_i^t > \theta_F\},$$

$$B_t = \{d_i^t \in C_t \mid \theta_B < s_i^t \leq \theta_F\}.$$

Foreground tokens dominate immediate conscious availability. Background tokens remain active without dominating the field.

10.2. Residue

Define residue as salient or affectively loaded material that is not well articulated and not adequately taken up:

$$\mathcal{R}_t = \sum_{i=1}^{n_t} s_i^t \, |v_i^t| \, (1 - a_i^t) \, (1 - u_i^t).$$

This quantity is central. It formalizes the case where something is experientially pressing or important without yet being clearly thinkable or well integrated.

11. Closure, Fragmentation, and Conflict

11.1. Closure

Let normalized uptake mass be

$$\pi_i^t = \frac{u_i^t}{\sum_{k=1}^{n_t} u_k^t}.$$

Define closure by concentration:

$$K_t = \sum_{i=1}^{n_t} (\pi_i^t)^2.$$

Interpretation:

- low K_t: uptake is distributed;
- high K_t: a narrow subset dominates the field.

High closure corresponds to fixation, premature simplification, or interpretive lock-in.

11.2. Fragmentation

Define fragmentation as lack of integration:

$$\mathcal{F}_t = 1 - \frac{\sum_{i \neq j} W_{ij}^t}{n_t(n_t - 1)}.$$

This is a simple density-based measure. Higher values indicate lower field integration.

11.3. Conflict

Let $\Gamma_{ij} \geq 0$ measure incompatibility between tokens i and j. Then define conflict:

$$\mathcal{D}_t = \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \Gamma_{ij} u_i^t u_j^t.$$

Conflict captures simultaneous active tension among incompatible framings or disclosures.

12. Correction Dynamics

Define correction pressure as a weighted combination of residue, conflict, and fragmentation:

$$\mathcal{P}_t = \lambda_R \mathcal{R}_t + \lambda_D \mathcal{D}_t + \lambda_F \mathcal{F}_t,$$

with nonnegative coefficients $\lambda_R, \lambda_D, \lambda_F$.

When correction pressure exceeds a threshold Θ,

$$\mathcal{P}_t > \Theta,$$

the field undergoes reorganization:

$$\mathcal{C}_{t+1} = \Gamma(\mathcal{C}_t),$$

where Γ is a correction operator.

12.1. Correction Operator

The correction operator may include some or all of the following:

- downweighting dominant closure,
- increasing articulation of previously weakly carried tokens,
- weakening incompatible links,
- strengthening neglected but recurrent tokens,

- generating reflexive higher-order tokens,
- redistributing uptake mass.

A minimal algebraic form is:

$$s_i^{t+1} \leftarrow (1 - \kappa_K K_t) s_i^{t+1} + \kappa_R \zeta_i^t,$$
$$a_i^{t+1} \leftarrow a_i^{t+1} + \kappa_A \zeta_i^t,$$

where ζ_i^t is a correction-priority score, for example

$$\zeta_i^t = s_i^t |v_i^t| (1 - a_i^t).$$

This makes correction sensitive to salient, affectively loaded, weakly articulated material.

13. Token Survival and Decay

A token survives into the next step when its persistence exceeds a survival threshold θ_p:

$$C_t^{\text{survive}} = \{d_i^t \in C_t \mid p_i^t > \theta_p\}.$$

Tokens below threshold are removed from the active field, though the model can be extended to include latent stores or memory reservoirs.

14. Canonical Regimes

The model is useful if distinct parameter regimes map to recognizable lived structures.

14.1. Rumination

High persistence, high closure, weak correction:

$$\alpha_p \uparrow, \qquad K_t \uparrow, \qquad \Theta \uparrow .$$

14.2. Anxiety

High affective amplification and high residue:

$$\beta_v \uparrow, \qquad \lambda_A \uparrow, \qquad \mathcal{R}_t \uparrow .$$

14.3. Insight

A previously weakly articulated but persistent token is elevated and integrated:

$$a_i \uparrow, \qquad W_{ij} \text{ restructured}, \qquad \mathcal{R}_t \downarrow .$$

14.4. Mood

A broad field-parameter shifts salience weighting across many tokens:

$$\beta_v = \beta_v(t), \qquad \omega_q = \omega_q(t).$$

14.5. Pre-Articulate Unease

Salient and pressure-laden, but weakly articulated:

$$s_i \uparrow, \qquad |v_i| \uparrow, \qquad a_i \downarrow .$$

15. Observables

The following field-level quantities are recommended as observables:

- **Foreground size:** $|F_t|$
- **Background size:** $|B_t|$
- **Residue:** $\mathcal{R}_t$
- **Closure:** K_t
- **Fragmentation:** $\mathcal{F}_t$
- **Conflict:** $\mathcal{D}_t$
- **Correction pressure:** $\mathcal{P}_t$
- **Mean articulation:** $\frac{1}{n_t} \sum_i a_i^t$
- **Mean persistence:** $\frac{1}{n_t} \sum_i p_i^t$

These make the model simulable and diagnostically useful.

16. Scope and Non-Claims

This formalization does *not* claim:

- that consciousness is exhausted by this model;
- that the toy variables are neurally literal;
- that the subject-position has been metaphysically explained;
- that phenomenal character is reducible to the state variables above;
- that all conscious phenomena are captured by salience, articulation, residue, closure, and correction alone.

The model claims only that these dimensions provide a tractable formal approximation of consciousness as a maintained selective field-order.

17. Minimal Summary

The model can be summarized in one line as:

$$\mathcal{C}_t = (C_t, W_t, U_t),$$

with dynamic updates for salience, articulation, persistence, relation, foregrounding, residue, closure, fragmentation, and correction.

In words:

> Consciousness is modeled as a dynamically maintained field of disclosure-tokens with variable salience, articulation, valence, persistence, and relational linkage, organized by a subject-relative uptake operator and exhibiting foregrounding, background retention, residue, closure, fragmentation, conflict, and correction.

18. Minimal Update Cycle

For simulation or implementation purposes, the model's update cycle is:

1. generate or import new tokens;
2. update salience;
3. update articulation;
4. update persistence;
5. update relational structure;
6. compute uptake;
7. partition foreground and background;

8. compute residue, closure, fragmentation, and conflict;
9. apply correction if pressure exceeds threshold;
10. remove non-surviving tokens.

This cycle is the toy dynamic core.