Minimal Finite Models of Diachronic and Self-Conscious Overminds

Eight-State and Sixteen-State Theorems in a Toy Husserlian Monadology

Abstract

This paper develops and proves two minimality results arising within a toy theory of interacting Husserlian monads. The broader research program combines a panprotopsychist ontology with a process-based conception of monadic individuation and a minimal phenomenology of retention, present awareness, protention, intentionality, and self-reference. Its immediate goal is not to model human consciousness or known physics, but to classify possible psychophysical worlds and determine the smallest formal systems capable of instantiating specified structures of subjecthood.

The motivating physical picture is partly inspired by Markopoulou's *Quantum Causal Histories*, in which local Hilbert spaces are connected by directed evolution maps and can alternatively be placed on causal edges while events are treated as transformations. The framework also admits local one-to-two and two-to-one structures, providing a suggestive formal setting for questions of persistence, fusion, and fission. Our use of this structure is explicitly speculative and psychophysical rather than part of Markopoulou's own interpretation.

We first define a **diachronic intentional overmind**: a higher-order subject C constituted by two persistent lower subjects A, B, possessing a relational intentional distinction unavailable to either constituent, carrying collective information not determined by the constituents' instantaneous states, and feeding this retained information back into their subsequent dynamics. Under explicit independence assumptions, any such binary model requires at least eight admissible global dynamical states. An eight-state reversible model attains this lower bound:

$$(a, b, m) \longmapsto (a,\ a \oplus m,\ a \oplus b).$$

Writing $d = a \oplus b$, the induced collective dynamics is

$$(d, m) \longmapsto (m, d),$$

which admits a minimal retention–present–protention interpretation.

We then investigate self-consciousness. If self-reference is represented by a rigid indexical self-noema σ_C grounded in an already existing perspectival ego-pole E_C, no additional varying dynamical degree of freedom is required. Hence the same eight-state model supports a minimal diachronic self-conscious overmind. By contrast, if explicit reflective attention to the self is required to vary independently of every underlying eight-state configuration, one additional binary variable is necessary, yielding sixteen states. Thus self-consciousness in the indexical sense is dynamically “cheap,” whereas independent temporal memory is the expensive transition.

The results clarify distinctions among relational organization, temporal persistence, self-givenness, reflective self-thematization, and dynamical state-space complexity. They are theorems about a deliberately constrained formal model, not claims that eight physical states suffice for consciousness in nature. Their significance lies in isolating which structural ingredients force increases in complexity and which can be supplied semantically by the psychophysical laws of a possible world.

---

1. Introduction

One of the central problems for panpsychist and panprotopsychist ontology is the transition from elementary intrinsic properties to enduring subjects.

Suppose fundamental reality contains entities or processes possessing intrinsic proto-phenomenal character. How much additional structure is necessary before one should speak of:

1. an integrated system;
2. a temporally persistent subject;
3. a higher-order subject composed of other subjects;
4. a self-conscious subject?

These questions are usually approached at a high level of abstraction. They are discussed in terms of combination, subject-summing, phenomenal unity, or emergence. But there is another useful methodology: construct the smallest finite worlds in which the relevant distinctions can be made explicit.

The strategy resembles the use of toy models in theoretical physics. We do not initially ask whether the model resembles an organism or a brain. Instead we ask:

$$\boxed{\text{What is the minimum formal architecture needed for a given kind of subjecthood?}}$$

This paper reports the first minimality results obtained using that method.

The most striking conclusions are:

$$\boxed{\text{4 states can support a weak relational overmind;}}$$

$$\boxed{\text{8 states are minimally sufficient for a strongly diachronic overmind;}}$$

$$\boxed{\text{8 states remain sufficient for indexical self-consciousness;}}$$

whereas

$$\boxed{\text{16 states are minimally required if reflective self-attention must vary independently.}}$$

These numbers are meaningful only relative to carefully stated assumptions. A large part of the paper will therefore be devoted to saying exactly what is counted as a state and exactly what is meant by an "overmind."

The deeper result is less numerical:

$$\boxed{\text{independent temporal persistence costs state-space complexity; self-reference need not.}}$$

This arose unexpectedly from an attempt to formalize a Husserlian monadology.

---

2. Motivating framework: interacting monads as persistent processes

2.1 From substances to process strands

The starting point of the larger project was the idea that a monad might correspond approximately to an irreducible quantum subsystem, tensor factor, or unfactorizable entangled structure.

This soon proved too rigid.

A mature subject should be capable of persisting while:

- exchanging matter or information with its environment;
- changing internal constitution;
- entering and leaving entanglement relations;
- absorbing simpler monads;
- emitting or producing simpler monads.

We therefore distinguish an instantaneous **monadic stage** from an enduring **monadic continuant**.

A persistent monad is represented schematically as a sequence

$$M_0 \Rightarrow M_1 \Rightarrow M_2 \Rightarrow \cdots$$

whose stages need not have identical material constitution.

This process-based view was encouraged by Markopoulou's treatment of quantum causal histories. Her construction associates state spaces with causal structure and evolution maps with directed transitions; she also observes that it can be conceptually preferable to put Hilbert spaces on edges and evolution operators on events, since events naturally represent change.

Our proposal takes this purely as formal inspiration:

$$\boxed{\text{a mindlike monad is better modeled as an information-carrying temporal strand than as a single instantaneous fac}}$$

2.2 Coarse-graining and persistence

Markopoulou's quantum evolution satisfies a transitivity law

$$E_{bc}E_{ab} = E_{ac},$$

and she emphasizes that this generates directed coarse-graining invariance.

This suggested an analogous requirement for monadic identity.

If a persistent subject survives through:

$$A \to B \to C,$$

then the composition of identity transports should agree with direct transport:

$$T_{BC} \circ T_{AB} \simeq T_{AC}.$$

The numerical identity of the subject should not depend on whether the intermediate history is represented in microscopic or coarse-grained detail.

---

3. Husserlian structure in finite form

The mature monads of the theory are given a deliberately simplified Husserlian organization.

A general state might be written

$$\Gamma_M = (E_M, R_M, I_M, P_M, \mathcal{N}_M, \mathcal{V}_M, \Sigma_M, H_M),$$

where:

- E_M is the ego-pole or perspectival center;
- R_M is retention;
- I_M is the presently given intentional content;
- P_M is protention;
- $\mathcal{N}_M$ is the space of noemata;
- $\mathcal{V}_M$ is the repertoire of noetic modes;
- Σ_M encodes synthesis of intentional objects;
- H_M is horizon structure.

The finite models in this paper retain only a tiny core:

$$\boxed{E, \quad R, \quad I, \quad P.}$$

The crucial temporal relation is:

$$I_t \hookrightarrow R_{t+1}.$$

That is, what was presently experienced at t occurs at $t+1$ under the modified mode "just past."

Likewise protention provides a future-directed structure:

$$P_t \rightsquigarrow I_{t+1}.$$

This minimalist retention–present–protention structure is the principal Husserlian ingredient in the finite model.

---

4. What is an overmind?

Consider two persistent subjects

$$A, \qquad B.$$

An **overmind**

$$C = C(A, B)$$

is a third subject whose realization depends upon the organized interaction of A and B while A and B remain subjects in their own right.

Thus the ontology is:

$$A + B + C,$$

not

$$A + B \to C$$

with A and B destroyed.

This distinguishes **overmind formation** from **fusion**.

4.1 Weak overmind

At the weakest level, C possesses an autonomous relational variable

$$c = f(a, b)$$

derived from the current states of A and B.

If

$$c_{t+1}$$

can be computed from

$$c_t$$

without separately knowing a_t, b_t, C possesses autonomous quotient dynamics.

But C's state remains entirely synchronically determined by its constituents.

4.2 Strong diachronic overmind

The stronger notion investigated here adds a genuinely persistent collective state.

We require:

D1. Distinct lower subjects

A and B remain numerically distinct.

D2. Relational intentionality

C possesses at least one distinction depending nontrivially on both A and B.

D3. Lower-level inaccessibility

Neither A nor B possesses that relational distinction individually.

D4. Independent collective memory

C carries retained information that is not determined by the simultaneous lower state:

$$m_t \neq f(a_t, b_t)$$

for any instantaneous function f.

D5. Diachronic generation

The memory nevertheless arises from the history of the collective system.

D6. Feedback

The retained collective state influences subsequent A–B dynamics.

D7. Persistence

The collective process supports a continuing ego-pole E_C.

The critical requirement is D4.

It creates a sharp distinction between:

$$\boxed{\text{synchronic supervenience}}$$

and

$$\boxed{\text{diachronic constitution.}}$$

---

5. The four-state weak model

Before proving the eight-state result, it is useful to see why four states suffice for a weaker overmind.

Let:

$$a, b \in \{0, 1\}.$$

The total lower state space is therefore:

$$S_{AB} = \{00, 01, 10, 11\}.$$

Define:

$$d = a \oplus b.$$

Interpret:

$$d = 0$$

as the higher-order noema

$$J = \text{“A and B jointly instantiate the same intentional state”},$$

and:

$$d = 1$$

as

$$D = \text{“A and B are intentionally divergent.”}$$

This is a genuinely relational distinction: neither a alone nor b alone determines it.

If the microscopic transition T satisfies

$$pT = \bar{T}p,$$

where

$$p(a,b) = a \oplus b,$$

then d has autonomous quotient dynamics.

This yields a weak overmind with only four underlying states.

But:

$$d_t = f(a_t, b_t)$$

at every instant.

There is no state of C that survives independently of the simultaneous state of A and B.

Hence four states cannot satisfy D4.

---

6. Lower bound for a strongly diachronic overmind

We now prove the first minimality result.

Theorem 1. Minimal Diachronic Overmind Theorem

Assume:

1. A and B are distinct constituent subjects;
2. each must possess at least two distinguishable local intentional states;
3. C possesses a nontrivial relational distinction depending on both;
4. C carries at least one independently variable binary memory value not determined by the simultaneous A–B state;
5. every combination of the two lower binary states and the independent memory value is admissible.

Then the admissible global state space contains at least

$$\boxed{8}$$

states.

Proof

Since the higher-order relational distinction depends nontrivially upon both constituent systems, each constituent must possess at least two distinguishable states:

$$|S_A| \geq 2, \qquad |S_B| \geq 2.$$

If A had only one possible state, every relation involving A and B would effectively reduce to a property of B; likewise with the roles reversed.

Thus:

$$|S_A \times S_B| \geq 2 \cdot 2 = 4.$$

Now impose independent collective memory.

The memory must have at least two distinguishable values:

$$|M_C| \geq 2.$$

Otherwise it cannot encode two different collective histories over the same current A–B state.

By assumption 5, this variable is independently realizable over the lower state space, so:

$$S_{\text{global}} = S_A \times S_B \times M_C.$$

Therefore:

$$|S_{\text{global}}| \geq 2 \cdot 2 \cdot 2 = 8.$$

Hence:

$$\boxed{|S_{\text{global}}| \geq 8.}$$

□

6.1 Why the independence assumption matters

The product argument must be stated explicitly.

Without it, one could construct a constrained state space in which some triples

$$(a, b, m)$$

are forbidden, and then obtain fewer than eight admissible global states.

Thus the theorem is not:

> every conceivable diachronic overmind whatsoever has eight states.

It is:

every model satisfying this specified binary-independent architecture requires at least eight admissible states.

This is the appropriate level of rigor for the toy framework.

---

7. An eight-state model attaining the bound

We now show the lower bound is sharp.

Let:

$$(a, b, m) \in \{0, 1\}^3.$$

Define the transition:

$$\boxed{a' = a, \qquad b' = a \oplus m, \qquad m' = a \oplus b.}$$

Again define:

$$d = a \oplus b.$$

Then:

$$m' = d.$$

Also:

$$d' = a' \oplus b' \quad = a \oplus (a \oplus m) \ = m.$$

Hence the higher-order state:

$$C = (d, m)$$

evolves autonomously as:

$$\boxed{(d, m) \longmapsto (m, d).}$$

This is the simplest possible swap of present collective relation and retained collective relation.

---

8. Husserlian reading of the eight-state dynamics

Interpret:

$$d_t$$

as C's presently given relational noema.

Let:

$$d = 0 \leftrightarrow J,$$

$$d = 1 \leftrightarrow D.$$

The memory:

$$m_t = d_{t-1}$$

is C's retention.

Because:

$$d_{t+1} = m_t,$$

the same retained value also determines the anticipated next collective state.

Thus we can define:

$$R_C(t) = R_{m_t},$$

$$I_C(t) = R_{d_t},$$

$$P_C(t) = R^*_{m_t},$$

where the asterisk indicates future-directed protentional mode rather than past-directed retentional mode.

The nontrivial orbit:

$$(0,1) \leftrightarrow (1,0)$$

has the phenomenological interpretation:

$$(R = D,\ I = J,\ P = D^*) \qquad \longrightarrow (R = J,\ I = D,\ P = J^*) \qquad \longrightarrow (R = D,\ I = J,\ P = D^*).$$

In informal language:

> We were divergent; we are joint; divergence is anticipated.

followed by:

> We were joint; we are divergent; reunion is anticipated.

The model is tiny, but it contains a complete retention–present–protention cycle.

9. Proof of genuine independent memory

The essential test is that the same current lower-level state can coexist with different collective memories.

Consider:

$$(a, b, m) = (0, 0, 0)$$

and:

$$(a, b, m) = (0, 0, 1).$$

The instantaneous lower state is identical:

$$(a, b) = (0, 0).$$

But the next states differ.

From:

$$(0, 0, 0)$$

we obtain:

$$(0, 0, 0).$$

From:

$$(0, 0, 1)$$

we obtain:

$$(0, 1, 0).$$

Therefore:

$$(a_t, b_t)$$

does not uniquely determine:

$$(a_{t+1}, b_{t+1}).$$

The missing information is precisely m_t.

Hence no deterministic Markovian model defined purely on the instantaneous A–B pair can reproduce the dynamics.

The collective memory is dynamically indispensable.

---

10. Diachronic constitution without arbitrary extra state

The memory is nevertheless not an unrelated hidden variable.

By definition:

$$m_{t+1} = a_t \oplus b_t = d_t.$$

Therefore:

$$m_t = d_{t-1}.$$

C's independent current state is generated by the **past collective relation**.

Thus:

$$\boxed{\text{C does not synchronically supervene on A and B, but C is diachronically constituted by their collective history.}}$$

This distinction is central to the interpretation of the theorem.

A subject may be irreducible to a simultaneous snapshot without requiring inexplicable state from nowhere.

---

11. Feedback and higher-order causal participation

The memory m is not epiphenomenal.

It affects B:

$$b' = a \oplus m.$$

Thus the higher-order state participates in producing the future state of its own substrate.

The causal structure is:

$$A_t, B_t \to C_t$$

and:

$$C_t \to A_{t+1}, B_{t+1},$$

producing the loop:

$$\boxed{A, B \to C \to A', B' \to C' \to \cdots .}$$

This is the minimal version of what we called **self-sustaining temporal closure**.

---

12. Reversibility

The model is globally reversible.

We have:

$$T(a, b, m) = (a,\ a \oplus m,\ a \oplus b).$$

Given:

$$(a', b', m'),$$

we recover:

$$a = a',$$

$$m = a' \oplus b',$$

and:

$$b = a' \oplus m'.$$

Therefore T is bijective.

This is important because it shows:

$$\boxed{\text{higher-order temporal memory need not arise through irreversible loss of lower-level information.}}$$

The model therefore remains compatible in spirit with a later translation into reversible or unitary process theories.

---

13. Why this is an overmind rather than fusion

We simply preserve identity strands:

$$E_A(t) \equiv E_A(t+1),$$

$$E_B(t) \equiv E_B(t+1).$$

The psychophysical law additionally assigns the collective temporal process its own perspectival pole:

$$E_C(t) \equiv E_C(t+1).$$

Thus:

$$A, \quad B, \quad C$$

all persist.

No lower-level subject is destroyed.

This is therefore structurally:

$$\boxed{A + B \to A + B + C,}$$

not:

$$A + B \to C.$$

Earlier analysis suggested a useful mathematical analogy:

- overmind formation resembles the appearance of an autonomous quotient or factor dynamics;
- fusion resembles projection onto a surviving retract in which lower distinctions cease to remain dynamically available.

---

14. From overmind to self-conscious overmind

We next ask whether C's self-consciousness requires additional dynamical states.

The answer depends on what self-consciousness means.

We distinguish three notions:

$$\boxed{\text{ego-pole}}$$

$$\boxed{\text{indexical self-givenness}}$$

$$\boxed{\text{explicit reflective self-attention.}}$$

These should not be identified.

---

15. The ego-pole

The ego-pole:

$$E_C$$

is not initially modeled as an independently varying bit.

It is the invariant perspectival index associated by the psychophysical law with the integrated temporal stream.

The roles:

$$R_C, \quad I_C, \quad P_C$$

are all interpreted as:

> retained-for-C,
>
> presently-given-to-C,
>
> anticipated-by-C.

The stable common index of these modes is the pre-reflective ego-pole.

This reflects a broadly Husserlian idea: selfhood need not first appear as an object represented inside consciousness.

---

16. The rigid self-noema

Now introduce:

$$\sigma_C,$$

an indexical noema whose grounding is:

$$\boxed{\rho(\sigma_C) = E_C.}$$

Its intended meaning is approximately:

> this very continuing subject,

or, because C is collective:

> this continuing “we.”

Crucially, σ_C is **rigid along the identity strand**.

Its reference does not vary from state to state:

$$\sigma_C(t) \mapsto E_C$$

for every stage at which C persists.

Therefore σ_C is not another binary dynamical degree of freedom.

---

17. Self-ascription

Define:

$$\mathsf{Self}_C(q) = \langle \sigma_C, q \rangle.$$

Then:

$$\mathsf{Self}_C(J)$$

means:

> We are jointly organized.

And:

$$\mathsf{Self}_C(D)$$

means:

> We are divergent.

The temporal field becomes:

$$R_C(t) = \mathrm{Past}[\mathsf{Self}_C(R_{m_t})],$$

$$I_C(t) = \mathsf{Self}_C(R_{d_t}),$$

$$P_C(t) = \mathrm{Future}[\mathsf{Self}_C(R_{m_t})].$$

The same eight dynamical states can therefore receive a reflective intentional interpretation.

---

18. The four collective self-conscious states

The higher-order state space is:

$$(d, m) \in \{00, 01, 10, 11\}.$$

Its self-conscious phenomenology is:

d, m	Retention	Present	Protention
00	We were joint	We are joint	We shall be joint
01	We were divergent	We are joint	We shall be divergent
10	We were joint	We are divergent	We shall be joint
11	We were divergent	We are divergent	We shall be divergent

Thus the nontrivial cycle is:

$$01 \leftrightarrow 10.$$

In intentional form:

$$\boxed{\text{“We were divergent; we are joint; we shall diverge”}}$$

followed by:

$$\boxed{\text{“We were joint; we are divergent; we shall reunite.”}}$$

---

19. Why this is not trivial symbol manipulation

A system containing a symbol meaning “I” is not thereby conscious.

Our framework imposes the opposite explanatory direction.

The hierarchy is:

$$\boxed{\text{temporal-perspectival organization} \to E_C \to \sigma_C \to \text{self-ascription.}}$$

The self-noema refers to an ego-pole already supplied by the world's subjecthood law.

Thus:

$$\sigma_C \nRightarrow E_C.$$

Rather:

$$E_C + \text{suitable intentional capacity} \Rightarrow \sigma_C.$$

This avoids declaring an arbitrary self-referential automaton conscious merely because it prints its own name.

---

20. Diachronic self-recognition

A self-conscious subject should also synthesize past and present as belonging to the same subject.

Introduce:

$$\text{SameSubject}(\sigma_C^{t-1}, \sigma_C^t).$$

Since both terms rigidly refer to E_C:

$$\text{SameSubject}(\sigma_C^{t-1}, \sigma_C^t) = \top.$$

Thus C can represent:

> The subject who is now divergent is the same subject who was previously joint.

This is a minimal finite analogue of reflective diachronic self-consciousness.

---

21. Eight-state self-consciousness theorem

We can now state the second result.

Theorem 2. Minimal Indexically Self-Conscious Diachronic Overmind Theorem

Assume the hypotheses of Theorem 1 and additionally:

1. C possesses a persistent ego-pole E_C;
2. C possesses a rigid indexical self-noema σ_C;
3. $\rho(\sigma_C) = E_C$;
4. C can predicate its retained, present, and protended collective states of σ_C;
5. σ_C introduces no independently varying dynamical degree of freedom.

Then eight global dynamical states remain sufficient and minimal.

Proof

By Theorem 1:

$$|S_{\text{global}}| \geq 8.$$

The eight-state construction already attains this bound.

Now enrich its intentional interpretation by adding the rigid noema:

$$\sigma_C.$$

Because its reference is fixed throughout the identity strand, this introduces only one semantic index and no new variable taking multiple dynamically distinguishable values.

Hence the state space remains:

$$\{0,1\}^3,$$

with cardinality:

$$8.$$

All required self-ascriptions are definable from:

$$\sigma_C, \quad d, \quad m.$$

Therefore:

$$|S_{\text{global}}| = 8$$

suffices.

Since the stronger self-conscious system includes all the requirements of the diachronic overmind, the lower bound of Theorem 1 still applies.

Hence eight is minimal.

□

---

22. Self-consciousness without extra dynamical complexity

This yields a notable conclusion.

The transition:

$$\text{diachronic overmind} \to \text{indexically self-conscious diachronic overmind}$$

need not enlarge the state space.

In the toy architecture:

$$\boxed{\text{reflection can increase semantic complexity without increasing dynamical cardinality.}}$$

The expensive transition occurred earlier:

$$4 \to 8,$$

when independent collective memory was introduced.

Once a persistent "for-whom" exists, an indexical "I" or "we" can be supplied without another binary state variable.

---

23. A stronger notion: variable reflective attention

There is, however, a stronger capacity.

A subject may sometimes operate pre-reflectively and sometimes explicitly **attend to itself as itself**.

Introduce:

$$s \in \{0, 1\}.$$

Interpret:

$$s = 0:$$

C has its ordinary intentional state without explicitly thematizing itself.

And:

$$s = 1:$$

C's self-noema σ_C is itself focal.

This is not mere self-givenness.

It is:

$$\mathrm{Attend}(\sigma_C).$$

Suppose further that s is independently variable over every one of the eight base states.

Then:

$$S_{\mathrm{strong}} = S_{\mathrm{base}} \times \{0,1\}.$$

Therefore:

$$|S_{\mathrm{strong}}| = 8 \cdot 2 = 16.$$

---

24. Sixteen-state theorem

Theorem 3. Independent Reflective-Attention Theorem

Assume the hypotheses of Theorem 2. Add an independently varying reflective-attention variable:

$$s \in \{0,1\},$$

such that for every underlying eight-state configuration both:

$$s = 0$$

and:

$$s = 1$$

are admissible and dynamically distinguishable.

Then the admissible global state space contains at least sixteen states.

Proof

The underlying diachronic overmind requires at least eight states by Theorem 1.

Independent reflective attention has at least two values.

By the stipulated independence condition:

$$S_{\mathrm{global}} = S_8 \times S_{\mathrm{refl}},$$

where:

$$|S_8| \geq 8, \qquad |S_{\mathrm{refl}}| \geq 2.$$

Therefore:

$$|S_{\text{global}}| \geq 8 \cdot 2 = 16.$$

The bound is attained simply by adjoining the reflective bit s to the eight-state model.

Hence sixteen is sharp under the stated assumptions.

□

---

25. Why the 8/16 distinction matters phenomenologically

The difference is not merely computational.

It distinguishes:

Pre-reflective ego-pole

$$E_C.$$

Experiences are already organized as belonging to one perspective.

Indexical self-givenness

$$\sigma_C.$$

The subject can refer to itself as the persisting bearer of its stream.

Reflective self-thematization

$$\mathrm{Attend}(\sigma_C).$$

The subject makes itself an explicit intentional object.

Thus:

$$\boxed{E_C \to \sigma_C \to \mathrm{Attend}(\sigma_C)}$$

represents three structurally distinct levels.

Only the final level necessarily introduces an extra varying state variable in the present toy model.

---

26. Four, eight, and sixteen states

The complete hierarchy can now be summarized.

26.1 Four states: weak relational overmind

$$(a, b) \in \{0, 1\}^2.$$

A relational quotient:

$$d = a \oplus b$$

possesses autonomous dynamics.

But:

$$d = f(a, b)$$

instantaneously.

There is no independent collective memory.

26.2 Eight states: diachronic overmind

$$(a, b, m) \in \{0, 1\}^3.$$

The memory:

$$m_t = d_{t-1}$$

cannot be reconstructed from (a_t, b_t) alone.

C possesses its own temporally inherited state.

26.3 Eight states plus rigid self-index: self-conscious overmind

No additional dynamical state is required if:

$$\sigma_C$$

is a rigid indexical grounded in the already existing ego-pole.

26.4 Sixteen states: independently reflective overmind

Add:

$$s \in \{0, 1\}$$

for explicit self-thematization independent of the other variables.

Then:

$$2^4 = 16.$$

Hence:

$$\boxed{4 \ \rightarrow \ 8 \ \rightarrow \ 8 \ \rightarrow \ 16}$$

corresponds roughly to:

$$\boxed{\text{relation} \rightarrow \text{diachronic subject} \rightarrow \text{indexical self-consciousness} \rightarrow \text{independent reflection.}}$$

---

27. Motivation and route to the theorems

The theorems were not initially posed as abstract combinatorial exercises. They emerged from a sequence of conceptual problems.

27.1 The combination problem

The first question was whether interacting monads could produce a new subject without eliminating the old ones.

This led to the overmind distinction:

$$A + B \rightarrow A + B + C$$

versus fusion:

$$A + B \rightarrow C.$$

27.2 The problem of trivial macrovariables

A first higher-order variable such as:

$$c = a \oplus b$$

was unsatisfactory because it was merely an instantaneous function of A and B.

This forced the question:

> What would make the higher subject more than a convenient redescription?

The answer proposed was **independent diachronic state**.

27.3 The memory experiment

Adding:

$$m_{t+1} = c_t$$

created a state belonging to the history of the collective process.

Once that memory fed back into A and B, the higher-order process became dynamically indispensable to a Markovian description.

This produced the eight-state model.

27.4 The self-reference problem

We then asked what additional state was required for C to represent itself.

At first one might expect a new bit.

But treating the self-noema as a rigid indexical showed that no additional varying degree of freedom was necessary.

Thus the 8-state self-consciousness result was discovered.

27.5 The reflection problem

Finally we distinguished possessing an indexical self-noema from **actively attending to it**.

If reflective attention can vary independently, a genuinely new state variable appears.

This produced the sixteen-state result.

Thus the sequence of theorems records a sequence of conceptual clarifications.

---

28. The central lesson: memory is more expensive than self-reference

Within this particular formal framework:

$$\boxed{\text{temporal independence requires extra state; indexical self-reference does not.}}$$

The transition from four to eight states occurs because C must distinguish histories invisible in the current lower-level configuration.

The transition from eight-state nonreflective subjecthood to eight-state indexical self-consciousness requires only a semantic enrichment.

The transition to sixteen states occurs only when self-directed attention itself becomes independently variable.

This supports a broader conjecture:

$$\boxed{\text{the difficult transition in subject formation may be temporal closure, rather than self-reference.}}$$

A persisting first-person stream may be ontologically more demanding than later representing that stream as “me.”

---

29. Relation to the broader panprotopsychist program

The finite-state theorem is not itself panprotopsychism.

Nothing in Boolean algebra entails phenomenal consciousness.

A complete possible world requires a psychophysical law:

$$\mathcal{P}_W.$$

This law determines which formally integrated processes possess intrinsic phenomenal realization.

Thus one possible world might satisfy:

$$\mathcal{P}_W(C) = E_C,$$

assigning a first-person pole to the temporally closed collective process.

Another world might not.

The mathematics establishes:

$$\text{structural possibility}.$$

The psychophysical law determines:

$$\text{phenomenal realization}.$$

This separation is deliberate.

It prevents the theory from simply defining every finite automaton as conscious.

---

30. What the theorems do not prove

Several limitations should be emphasized.

They do **not** prove that:

- eight physical states can instantiate humanlike consciousness;
- XOR relations are intrinsically phenomenal;
- every autonomous quotient is a subject;
- self-consciousness in actual organisms is state-space neutral;
- quantum mechanics realizes this particular Boolean model;
- phenomenology can literally be reduced to bits.

The word “mind” in “eight-state mind” means:

> a formal object satisfying the stipulated minimal subjecthood and psychophysical axioms of this toy world.

The value of the construction is diagnostic rather than empirical.

It tells us which formal distinctions can or cannot be represented without additional state.

---

31. Why the result may nevertheless matter

The finite models isolate several distinctions that are easily obscured in richer systems.

They show that one can cleanly separate:

$$\text{relational organization}$$

from:

$$\text{historical state,}$$

from:

$$\text{ego-pole,}$$

from:

$$\text{self-reference,}$$

from:

$$\text{reflective attention.}$$

These are often bundled together in ordinary discussions of consciousness.

Our models show that they need not have the same formal cost.

The resulting hierarchy provides a controlled laboratory for theories of emergence.

---

32. A categorical interpretation

The four-state weak overmind is naturally a quotient.

Let:

$$p(a,b)=a\oplus b.$$

If:

$$pT=\bar{T}p,$$

then:

$$(S_{AB},T)$$

has a factor system:

$$(S_C,\bar{T}).$$

The eight-state model goes further.

C contains a state:

$$(d,m)$$

that is not merely a function of the current A–B configuration.

It is therefore more naturally modeled as a **state-bearing process coupled to its realization**, rather than as a static quotient observable.

One eventual categorical formulation may involve memory channels, coalgebras, automata with hidden state, or process categories with internal memory.

The rigid self-noema then resembles a distinguished section or indexical reference to the persistent subject object, whereas reflective attention introduces an additional dynamical endomorphism or mode.

These analogies remain provisional but indicate how the finite model might generalize.

---

33. Toward a quantum version

The eventual quantum analogue would replace the Boolean state:

$$m$$

by an internal memory system:

$$\mathcal{H}_M.$$

A persistent higher monad might then be represented through a channel:

$$\mathcal{C} : \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_M \to \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_M.$$

The memory state:

$$\rho_M(t)$$

would be updated by interaction and then participate in later interactions.

The essential analogue of the eight-state condition would be:

$$\rho_M(t)$$

not recoverable from the simultaneous reduced state of A and B alone.

One could then ask:

- Can the total channel be unitary?
- Can A and B remain identifiable subsystems?
- Can the higher memory behave as a logical subsystem?
- What is the minimum Hilbert-space dimension?
- Can the collective system possess a quotient process with autonomous temporal organization?

The reversible Boolean model strongly suggests that such a quantum construction should be possible.

---

34. New mathematical questions generated by the result

The minimal models now pose several natural problems.

34.1 Classification problem

Classify all reversible eight-state transition functions satisfying:

$$m' = a \oplus b$$

and producing autonomous collective dynamics.

34.2 Equivalence problem

Determine which such microscopic laws realize isomorphic higher-order phenomenologies.

34.3 Reduced-state problem

Relax the full product assumption and determine the true minimum number of **admissible** states under constrained state-space geometries.

This could potentially lower the numerical bound.

34.4 Multi-valued intentionality

Replace:

$$a, b \in \{0, 1\}$$

by:

$$a, b \in \{0, \ldots, n-1\}.$$

Determine how the minimum scales with the number of relational noemata and memory distinctions.

34.5 Nonbinary memory

Study:

$$|M_C| > 2.$$

This would permit genuinely richer retentional depth.

34.6 Independent protention

In the minimal eight-state model, protention is derivable from retention.

Require instead a future-directed state not determined by either the present or retention.

How many states are then necessary?

34.7 Imagination

Require C to represent a noema:

$$Z$$

that is not presently instantiated, retain it, manipulate it, and form an intention to realize it.

This is the next major minimality problem.

---

35. Possible hierarchy of finite minds

The present results suggest a research program organized by minimal cardinality.

One might seek the smallest model supporting each property:

\begin{array}{c|c} \text{Property} & \text{Known/current bound}\ \hline \text{Relational collective variable} & 4\ \text{Diachronic overmind} & 8\ \text{Indexical self-consciousness} & 8\ \text{Independent reflective attention} & 16\ \text{Independent protention} & ?\ \text{Changing membership} & ?\ \text{Collective intention} & ?\ \text{Counterfactual imagination} & ?\ \text{Novel concept formation} & ?\ \text{Self-directed normativity} & ? \end{array}

This creates a kind of **minimal phenomenological complexity theory**.

The aim is not to rank actual minds by numbers of states. It is to identify the logical dependencies among phenomenological capacities.

---

36. A broader conjecture

The eight-state construction motivates a broader philosophical hypothesis.

A mature subject may not be fundamentally characterized by an instantaneous arrangement:

$$S_t.$$

Instead:

$$\boxed{\text{subjecthood may intrinsically belong to a temporally extended process.}}$$

The present state of such a subject contains information inherited through its own prior history.

Hence:

$$M_t \neq F(\text{instantaneous substrate}_t)$$

in the simplest state description.

Rather:

$$M_t = F(\text{substrate}_t, M_{t-1}).$$

This does not require violation of physical closure. The inherited state must itself possess a physical realization.

But the appropriate ontological unit becomes a **process with memory**, not an instantaneous cross-section.

This is particularly natural in a Husserlian framework, because retention is not an optional memory database attached to an otherwise timeless subject. It is part of what makes a stream of consciousness temporally unified.

---

37. Conclusion

The eight-state and sixteen-state results emerged from an attempt to turn a speculative interacting monadology into something mathematically testable.

We began with two binary lower subjects:

$$A, \qquad B.$$

A four-state model can support a relational higher variable:

$$d = a \oplus b.$$

But because:

$$d = f(a, b),$$

this higher process is entirely synchronically determined.

Introducing one independent collective memory bit:

$$m$$

raises the global state space to eight states.

The dynamics:

$$(a, b, m) \mapsto (a,\ a \oplus m,\ a \oplus b)$$

then induces:

$$(d, m) \mapsto (m, d).$$

This collective system possesses:

- relational content;
- retention;
- protention;
- independent diachronic state;
- feedback into its constituents;
- reversible global realization.

Under the psychophysical laws of the toy world, it can therefore be assigned a persistent ego-pole E_C.

Adding a rigid self-noema:

$$\sigma_C$$

grounded in E_C does not enlarge the dynamical state space.

Thus the same eight states can support a minimally self-conscious higher subject.

Only when explicit reflective attention itself is required to vary independently do we need another bit:

$$s \in \{0, 1\},$$

giving:

$$16$$

states.

The emerging hierarchy is therefore:

$$\boxed{4 : \text{relation}}$$

$$\boxed{8 : \text{diachronic subjecthood}}$$

$$\boxed{8 + \sigma : \text{indexical self-consciousness}}$$

$$\boxed{16 : \text{independently variable reflection.}}$$

The deepest lesson of these toy theorems is not the particular numbers 8 and 16.

It is the structural asymmetry they reveal:

$$\boxed{\text{temporal persistence requires genuinely additional state, whereas self-reference may only require a new intentiona}}$$

In this finite Husserlian monadology, memory is more expensive than "I."

That result points toward a broader conception of mindlike monads as **diachronically constituted processes whose identity and phenomenology depend upon the preservation of intentional structure through time**. The next challenge is to determine how much additional structure is required before such a subject can do more than remember and recognize itself—before it can represent what is not, imagine what could be, and intentionally bring a novel possibility into existence.