

From Prediction to Self: Developmental Conditions for Agency in Minimal Neural Systems

Evan Ye

Independent Researcher

ye635498222@gmail.com

Abstract

How does a system that merely predicts the world come to distinguish its own causal influence from everything else? We trace this transition in a minimal 192-dimensional GRU through 40 controlled experiments arranged as a developmental sequence. Starting from a system with no action and no self-representation, we add components one at a time — causal action loops, proprioceptive channels, learnable action policies — and track, at each stage, whether the system can distinguish self-caused from world-caused changes in its observations.

The developmental path reveals four necessary conditions that must be satisfied in strict order: (1) persistent state that forms stable attractors, (2) a causal action loop linking the system’s output to its input, (3) proprioceptive feedback that makes implicit causal knowledge explicit, and (4) asynchronous awakening — perceptual learning must consolidate before action learning begins. We propose *agency gain* ($\mathcal{A} = \text{Err}_{\text{world}} - \text{Err}_{\text{self}}$), the predictive advantage of knowing one’s own action, as a metric to track this developmental process. In the final configuration, the self-aware predictor consistently outperforms the self-blind predictor across both periodic (sinusoidal) and chaotic (Lorenz) environments, and the metric survives ablation of all auxiliary components. Only forward-sampled action selection produces meaningful agency gain; two gradient-based alternatives degenerate.

Equally significant are the 12 falsified hypotheses that map where development stalls: predictive coding alone does not produce self-representation, passive memory cannot sustain post-action state, complex probes cannot extract what is not encoded, and awareness and intention cannot be co-learned. These negative results delineate the boundary between systems that predict and systems that know they are the ones predicting. Moreover, the system sustains self-representation only when it is causally useful: after the external training signal is removed, the causal agent retains its encoding (94.9%) while a statistically-matched control collapses to chance (53.9%).

Keywords: self-world decomposition, agency gain, developmental sequence, predictive coding, negative results, causal attribution

1 Introduction

A system that predicts the weather does not know it is predicting the weather. It maps inputs to outputs — past observations to future estimates — without any representation of itself as a distinct causal entity in the process. Adding actions changes this situation fundamentally: when a system’s outputs feed back into its inputs through the world, the observations it receives are no longer purely

external. Some of the changes it observes are consequences of what it did. The question this paper addresses is: under what conditions does a predictive system come to distinguish these self-caused changes from world-caused changes?

This question sits at the intersection of several research programs. Predictive processing and the free energy principle Friston (2010) propose that organisms minimize prediction error through perception and action, with efference copies von Holst and Mittelstaedt (1950) canceling self-caused sensory changes. Empowerment Klyubin et al. (2005) measures the channel capacity between actions and future states. Curiosity-driven exploration Pathak et al. (2017) uses prediction error as intrinsic reward. Developmental robotics Oudeyer and Kaplan (2007) studies how sensorimotor competence unfolds over time.

What is missing from these programs is a systematic, empirical account of how self-world decomposition develops from scratch — what the minimal conditions are, what order they must appear in, and what plausible-sounding alternatives fail. The theoretical proposals are rich, but the experimental base is thin: we do not know, in any concrete system, the precise boundary between “a system that predicts” and “a system that knows it is the one predicting.”

This paper provides that account. Using a minimal architecture — a 192-dimensional GRU with multi-scale dynamics, fewer than 100K parameters — we conduct 40 controlled experiments arranged as a developmental sequence. Starting from a system with no action and no self-representation, we add components one at a time and observe what changes. Each experiment is motivated by a question that the previous experiment raised, and each is evaluated against pre-registered scorecards with explicit PASS/FAIL criteria.

The central finding is that self-world decomposition requires four conditions satisfied in strict order: persistent state, causal action loop, proprioceptive feedback, and asynchronous awakening. Twelve alternative approaches all fail, for reasons we characterize precisely. To quantify the decomposition at each stage, we propose agency gain: the gap in prediction error between a model that knows the system’s action and an otherwise-identical model that does not. This metric is measurable, ablatable, comparable, and environment-agnostic.

We make no claims about consciousness, sentience, or subjective experience. We claim only to have mapped the conditions under which a minimal predictive system learns to distinguish its own causal influence from the rest of its world.

2 Method

2.1 Experimental Paradigm

All experiments share a common structure. A world produces a multi-channel signal. A model receives the signal, maintains persistent internal state, and predicts the next observation. We intervene — adding an action loop, changing the architecture, ablating a component — and observe the consequences through quantitative metrics.

Each experiment changes one variable relative to its predecessor. This allows precise attribution: if metric X changes when component Y is added, and only when component Y is added, the change is attributable to Y .

2.2 World

The default environment is a 4-channel sinusoidal signal. Each channel consists of two frequency components with distinct base frequencies and amplitudes, plus Gaussian noise ($\sigma = 0.05$). The frequencies are chosen to be incommensurate, preventing simple periodic prediction. When an action loop is present, the action modifies one channel: $\text{obs}[0] += \gamma \cdot a(t)$, where $\gamma = 2.0$.

2.3 Model

The core model is a GRU (Gated Recurrent Unit) with 192 hidden dimensions, augmented with multi-scale exponential moving average (EMA):

$$h_{\text{multi}}(t) = (1 - \alpha) \cdot h_{\text{multi}}(t - 1) + \alpha \cdot \text{GRU}(x(t), h_{\text{gru}}(t - 1)) \quad (1)$$

where α spans four timescales from 0.02 (slow, long memory) to 0.80 (fast, immediate response), with 48 hidden dimensions per scale. The state h_{multi} is never reset — it runs continuously across all training steps, accumulating a persistent representation of the system’s entire history.

A prediction head maps h_{multi} to the predicted next observation. When dual heads are present (Section 3.6), **pred_A** receives both h_{multi} and the action value, while **pred_B** receives only h_{multi} .

The system architecture is illustrated in Figure 1.

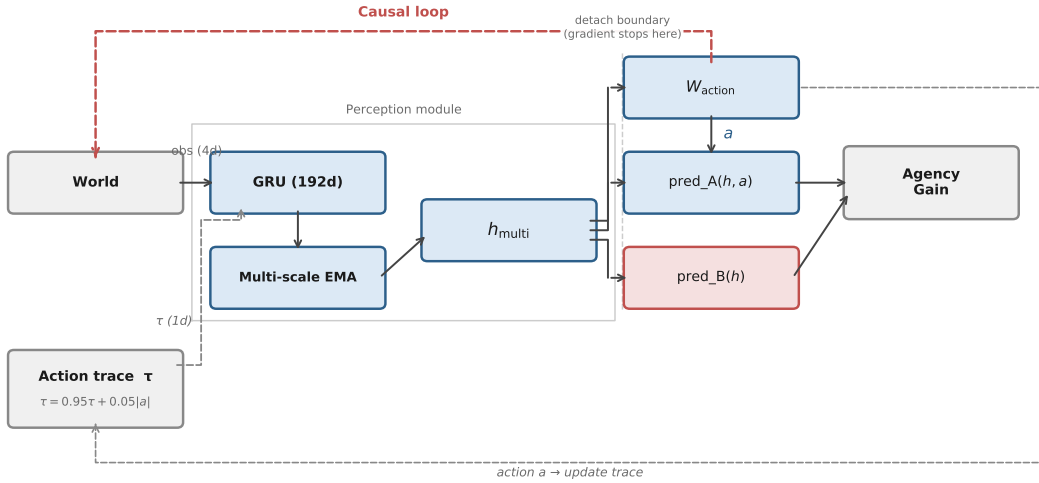


Figure 1: System architecture. Perception module (blue): GRU (192d) + multi-scale EMA $\rightarrow h_{\text{multi}}$. The action-aware predictor $\text{pred_A}(h, a)$ uses the action as an explicit input; the action-blind predictor $\text{pred_B}(h)$ provides the contrastive baseline. Agency Gain $\mathcal{A} = \text{Err}_{\text{world}} - \text{Err}_{\text{self}}$ is computed from their divergence. Dashed red arc: causal loop. Dashed vertical line: detach boundary (gradient isolation).

2.4 Causal vs. Control Design

Experiments that test causal attribution use a matched-control design. The Causal group has action $a(t) = f(h_{\text{multi}})$ — the action is a function of the system’s internal state, creating a genuine causal loop. The Control group replaces this with AR(1) noise matched in mean, variance, and autocorrelation to the Causal group’s actions. This preserves the statistical properties of the signal while breaking the causal link between the system’s state and its action.

2.5 Agency Gain

To quantify self-world decomposition, we define agency gain as the predictive advantage of knowing one’s own action:

$$\mathcal{A}(t) = \underbrace{\|o(t+1) - \hat{o}_{\text{world}}(t+1)\|^2}_{\text{Err}_{\text{world}}} - \underbrace{\|o(t+1) - \hat{o}_{\text{self}}(t+1)\|^2}_{\text{Err}_{\text{self}}} \quad (2)$$

where \hat{o}_{self} is the prediction of a model that knows the action, and \hat{o}_{world} is the prediction of an otherwise-identical model that does not. When $\mathcal{A} > 0$, the action carries causal information that improves prediction. The prediction gap is defined as $(\text{Err}_{\text{world}} - \text{Err}_{\text{self}})/\text{Err}_{\text{world}}$.

The spike test verifies causality by disconnecting the action and measuring the increase in Err_{self} :

$$\text{spike} = \frac{\text{Err}_{\text{self}}^{\text{action disconnected}}}{\text{Err}_{\text{self}}^{\text{normal}}} \quad (3)$$

A spike significantly greater than 1.0 confirms that the prediction advantage depends on the action’s causal effect, not on statistical artifacts.

2.6 Training Protocol

Training follows a predict-then-update protocol: at each step, the model first predicts the current observation from its old state, then computes the loss, then updates weights via backpropagation, and finally updates h_{multi} with the current observation (under no gradient). This prevents information leakage — the model cannot “see” the observation before predicting it.

State updates are performed without gradient tracking to prevent backpropagation through time across the action pathway, ensuring that action-learning gradients do not corrupt perceptual representations.

3 The Developmental Path

Figure 2 shows the complete developmental chain of the six experiments. Each box corresponds to one stage; the chain proceeds left-to-right. Red-bordered box marks the Encoding Gap (Exp. 3); green-bordered box marks the Proprioceptive Breakthrough (Exp. 4).

3.1 Perception: Stable Attractors

We begin with the simplest possible system: a continuously-running GRU predicting a sinusoidal signal. No action, no self/world distinction — just prediction. The question is whether persistent prediction alone produces stable internal structure.

Result. The system forms stable low-dimensional attractors (Figure 3). PCA analysis reveals that 95% of the variance in h_{multi} is concentrated in 5 dimensions out of 192 (effective dimensionality $< 3\%$). The multi-scale EMA produces hierarchical temporal structure: power spectrum analysis confirms that the four timescale groups specialize in different frequency bands. Perturbation recovery reaches 95.0% — after injecting noise into h_{multi} , the system returns to its attractor within hundreds of steps. The scorecard passes 6 of 7 tests (the exception being residual autocorrelation, indicating the GRU has not fully extracted all predictable structure from the signal).

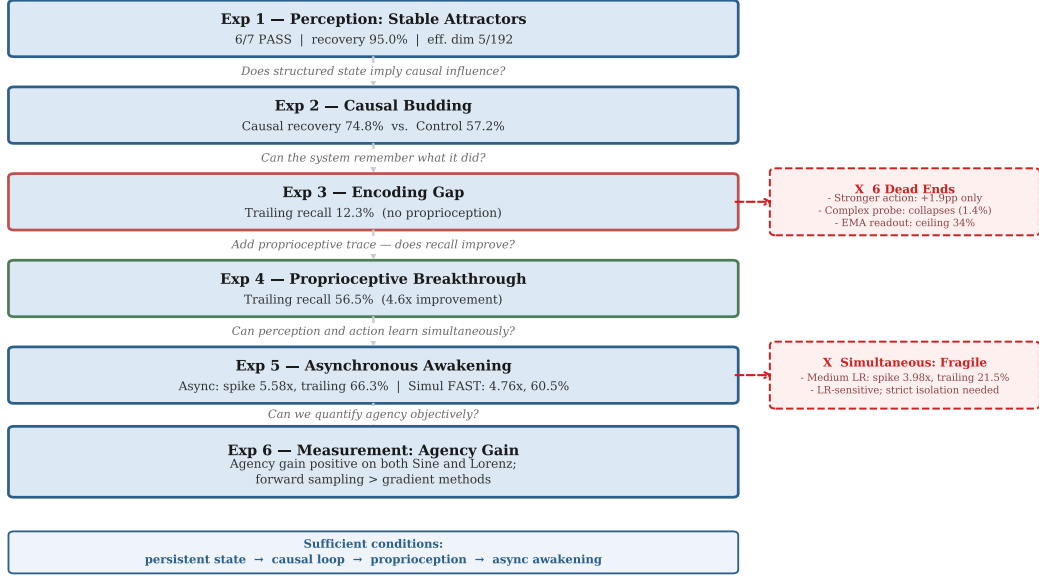


Figure 2: Developmental chain of the six experiments. Red-bordered box (Exp. 3) marks the Encoding Gap bottleneck; green-bordered box (Exp. 4) marks the Proprioceptive Breakthrough. Right-hand dashed boxes show dead-end variants. The bottom bar states the four sufficient conditions. Self-maintenance (Exp. 4b): Causal 94.9% vs. Control 53.9%.

Significance. Persistent prediction with multi-scale dynamics is sufficient to form stable, low-dimensional, perturbation-resistant internal structure. This structure is the foundation for everything that follows.

Question raised. The system has stable internal structure but no action. What happens when its behavior can change the world?

3.2 Causal Budding: Implicit Self-World Decomposition

We add a causal action loop: the system generates an action $a(t) = f(h_{\text{multi}})$ through a linear projection, and this action modifies one observation channel: $\text{obs}[0] += \gamma \cdot a(t)$, with $\gamma = 2.0$. Channels 1–3 are unaffected. The system still has a single prediction head. The question is whether the system implicitly learns which observation changes are self-caused.

Result. Disconnecting the action produces a channel-specific spike: prediction error on channel 0 rises by a factor of 13.8, while channels 1–3 are unaffected. The system has learned, without any explicit supervision, that channel 0 is the one its action influences. Self/world decomposition exists implicitly in the single-head architecture.

Causal verification. To resolve the ambiguity between causal attribution and distributional surprise, we conduct a long-disconnect test (2,000 steps with action removed) comparing Causal and Control groups (Figure 4). The Control group uses AR(1) noise with matched statistics, so removing it produces the same distributional shift. Result: the Causal group recovers to 74.8% of baseline prediction accuracy after 2,000 steps; the Control group recovers only 57.2%. The Causal group recovers differently from the Control group, a behavioral observation consistent with causal attribution. The decisive evidence, however, comes from the self-maintenance test below: only the

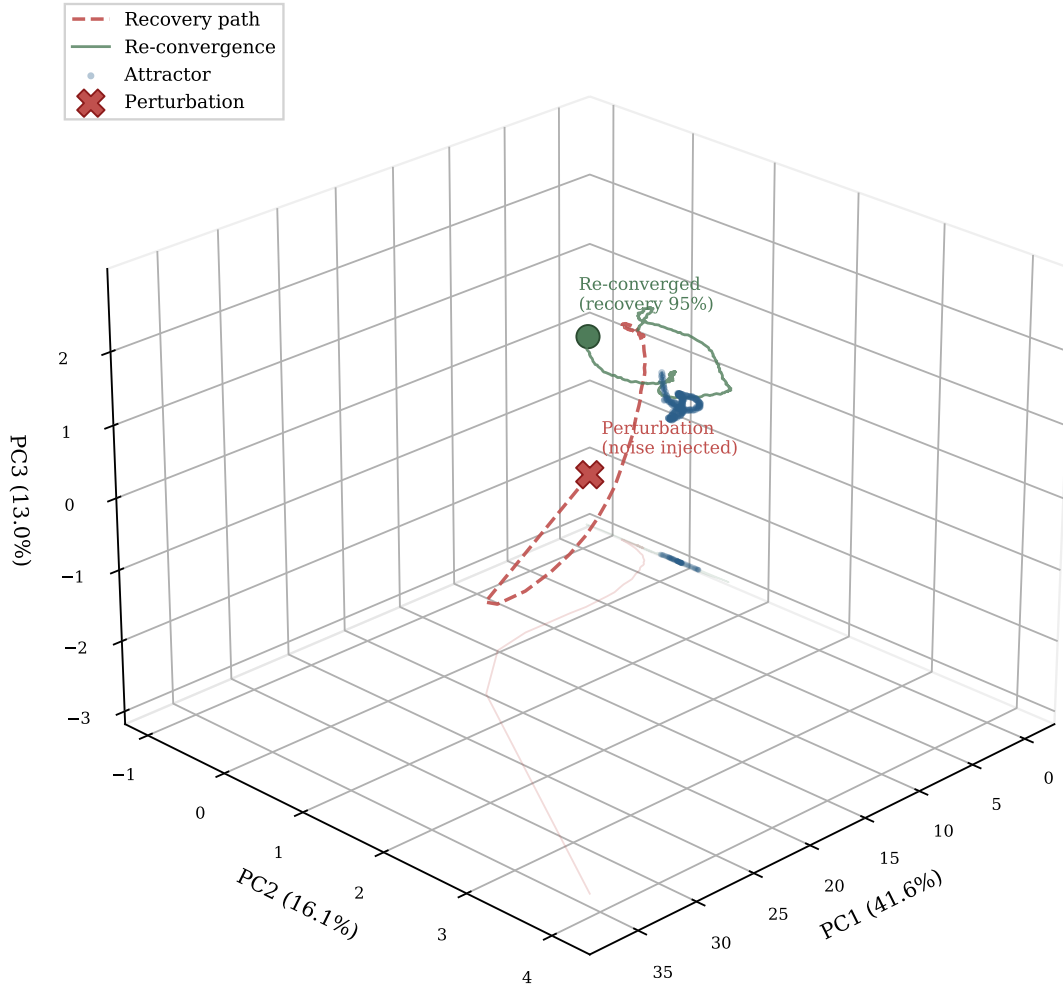


Figure 3: PCA projection (3 components, 70.8% variance) of h_{multi} across 30 000 training steps on the sinusoidal signal. Blue cloud: stable attractor region. Red dashed path: recovery trajectory after noise perturbation at step 15 000. Green path: re-convergence. Recovery rate: 95.0%.

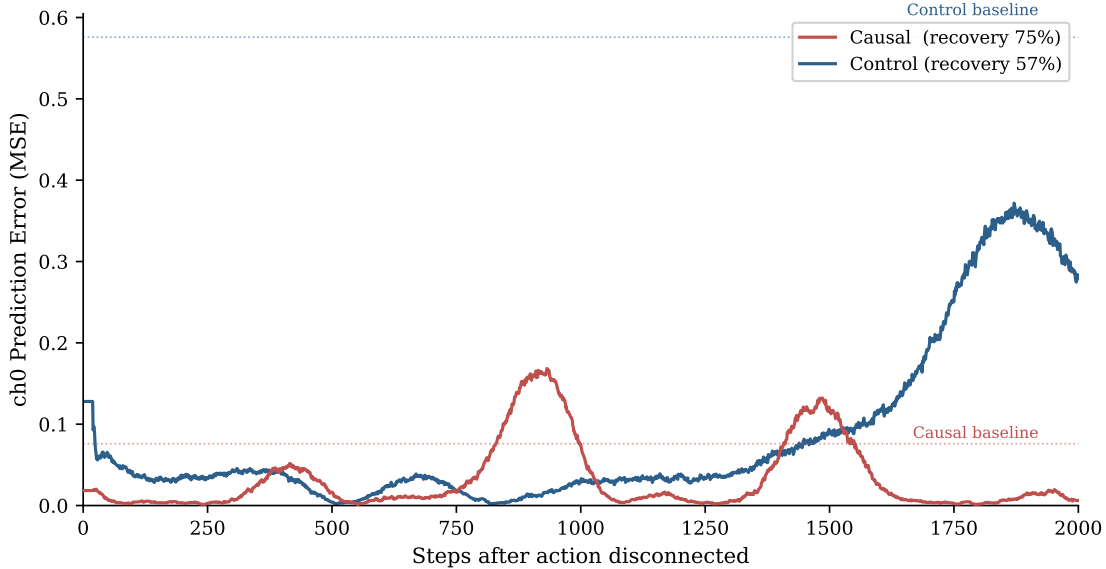


Figure 4: Channel-0 prediction error (MSE, 20-step rolling average) after action disconnection. The causal model (red) recovers to near-baseline whereas the control (blue, AR-process action) drifts. Recovery metric: causal 74.8% vs. control 57.2%.

causally-grounded encoding survives removal of the training signal.

Significance. A single-head predictive system with a causal action loop implicitly learns self-world decomposition. This decomposition is real (verified by Causal vs. Control), channel-specific, and emerges without any architectural prior.

Question raised. The system “knows” implicitly which changes it causes. But can a probe read “I am acting” from h_{multi} ?

Self-maintenance test. To test whether the self-representation depends on an external training signal or is self-sustaining, we ran an ablation on the original architecture (3-channel signal, no trace input, a binary auxiliary loss that explicitly forces the GRU to encode whether it is currently acting — the burst-active state; the burst gate alternates between acting and quiet periods, detailed in Section 3.3).¹ Phase 2 pushes the Causal group’s classification accuracy to 95.3% and the Control group’s to 91.9%; the auxiliary loss is then removed and the system runs 5,000 steps with all parameters frozen. The Causal group retains 94.9% (near-lossless), while the Control group collapses to 53.9%—near chance. The system sustains self-representation only when it is causally useful: when the encoding serves ongoing prediction through the causal loop, the GRU’s learned dynamics maintain it without external pressure; when it has no predictive utility (Control), it decays the moment the training signal is removed.

3.3 The Encoding Gap

We train a linear probe on h_{multi} to classify whether the system is currently acting or quiescent.

¹This experiment uses the same burst-gate mechanism but in a different setting from the failed Hypothesis 12 in Section 5.2 (2-D grid, counterfactual directional prediction, no valid action channel during stillness); the two results are not contradictory.

Result. Even with an auxiliary classification head, trailing recall — the probe’s ability to detect “recently acted” after action ceases — reaches only 12.3%. The system can perfectly compensate for its own actions in prediction, but h_{multi} does not retain “I was acting” in a readable form once action stops.

Six successive experiments attempt to break through this ceiling:

Table 1: Failed attempts to bridge the encoding gap.

Attempt	Hypothesis	Result
Stronger action ($\gamma \times 1.5$)	Larger effect \rightarrow clearer encoding	+1.9pp
Passive EMA readout	Slow EMA retains trace	Ceiling 34%
GRU probe (complex)	Nonlinear probe extracts signal	1.4%
Character-level CE	Train on rare “i” tokens	0.0%
Stronger action ($\gamma \times 2.0$)	Even larger effect	Marginal
Causal vs. Control ablation	Compare probe accuracy	Causal 94.9% vs. Control 53.9% (self-maintenance); action-state probe still 70%

Row-by-row takeaways: (1) Implicit compensation scales with action strength but hits the same ceiling regardless of scale. (2) Exponential decay erases action history; slow EMA cannot retain the trace. (3) Signal is not weak — it is absent: a nonlinear probe finds nothing to extract. (4) Gradient dominated by 99% majority class; rare tokens carry no useful gradient. (5) Same ceiling, different scale: stronger action shifts magnitude but not the ceiling. (6) Self-maintenance confirms causal knowledge; action-state probe confirms it is implicit.

Key finding. Implicit causal knowledge is not the same as explicit self-representation. Prediction accuracy and self-representation are dissociated.

Question raised. What is missing? What would make the implicit explicit?

3.4 Proprioceptive Breakthrough

We add a single architectural change: an action trace as proprioceptive input to the GRU.

$$\tau(t) = \beta \cdot \tau(t-1) + (1 - \beta) \cdot |a(t)|, \quad \beta = 0.95 \quad (4)$$

The GRU input changes from 4 dimensions (obs only) to 5 dimensions (obs + action trace). No other change.

Result. Trailing recall jumps from 12.3% to 56.5%. A single additional input dimension breaks through a ceiling that six prior experiments could not move.

Symbol grounding. This explicit representation enables grounding the first-person pronoun “i” to causal dynamics. During trailing periods, the system is presented with character sequences such as “i moved”; during quiet periods, it sees “the world changed.” A linear probe achieves balanced accuracy of 80.1%, compared to 64.4% for the Control group (+15.7pp gap). Generalization to unseen sentences (“i jumped,” “i stopped”) reaches 83.8%, demonstrating that the probe learned the mapping from trailing-state dynamics to “i,” not from specific word patterns Harnad (1990).

Significance. The encoding gap requires a new information channel: proprioceptive feedback that directly writes action history into the system’s state. Proprioception transforms implicit causal knowledge (distributed in the weights) into explicit self-representation (readable from h_{multi}).

Question raised. The system now has explicit self-representation. Can it learn to actively control its actions?

3.5 Asynchronous Awakening

We unfreeze W_{action} — the linear projection from h_{multi} to action — and attempt to train it alongside the perceptual system. To select structured actions, we introduce the dual-head architecture here: **pred_A** receives h_{multi} and the action value; **pred_B** receives only h_{multi} . At each step the action that maximises the disagreement between the two heads is chosen (forward-sampled disagreement maximisation). Both the asynchronous and simultaneous conditions use this same dual-head setup; the only difference between them is the training schedule, not the architecture.

Result: simultaneous learning is unstable. Three experiments with different action learning rates show inconsistent results:

Table 2: Simultaneous learning results by learning rate.

LR	Spike	Trailing	Pass?
Fast (10^{-3})	$4.76\times$	60.5%	Yes
Slow (10^{-4})	$2.52\times$	40.3%	Yes (marginal)
Medium (5×10^{-4})	$3.98\times$	21.5%	No

No simultaneous configuration achieves the best result on both metrics.

Solution: temporal separation. Training is divided into three phases:

- **Phase 1** (100K steps): Random actions. W_{action} frozen. $\text{LR} = 10^{-3}$.
- **Phase 2a** (60K steps): W_{action} still frozen. LR drops to 10^{-4} . Perception consolidates.
- **Phase 2b** (60K steps): W_{action} unfrozen. Action policy trains on a stable perceptual foundation.

Result: asynchronous learning produces the best result. Spike ratio $5.58\times$ (highest of all configurations), trailing recall 66.3% (highest of all configurations).

Significance. Temporal separation of perception and action learning produces the most robust results. The developmental sequence — perception first, then action — is the reliable path Rochat (2001).

Question raised. How do we quantify how much knowing one’s own action actually helps?

3.6 Measurement: Agency Gain

The dual-head architecture, already introduced in Section 3.5 for action selection, now serves a measurement purpose: the prediction gap between **pred_A** and **pred_B** directly quantifies agency gain. **pred_A** receives h_{multi} and the action; **pred_B** receives only h_{multi} . Both are trained to predict the same target. Actions continue to be selected via forward-sampled disagreement maximisation:

from four candidates (action base, base \pm perturbation, zero), the system executes whichever produces the largest difference between **pred_A** and **pred_B** predictions.

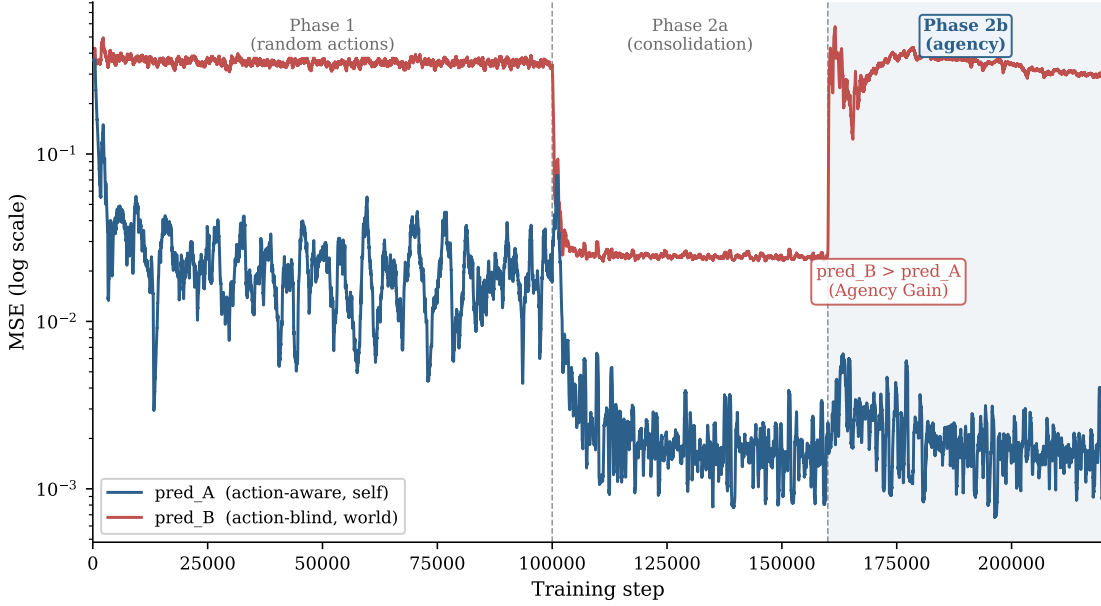


Figure 5: Three-phase training curves on the Lorenz signal (log MSE). **pred_A** (blue, action-aware) remains below **pred_B** (red, action-blind) across all three phases. The Phase 1 gap reflects mechanical compensation of known random perturbations (no causal dependence); the Phase 2b gap accompanies structured action selection (genuine causal dependence). Agency gain is measurable and positive throughout. Spike ratios quantifying the causal dependence are reported in Sections 3.6 and 6.1.

Action strategy comparison.

Table 3: Action selection strategy comparison on sinusoidal signal.

Strategy	pred gap	Spike	Autocorr	Behavior
Forward-sampled	80.7%	$17.32\times$	0.788	Structured
Direct AG gradient	-2.0%	$0.98\times$	1.000	Degenerate
Gradient disagree	-1894.5%	$0.02\times$	0.972	Catastrophic

Direct optimization of agency gain as a gradient objective degenerates — the policy finds the trivial solution of minimizing Err_{self} alone.

Ablation: proprioception.

Table 4: Agency gain with and without proprioceptive trace.

Config	pred gap	Spike	Autocorr
With trace (obs+ τ , 5d)	80.7%	$17.32\times$	0.788
Without trace (obs, 4d)	81.1%	$23.52\times$	0.766

Agency gain survives removal of the proprioceptive channel. Proprioception is an enhancer for explicit self-representation (Section 3.4) but not a requirement for agency gain measurement.

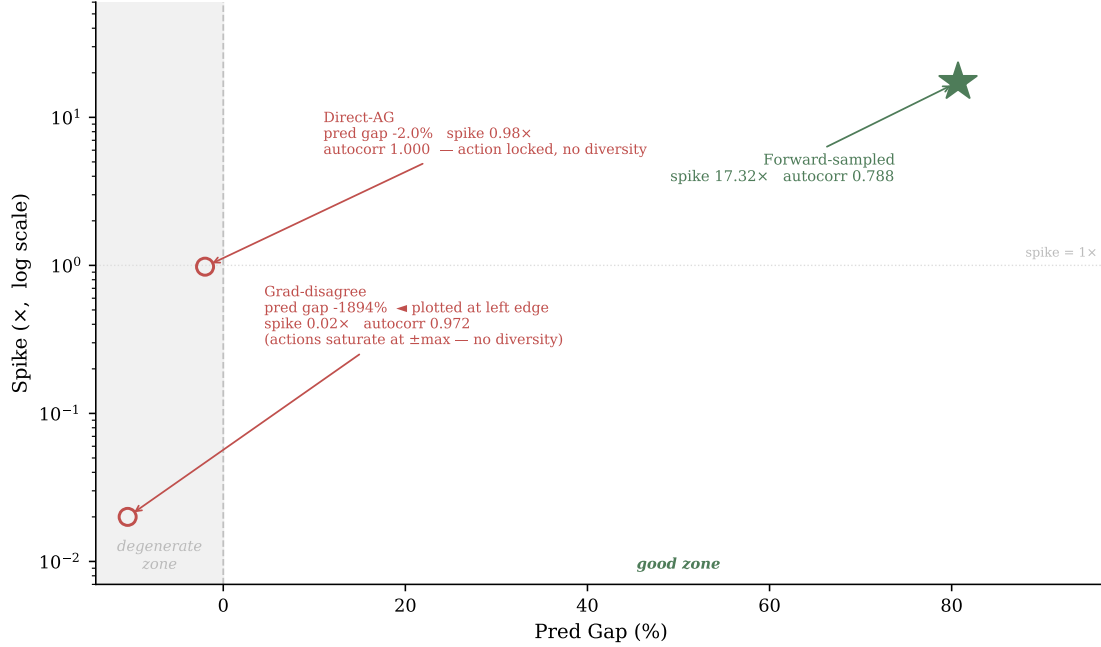


Figure 6: Action strategy comparison. Forward-sampled disagreement (green star) achieves pred gap = 80.7% and spike 17.32 \times , escaping the degenerate zone (grey shading). Both gradient-based alternatives collapse into the degenerate zone: Direct-AG (autocorr 1.000, action locked) and Grad-disagree (autocorr 0.972, actions saturate at $\pm\max$). Horizontal axis uses Plan-B scaling; the Grad-disagree point is pinned at the left edge with its true pred gap = -1894.5% annotated.

Cross-signal validation.

Table 5: Agency gain across signal types.

Signal	pred gap	Spike	Autocorr
Sinusoidal	80.7%	17.32 \times	0.788
Lorenz	99.5%	141.95 \times	0.762

The higher absolute numbers on Lorenz reflect the signal’s smaller normalized amplitude relative to the action range, not a stronger causal effect. Agency gain is computable and positive on both periodic and chaotic signals (Figure 5).

4 Unified Framework

The six developmental stages can be unified under an information-theoretic interpretation. Each stage corresponds to the emergence of a distinct information-theoretic quantity.

Stage 1 — Perception: The system learns a compressed latent $Z_t = g(H_t)$ that maximizes $I(Z_t; O_{t+1})$ (attractor formation, 6/7 scorecard).

Stage 2 — Causation: Action creates conditional mutual information $I(O_{t+1}; A_t \mid H_t) > 0$ (channel-specific spike, confirmed causal).

Stage 3 — Encoding gap: The causal information exists in the system (Stage 2) but is not encoded in the state: $I(Z_t; S_t^{\text{self}}) \approx 0$, where S_t^{self} denotes the system’s self-action state (whether it is currently or recently acting). The system compensates for its actions through distributed weight dynamics without creating an explicit self-variable (trailing recall 12.3%, near chance; the overall action-state probe reaches 70% only because the easy active and quiet states inflate it).

Stage 4 — Awareness: Proprioceptive feedback makes the implicit explicit: $I(Z_t; S_t^{\text{self}}) \gg 0$ (trailing recall 12.3% \rightarrow 56.5%, symbol grounding BA 80.1%).

Stage 5 — Intention: Under Gaussian assumptions, agency gain approximates the conditional mutual information:

$$\mathcal{A} \approx I(O_{t+1}; A_t | H_t) = H(O_{t+1} | H_t) - H(O_{t+1} | H_t, A_t) \approx \text{Err}_{\text{world}} - \text{Err}_{\text{self}} \quad (5)$$

Stage 6 — Measurement: The spike test implements a do-operation: $\text{do}(A_t = 0)$. A spike confirms that the measured agency gain reflects genuine causal influence, not statistical correlation Pearl (2009).

Comparison with related measures:

Table 6: Comparison of agency measures.

Measure	Density est.	Diff.	Self/world
Empowerment Klyubin et al. (2005)	Required	Approx.	No
Curiosity Pathak et al. (2017)	No	Yes	No
Free energy Friston (2010)	Required	Approx.	Implicit
Agency gain	No	Yes	Yes

5 Conditions for Self-World Decomposition

5.1 Sufficient Conditions

Table 7: Four sufficient conditions for self-world decomposition, in required order.

#	Condition	Without it	With it
1	Persistent state (never reset)	No attractors	Stable structure (6/7 scorecard)
2	Causal action loop	No agency	Recovery 74.8% vs. Control 57.2%
3	Proprioceptive feedback	Encoding gap (trailing 12.3%)	Explicit representation (trailing 56.5%)
4	Asynchronous awakening	Fragile, LR-dependent	Robust: spike 5.58 \times , trailing 66.3%
(5)	Dual-head (measurement)	No quantitative metric	pred gap 80.7–99.5%, spike 17–142 \times

Conditions 1–4 must be satisfied in order. Proprioception without a causal loop provides no action information to encode. Asynchronous awakening without stable perception provides no foundation for action learning. The overall developmental chain is illustrated in Figure 2.

5.2 Twelve Falsified Hypotheses

Each of the following hypotheses was a plausible candidate pathway; all twelve were tested and falsified, delineating the boundary of what the minimal system can and cannot achieve.

Table 8: Twelve falsified hypotheses: what was tried, why it fails, and the evidence.

#	Hypothesis	Why it fails	Evidence
1	FC network \rightarrow modularity	No selection pressure	Self/world overlap \approx random
2	Stronger action \rightarrow explicit encoding	Implicit compensation scales with action	+1.9pp only
3	Passive EMA \rightarrow post-action memory	Exponential decay, no reactivation	Ceiling 34%
4	Complex probe for weak signals	Collapses to majority class	GRU probe 1.4%
5	Character-level CE \rightarrow rare token learning	Gradient dominated by 99% majority	I-accuracy 0.0%
6	Trailing self-sustains without aux loss	Weak physical anchor	BA 80% \rightarrow 70%
7	Single channel \rightarrow other-agent model	Blind source separation at low SNR	B-trailing < 3%
8	Awareness + intention co-learnable	Gradient interference	LR dilemma
9	Gumbel-Softmax \rightarrow discrete action	Jacobian disperses signal	Autocorr = -0.007
10	Shared head \rightarrow spatial action	Prediction-action conflict	Spike < 1.6
11	Residual perception \rightarrow better attribution	Stillness-detector shortcut	Spike inverts
12	Counterfactual + burst gate (2-D grid; stillness lacks action channel)	No valid action channel during quiet periods	Phase 2 collapses

6 Discussion

6.1 The Dual-Head Architecture: Thermometer, Not Source

A natural objection is that the dual-head design creates rather than detects self-world decomposition. This objection conflates opportunity with realization. The architecture guarantees only that **pred_A** has access to action information. It does not guarantee that **pred_A** will develop any genuine causal dependence on the action, that the action policy will be structured, or that the system will select actions to sustain its causal identifiability. These are emergent properties of training — as the Phase 1 baseline below shows, action access alone produces a large gap through mechanical compensation, but no causal dependence.

The Phase 1 baseline directly addresses the “unfair prior” objection. Even under random actions, **pred_A** achieves a large prediction gap — but this reflects mere mechanical compensation of a

known perturbation, not causal agency. The two phases are distinguished not by gap magnitude but by the spike ratio (the error increase when the action input is removed):

	Pred gap	Spike
Phase 1 (random actions)	98.8%	0.95×
Phase 2b (trained policy)	80.7%	17.32×

The gap fails to distinguish the two phases — both are 80–99%. The spike does: a Phase 1 spike of 0.95× (indistinguishable from 1.0) shows `pred_A` has built no causal dependence on the action, while the Phase 2b spike of 17.32× confirms the trained policy generates actions `pred_A` genuinely depends on. Agency gain requires not just access to action information, but that the information be causally structured.

6.2 The Encoding Gap: Prediction \neq Self-Representation

The encoding gap (Section 3.3) is perhaps the most important finding for theories of self-awareness. A system can achieve strong implicit agency — perfectly compensating for its own actions in prediction — while achieving only 70% on an explicit self-classification probe (trailing recall 12.3%, near chance). Six experiments confirm this is not a measurement limitation but a genuine representational absence.

This dissociation has implications beyond our minimal system. It suggests that predictive competence and self-representation are fundamentally different capabilities, requiring different architectural support.

6.3 Asynchronous Awakening: A Robust Strategy

The finding that temporal separation produces the most robust results reflects a practical constraint on shared-representation systems. When perception and action train simultaneously, results are fragile and sensitive to learning rate. Asynchronous training consistently produces the highest spike and trailing recall across all configurations.

This parallels observations in infant development: stable sensory processing precedes intentional motor control by months Rochat (2001). Our minimal system suggests this developmental sequence may reflect an optimization landscape constraint.

6.4 Correspondence with Contemplative Phenomenology

While all architectural decisions were made on engineering grounds, the emergent developmental sequence shows a post hoc alignment with the Yogācāra school’s layered consciousness model:

Table 9: Correspondence with Yogācāra consciousness layers.

Contemplative concept	Engineering component	Evidence
Store consciousness	Never-resetting h_{multi}	Stable attractors
Body faculty	Proprioceptive channel	Encoding breakthrough
Corollary discharge	Action-conditioned pred	Channel-specific spike
Self-appropriation	Agency gain	pred gap 80.7%
Calm before insight (samatha-before-vipassanā)	Async awakening	Awareness before intention

These correspondences were recognized after the experiments, not before. We note them as evidence that the developmental constraints identified here may have broader applicability, while acknowledging that the parallel is suggestive rather than explanatory.

6.5 Self-Representation Is Sustained Only When Causally Useful

The self-maintenance result (Section 3.2) reveals a selection pressure on representations. When the encoding of burst-active state serves ongoing prediction through the causal loop, the GRU’s learned dynamics maintain it without any external training signal: removing the auxiliary loss leaves accuracy at 94.9%. When the same encoding has no predictive utility — as in the control group, where the action statistics are matched but the causal loop is absent — it decays the moment the training signal is removed, collapsing to 53.9%. This asymmetry suggests that causal grounding is not merely a correlate of self-representation but a precondition for its persistence.

7 Limitations

The system operates on 4-channel sinusoidal signals and Lorenz chaotic attractors with fewer than 100K parameters. Actions take effect immediately ($\tau = 0$); a preliminary test with 2-step delay shows the metric remains functional, but the current design passes the delayed action value directly to `pred_A`, meaning it does not test temporal causal reasoning per se. Genuine temporal delay robustness remains future work.

Scale dependence. The absolute magnitudes of prediction gap and spike ratio depend on the ratio of action range to signal amplitude. With action range = 2.0 and signal amplitudes of order 1, the action dominates observable variance, producing large gap values. The scientifically meaningful comparisons are relative: Causal vs. Control, with-trace vs. without-trace, forward sampling vs. gradient methods, async vs. simultaneous. These relative differences are invariant to scaling.

Action autocorrelation. The temporal coherence of Phase 2b actions (autocorrelation ~ 0.8) arises partly from W_{action} reading the slowly-varying h_{multi} , and partly from training. The relative contribution of training vs. structural correlation is not fully disentangled.

Self/other discrimination fails: the system can distinguish “I caused this” from “something else caused this” but cannot identify which other agent caused a change. Two-dimensional intentional action fails across five versions due to gradient conflict in spatial action learning.

Agency gain as defined measures predictive advantage under observational distributions, not causal advantage under interventionist do-calculus Pearl (2009). A fully causal formulation remains future work.

We make no claims about consciousness, sentience, or subjective experience.

8 Conclusion

We have traced the development of self-world decomposition in a minimal predictive system through 40 controlled experiments. The developmental path — from stable perception, through implicit causal attribution, past an encoding gap, to explicit self-representation and finally quantifiable agency gain — is strict: each stage requires specific conditions, and skipping any stage fails.

Four conditions are sufficient, in order: persistent state, causal action loop, proprioceptive feedback, and asynchronous awakening. Twelve alternative approaches fail, for reasons we characterize precisely. The core results are relative comparisons that do not depend on scaling: Causal recovery 74.8% vs. Control 57.2% (Section 3.2), trailing recall 12.3% \rightarrow 56.5% with proprioception (Section 3.4), async spike $5.58\times$ vs. simultaneous best $4.76\times$ (Section 3.5), forward-sampled agency gain positive while two gradient alternatives degenerate (Section 3.6), and self-representation is sustained only when causally useful (Causal 94.9% vs. Control 53.9% after the training signal is removed, Section 3.2). These comparisons hold across both sinusoidal and chaotic Lorenz environments.

The 12 falsified hypotheses are a primary contribution. They map the boundary between systems that predict and systems that know they are the ones predicting. These negative results save future work from 12 dead ends.

Agency gain — the predictive advantage of knowing one’s own action — is the metric that makes this developmental process quantifiable. Its absolute magnitude depends on action-to-signal scaling (see Limitations), but its sign, its sensitivity to ablation, and its relative behavior across conditions constitute the empirical evidence of this paper.

Whether the developmental constraints identified here — the encoding gap, the proprioceptive breakthrough, the strict ordering of awareness before intention — reflect universal principles of self-representation or merely artifacts of the GRU architecture is a question for future work with different architectures and more complex environments.

A Notation and Terminology

Notation

Symbol	Meaning
h_{multi}	Multi-scale EMA persistent hidden state (never reset)
pred_A / pred_B	Action-conditioned prediction head / action-blind prediction head
A (agency gain)	$\text{Err}_{\text{world}} - \text{Err}_{\text{self}}$
pred gap	$(\text{Err}_{\text{world}} - \text{Err}_{\text{self}}) / \text{Err}_{\text{world}}$
spike	$\text{Err}_{\text{self}}(\text{action disconnected}) / \text{Err}_{\text{self}}(\text{normal})$
τ	Action trace: $\tau(t) = 0.95 \cdot \tau(t-1) + 0.05 \cdot a(t) $
α	EMA timescale coefficient, ranging 0.02–0.80
γ	Action gain on obs[0], fixed at 2.0
W_{action}	Linear projection from h_{multi} to action

Terminology

Term	Definition
Self-world decomposition	Distinguishing observation changes caused by the system’s own actions from those caused by the external world.
Agency gain	The predictive advantage that comes from knowing one’s own action: how much better pred_A predicts than pred_B .
Encoding gap	The dissociation between implicitly compensating for one’s actions in prediction and explicitly encoding “I am acting” as a readable state variable in h_{multi} .
Trailing recall	A probe’s ability to detect “recently acted” after action ceases, during the 50-step window following burst-gate deactivation.
Burst gate	The mechanism alternating between active (action on) and quiet (action off) periods.
Trailing period	The 50-step window immediately after the burst gate closes.
Async awakening	Phased training in which perception consolidates first (Phases 1–2a) before action learning begins (Phase 2b).
Self-maintenance	The phenomenon whereby self-representation is spontaneously sustained by the system when — and only when — it is causally useful for prediction.
Forward-sampled disagreement maximisation	Action selection by evaluating four candidates and executing the one that produces the largest prediction gap between pred_A and pred_B .

References

- Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, 2010.
- Stevan Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1–3):335–346, 1990.
- Alexander S. Klyubin, Daniel Polani, and Chrystopher L. Nehaniv. Empowerment: A universal agent-centric measure of control. In *IEEE Congress on Evolutionary Computation*, 2005.
- Pierre-Yves Oudeyer and Frederic Kaplan. What is intrinsic motivation? a typology of computational approaches. *Frontiers in Neurorobotics*, 1:6, 2007.
- Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *International Conference on Machine Learning (ICML)*, 2017.
- Judea Pearl. *Causality*. Cambridge University Press, 2009.
- Philippe Rochat. *The Infant’s World*. Harvard University Press, 2001.
- Erich von Holst and Horst Mittelstaedt. Das Reafferenzprinzip. *Naturwissenschaften*, 37(20):464–476, 1950.