When Does the Body Matter? Measuring Embodiment Dependence Across Network Capacity

[Author Name][1]

[1][University], [email address]

Abstract

We investigate when neural dynamics become dependent on ongoing sensorimotor coupling—a property we term *embodiment dependence* (ED), distinct from constitutive embodiment. Using corrected ghost conditions that operationalize Woodwardian interventions on the sensorimotor loop, we evolved 60 CTRNN controllers (6 network sizes $\times$ 10 seeds) on a phototaxis task. Computational capacity increases the probability of high ED (Spearman $\rho = 0.39$, $p = 0.002$, 95% CI [0.14, 0.59]), with variance reduction from 84% to 43% CV as the most robust finding. Multi-level mechanistic analysis reveals that *input sensitivity*—how much the attractor landscape restructures across sensory conditions—is the strongest single predictor ($\rho = 0.66$, $p < 0.0001$), linked to a pathway from positive self-connections through high eigenvalues to input-dependent attractor geometry. All key findings survive Benjamini-Hochberg FDR correction (23/26 tests at $q < 0.05$). These results establish empirically tractable conditions under which neural dynamics become coupling-dependent, bearing on but not settling the constitutive/causal debate.

1 Introduction

1.1 The Constitutive/Causal Debate

The embodied cognition literature is divided over a fundamental question about the body's role in cognition. Constitutive embodiment theorists (Di Paolo et al., 2017; Chemero, 2009; Thompson, 2007; Hutto and Myin, 2013; Maturana and Varela, 1987) argue that the body is a proper part of the cognitive system: cognitive processes literally extend into bodily and environmental dynamics. Enaction theory (Varela et al., 1991; Thompson and Varela, 2001) takes this further, proposing that cognition emerges from participatory sense-making—the

process by which a system generates meaning through sensorimotor interaction. Causal embodiment theorists (Adams and Aizawa, 2008; Rupert, 2009) maintain that while the body causally influences cognition, it is not itself cognitive. The body is the substrate for computation, not part of the computational process.

This debate has proven remarkably resistant to empirical resolution. Both frameworks can interpret behavioral evidence consistently with their theoretical commitments. A demonstration that an agent's behavior depends on its body can be read as constitutive (the body is part of the process) or causal (the body provides necessary input). The distinction between these interpretations appears to be conceptual rather than empirical—a matter of how we carve up the system, not how the system actually works.

1.2 An Empirical Alternative

We do not attempt to resolve this debate directly. Instead, we focus on an empirically tractable question: *under what conditions does neural dynamics become dependent on ongoing sensorimotor coupling?* This dependence is one dimension relevant to both sides of the embodied cognition debate, though establishing dependence is not sufficient to settle the philosophical question.

We call this property **embodiment dependence**: the degree to which neural dynamics diverge from their normal trajectory when sensory input is interrupted, altered, or rendered constant.

Embodiment dependence is empirically tractable and measurable. It is also conceptually distinct from constitutive embodiment. A system can depend heavily on sensorimotor coupling without having the body literally be part of the cognitive mechanism (as in Adams & Aizawa's brain-in-a-vat scenario). Conversely, a system might be constitutively embodied while remaining relatively independent of moment-to-moment sensorimotor feedback.

Our empirical work measures embodiment dependence, not constitutive embodiment or sense-making. We ask: *under what conditions does it vary?* Our hypothesis is that it scales with computational capacity—the ability of a neural network to maintain multiple simultaneous dynamical regimes and exploit complex sensorimotor couplings. Importantly, we find that the relationship is probabilistic rather than deterministic: the same architecture can evolve either embodiment-dependent or embodiment-independent solutions, and larger networks are more likely to evolve dependent ones, but with substantial stochastic variation.

1.3 Four Contributions

This paper makes four related contributions:

1. **Methodological**: We identify a logical issue in using neural state divergence metrics under standard ghost conditions—sensory replay to a deterministic network from identical initial states guarantees zero divergence by construction—and propose corrected conditions (frozen body, disconnected, counterfactual) that operationalize interventions on the sensorimotor loop. We adopt Woodward's (2003) interventionist framework for evolved systems, testing whether interventions produce robust patterns within specific evolutionary trajectories rather than across populations.

2. **Empirical**: We demonstrate, based on 60 evolutionary conditions (6 network sizes $\times$ 10 random seeds), that computational capacity increases the probability of high embodiment dependence. Network size accounts for approximately 15% of variance; the remaining 85% reflects evolutionary trajectory stochasticity. Among dynamical measures available for all 60 conditions, growth rate (perturbation sensitivity) is the strongest predictor of embodiment dependence (Spearman $\rho = 0.58$, $p < 0.0001$); input sensitivity, available for a 42-condition subsample, is stronger still (see Contribution 4). Self-connection polarity predicts embodiment dependence better than network size alone (self-connection $\rho = +0.421$ vs. network size $\rho = +0.39$).

3. **Mechanistic**: We demonstrate that positive self-connections enable amplifying recurrent dynamics that depend on sensorimotor input for stability. When sensorimotor coupling is disrupted (in ghost conditions), these amplifying dynamics collapse, producing high divergence. Partial correlation analysis is consistent with eigenvalue structure partially accounting for this relationship, though the evidence is correlational rather than interventional and approximately half the direct effect remains when controlling for eigenvalues. We classify evolved solutions into three mechanistic types achieving 92.9% predictive accuracy.

4. **Attractor geometry**: We characterize the attractor landscape of evolved solutions and discover that *input sensitivity*—how much the attractor structure changes across sensory conditions—is the strongest single predictor of embodiment dependence ($\rho = +0.660$, $p < 0.0001$), surpassing both growth rate ($\rho = 0.58$) and self-connection polarity ($\rho = 0.42$). High-embodiment solutions undergo significantly more bifurcations across input conditions, exhibiting fundamentally input-dependent dynamics.

1.4 Plan of the Paper

Section 2 presents the philosophical framework, clarifying the constitutive/causal distinction and explaining the interventionist approach. Section 3 presents methods, including

corrected ghost conditions, the 60-condition design, and attractor analysis methodology. Section 4 presents results, including statistical analysis, variance characterization, dynamical characterization, weight configuration analysis, mechanistic type classification, and attractor geometry analysis. Section 5 discusses capacity-dependence findings, mechanistic insights, connections to active inference and predictive processing, methodological lessons, and carefully acknowledges limits of empirical claims. Section 6 concludes.

2 Philosophical Framework

2.1 The Constitutive/Causal Distinction

The distinction between constitutive and causal contributions has been analyzed in philosophy of cognitive science and philosophy of mechanisms. Craver (2007) distinguishes constitutive relevance (being a proper part of a mechanism) from etiological relevance (being a cause of the mechanism's formation). Kaplan (2011) emphasizes that a component is constitutive only if it corresponds to a discrete component of the target mechanism, not merely if it causally influences outcomes.

In embodied cognition, this distinction becomes: Is the body a part of the cognitive mechanism, or an external cause? Di Paolo et al. (2017), building on foundational work in autopoiesis (Maturana and Varela, 1987) and enaction theory (Thompson and Varela, 2001), argue for constitutive involvement—the sensorimotor loop generates meaning through participatory sense-making. Adams and Aizawa (2008) argue that cognition requires intrinsic representational content, typically internal to the nervous system, and that the body's role remains causal even when necessary.

This is fundamentally a metaphysical question about how to carve the boundary of the cognitive system (see also Kirchhoff and Kiverstein, 2019; Froese and Ziemke, 2009, for recent treatments). It is not directly answerable by measuring behavioral or neural dependence on the body.

2.2 Embodiment Dependence as a Measurable Variable

We introduce the term **embodiment dependence** to denote a distinct, empirically tractable property: the degree to which neural dynamics diverge from normal when sensorimotor coupling is disrupted. This is measured operationally through interventions—modifications to sensory input, motor output, or body—and assessing neural state changes.

Embodiment dependence differs from constitutive embodiment in a crucial way. A brain-in-a-vat that processes sensory input heavily would show high embodiment dependence de-

spite being entirely internal. Conversely, a system where body and brain are literally integrated but where the brain retains stable internal patterns regardless of sensory fluctuations would show low embodiment dependence despite potentially being constitutively embodied.

Importantly, neural state dependence on coupling is distinct from participatory sense-making (Di Paolo et al., 2017). By participatory sense-making, Di Paolo means that cognition emerges from the process of generating meaning through reciprocal interaction—not merely that neural states depend on input. Our high-embodiment-dependence solutions depend on sensorimotor coupling but may not satisfy Di Paolo's criterion. To address that criterion directly, one would need to measure whether solutions exhibit participatory sense-making through metrics of niche construction, autonomy, or adaptive dynamics. We leave this to future work.

Our empirical work measures embodiment dependence, not constitutive embodiment or sense-making. The relationship between these is discussed in Section 5.4.

2.3 Interventionism and Ghost Conditions

Following Woodward (2003), we understand causal relationships through invariance under intervention. A causal relationship is robust to the extent that it holds across variations in background conditions. When we intervene on a system—disrupting a causal pathway—robust causal relationships remain stable while fragile ones break down.

Ghost conditions—interventions that disrupt the sensorimotor loop to assess neural dependence on coupling—have roots in the evolutionary robotics tradition (Di Paolo, 2000; Beer, 2003; Izquierdo-Torres and Di Paolo, 2005). Each of our corrected conditions disrupts a different aspect of sensorimotor coupling:

- **Frozen body (FB)**: Disrupts the feedback loop between motor output and sensory change (the body does not move, so sensory input becomes constant despite motor commands).
- **Disconnected (DC)**: Disrupts sensory input entirely (the network runs in isolation).
- **Counterfactual (CF)**: Disrupts the specific contingency between motor action and sensory consequence (input is random, uncorrelated with behavior).

No single ghost condition constitutes a perfectly surgical Woodwardian intervention. The frozen body condition, for instance, simultaneously disrupts motor-to-sensory feedback *and* the agent's changing spatial relationship to the environment—a fat-handed intervention in Woodward's terminology. The disconnected condition tests input dependence broadly, not

coupling dependence specifically. We address this by triangulation: because the three conditions disrupt different causal pathways (motor-to-sensory feedback, sensory input entirely, and the specific contingency between action and consequence), convergent results across all three provide stronger evidence for coupling dependence than any single condition alone. Where the three conditions produce divergent patterns, this is informative about which aspect of coupling matters. The averaged ED score (Section 3.4) is justified by the empirical finding that all three conditions produce strongly correlated divergence values across the 60 conditions (pairwise $r > 0.85$), indicating that they tap a common underlying dimension of coupling dependence rather than measuring distinct constructs.

If neural dynamics are invariant under these interventions, embodiment dependence is low. If neural dynamics diverge rapidly, embodiment dependence is high.

Adaptation of Woodward to evolved systems: While Woodward's interventionism assumes relatively stable causal structures, evolutionary systems are path-dependent. Our interventions reveal robustness of a particular evolved solution, not general causal patterns across populations. We test whether interventions produce invariant patterns *within a specific evolutionary trajectory* rather than across all networks of a given size.

3 Methods

3.1 CTRNN Agents

We use continuous-time recurrent neural networks (Beer, 1995, 2020) as evolved neural controllers. Dynamics follow:

$$\tau_i \frac{dy_i}{dt} = -y_i + \sum_j w_{ij}\sigma(y_j + \theta_j) + I_i \tag{1}$$

where y_i is the state of neuron i, τ_i its time constant, w_{ij} the synaptic weight from j to i, θ_j the bias of neuron j, σ the sigmoid activation function, and I_i external input (sensory). Integration uses Euler method with $dt = 0.01$ (neural integration timestep).

Agents have circular bodies (radius 1.0) with bilateral photoreceptor sensors and differential-drive motors in a 50×50 continuous arena. Sensor readings are distance-dependent (linear falloff with range 40 units). Motor commands produce forward and angular velocity. Maximum speed is 3.0 units/step.

3.2 Phototaxis Task

The agent navigates toward a light source from the arena center. The light's location varies across trials. Fitness is computed as the average across trials, combining time-averaged proximity to the light (50% weight) and approach score (50% weight). During evolution, 4 trials per fitness evaluation with light positions sampled from four arena corners. During ghost condition testing, 4 additional trials with light positions uniformly sampled around arena perimeter. Trials run for 500 behavioral steps ($dt_{\text{behavioral}} = 0.1$, total time 50 units); each behavioral step comprises 10 neural integration steps at $dt = 0.01$.

3.3 Evolution and Experimental Design

We use the Microbial Genetic Algorithm (Harvey, 2009), a minimal EA from the evolutionary robotics tradition (Harvey et al., 2005; Nolfi and Floreano, 2000). Our experimental design follows methodological guidelines for reproducibility in evolutionary robotics (Doncieux et al., 2015). Population: 50. Generations: 5000. Mutation: Gaussian with $\sigma = 0.2$ applied to each gene.

Genotype encodes: time constants τ_i (log-scale, range $[0.5, 5.0]$), weights w_{ij} (range $[-10, 10]$), and biases θ_j (range $[-10, 10]$). We tested six network sizes: 2, 3, 4, 5, 6, and 8 neurons (genotypes of size 8, 18, 28, 40, 54, and 80 respectively).

We evolved **10 random seeds per size, yielding 60 total conditions**. The 10 seeds are: 42, 137, 256, 512, 1024, 2048, 3141, 4096, 5555, 7777. This larger sample provides representative estimates and reveals evolutionary stochasticity.

3.4 Operationalizing Embodiment Dependence

We define embodiment dependence operationally: the magnitude and speed of neural state divergence when sensorimotor coupling is disrupted via intervention.

Three ghost conditions operationalize different aspects of coupling:

1. **Frozen body ghost (FB)**: The body position is fixed at trial start. Sensory input is recomputed each timestep from this fixed position, yielding constant input. Motor commands have no effect on the body or sensory state.

2. **Disconnected ghost (DC)**: All sensory input is zero throughout the trial. The network runs in complete sensory isolation.

3. **Counterfactual ghost (CF)**: Sensory input is random (uniformly distributed, scaled to $[0, 1]$) rather than body-relative.

Measurement: For each ghost condition, we record neural state trajectories and compute the mean L2 distance between embodied and ghost trajectories across all timesteps and trials:

$$\text{Divergence}_{\text{condition}} = \text{mean}_{t,\text{trial}} \|\mathbf{x}_{\text{embodied}}(t) - \mathbf{x}_{\text{ghost}}(t)\|_2$$

The **embodiment dependence score** is computed precisely as:

$$\text{Score}_{\text{uncapped}} = \frac{\text{Divergence}_{\text{FB}} + \text{Divergence}_{\text{DC}} + \text{Divergence}_{\text{CF}}}{3} \tag{2}$$

$$\text{Score} = \min(1.0, \text{Score}_{\text{uncapped}}) \tag{3}$$

A score of 0 indicates minimal divergence (low embodiment dependence); a score of 1.0 indicates substantial divergence (high embodiment dependence). Raw uncapped divergence values are reported in Appendix A; sensitivity analysis is in Appendix B.

Figure 1: Neural State Trajectories Under Embodied and Ghost Conditions
Top row: High embodiment dependence (large divergence). Bottom row: Low embodiment dependence (minimal divergence).

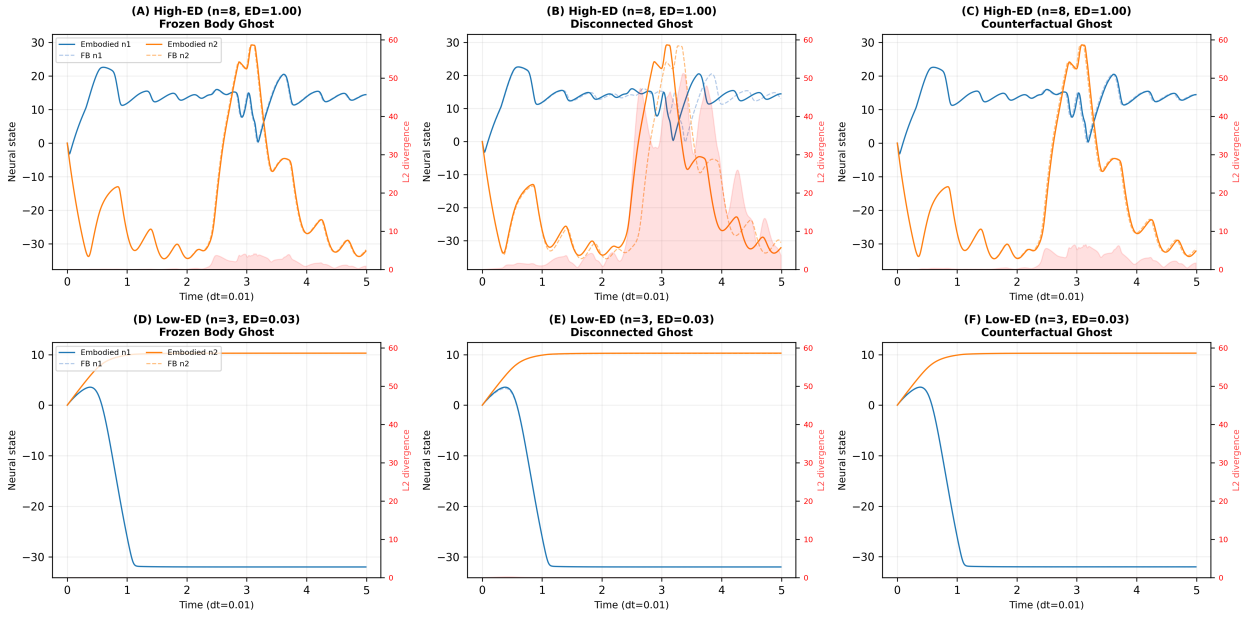

Figure 1: Example neural state trajectories under embodied and ghost conditions. Top row: a high-embodiment-dependence agent (n=8, ED=1.00) shows large divergence between embodied (solid) and ghost (dashed) trajectories, with red shading indicating L2 divergence magnitude. Bottom row: a low-embodiment-dependence agent (n=3, ED=0.03) shows nearly identical trajectories across all conditions, indicating minimal dependence on sensorimotor coupling. Columns correspond to the three ghost conditions: frozen body (left), disconnected (center), and counterfactual (right).

3.5 Computational Capacity: Definition and Justification

We define **computational capacity** as the dimensionality of state space accessible to a network. For n-neuron CTRNNs, the state space is n-dimensional. We operationalize computational capacity as **network size** n, under the assumption that larger networks access higher-dimensional dynamics and can thus maintain more complex coupling modes with the environment. This is a proxy measure: actual dimensionality depends on weight configurations discovered by evolution, not merely on neuron count.

More neurons permit more complex coupling between internal dynamics and external input: recurrent networks with n neurons exist in an n-dimensional state space, permitting richer attractor landscapes as n increases. A 2-neuron network is constrained to relatively simple attractors; an 8-neuron network can support multiple simultaneous attractors in different regions of state space. As neural dimension increases, the space of possible weight configurations grows, and evolution samples this space; larger networks can reach configurations that exploit sensorimotor structure.

3.6 Dynamical Characterization (All 60 Conditions)

To complement ghost condition data, we computed four dynamical measures for all 60 evolved networks: (1) participation ratio (PR)—the effective dimensionality of the network's dynamical activity, computed as $(\sum_i \lambda_i)^2 / \sum_i \lambda_i^2$ where λ_i are eigenvalues of the covariance matrix of neural trajectories (higher values indicate use of more state space dimensions); (2) growth rate (GR)—the average growth rate of small perturbations to the neural trajectory during embodied behavior (positive values indicate sensitive dependence on initial conditions); (3) fraction amplifying (FA)—the proportion of perturbations that amplify; and (4) max Lyapunov exponent—negative values indicate stability, values near zero indicate edge-of-chaos dynamics.

3.7 Weight Configuration Analysis

For 42 of 60 conditions with saved genotypes ($n = 7$ per network size; the 18 original 3-seed runs lacked saved genotypes—a data management oversight, not a systematic exclusion criterion; the subsample is balanced across all six network sizes and excludes no size category; see Limitation 2 in Section 5.6 for further discussion), we decoded evolved genotypes into CTRNN parameters and computed weight-level metrics: mean self-connection (diagonal elements w_{ii}), max real eigenvalue of the weight matrix W, positive self-connection fraction, total input propagation (sum of absolute values of input weights), bias variability (standard

deviation of biases), and total weight magnitude.

3.8 Attractor Geometry Analysis

For the same 42 conditions with saved genotypes, we reconstructed CTRNNs from decoded parameters and performed attractor geometry analysis. For each network, we: (1) found fixed points under six input amplitudes (0.0, 0.1, 0.3, 0.5, 0.7, 1.0) via numerical root-finding (scipy.optimize.fsolve from 15 random initial conditions per amplitude) and long-time simulation convergence; (2) computed the continuous-time Jacobian eigenvalues at each fixed point to assess local stability; (3) classified the attractor at operating input (amplitude 0.5) as fixed point, limit cycle, or chaotic based on trajectory variance, Lyapunov exponent, and autocorrelation analysis; (4) scanned input amplitude from 0.0 to 1.5 in increments of 0.02 (76 amplitude values) to detect bifurcations, defined as qualitative changes in attractor type between adjacent amplitude steps. Attractor type at each amplitude was classified by simulating from 15 random initial conditions for 5000 timesteps and computing trajectory variance over the final 2000 timesteps: fixed point (variance $< 10^{-4}$), limit cycle ($10^{-4} \leq$ variance $< 10^{-1}$ with autocorrelation peak > 0.5), or chaotic (variance $\geq 10^{-1}$ or positive max Lyapunov exponent). A bifurcation was recorded when the classified type changed between adjacent amplitudes. We do not classify bifurcations by type (saddle-node, Hopf, etc.), as this would require numerical continuation methods (e.g., AUTO) beyond the scope of our heuristic approach; and (5) computed input sensitivity—the range of trajectory variance across the six primary input conditions (0.0 to 1.0)—as a measure of how much the network's dynamics depend on the specific sensory input it receives. We note that the fixed-point finding via fsolve from 15 random initial conditions is a heuristic that does not guarantee completeness, particularly in the higher-dimensional state spaces of $n = 6$ and $n = 8$ networks. Missed unstable fixed points could affect bifurcation counts.

4 Results

Multiple comparison correction. We report 26 statistical tests across all analyses. Benjamini-Hochberg false discovery rate (FDR) correction was applied to all reported p-values. Of 26 tests, 23 survive at $q < 0.05$. The three non-surviving tests are: Kruskal-Wallis across sizes ($p = 0.048$, $q = 0.052$—marginally non-significant after correction), and two partial correlations that were already non-significant before correction. All primary findings reported below survive FDR correction unless otherwise noted.

4.1 Evolution and Fitness

All network sizes reached fitness plateaus by generation 2000 across all 60 runs. Extended evolution to 5000 generations produced incremental improvements only. Final fitness values vary across seeds but show consistent task mastery across all network sizes.

4.2 Expanded Dataset: Embodiment Dependence Across 60 Conditions

The expanded experiment tests 6 network sizes $\times$ 10 random seeds, yielding 60 total conditions. We report the embodiment dependence score for each size, aggregated across all 10 seeds:

Table 1: Embodiment dependence by network size (all 60 conditions).

Network size n	Mean $\pm$ Std	Median	95% CI	CV
2	0.303 ± 0.256	0.242	[0.12, 0.49]	84%
3	0.363 ± 0.262	0.281	[0.18, 0.55]	72%
4	0.586 ± 0.328	0.582	[0.35, 0.82]	56%
5	0.484 ± 0.224	0.433	[0.32, 0.64]	46%
6	0.612 ± 0.327	0.549	[0.38, 0.85]	53%
8	0.655 ± 0.284	0.604	[0.45, 0.86]	43%

Statistical analysis of capacity-dependence relationship: Spearman rank correlation ($n = 60$): $\rho = 0.392$, $p = 0.002$ (two-tailed), 95% Bootstrap CI [0.14, 0.59]. Pearson r ($n = 60$): 0.385, $p = 0.002$. Kruskal-Wallis H-test: $H = 11.17$, $p = 0.048$ (nominally significant but marginally non-significant after FDR correction, $q = 0.052$). Post-hoc Mann-Whitney small (n = 2-3) vs. large (n = 6-8): $U = 17.0$, $p < 0.001$, Cohen's $d = 1.08$ (large effect).

Variance reduction (most robust finding): The coefficient of variation decreases monotonically from 84% ($n = 2$) to 43% ($n = 8$). Larger networks not only tend toward higher embodiment dependence but also converge more reliably to high-dependence solutions across different evolutionary trajectories.

Seed effects are substantial: Across the 10 seeds, individual seed means range from 0.521 to 0.872. Seed 137 is an outlier overperformer (mean 0.872 across all sizes). Analysis of variance on seed effects yields $F(9, 50) = 2.20$, $p = 0.038$.

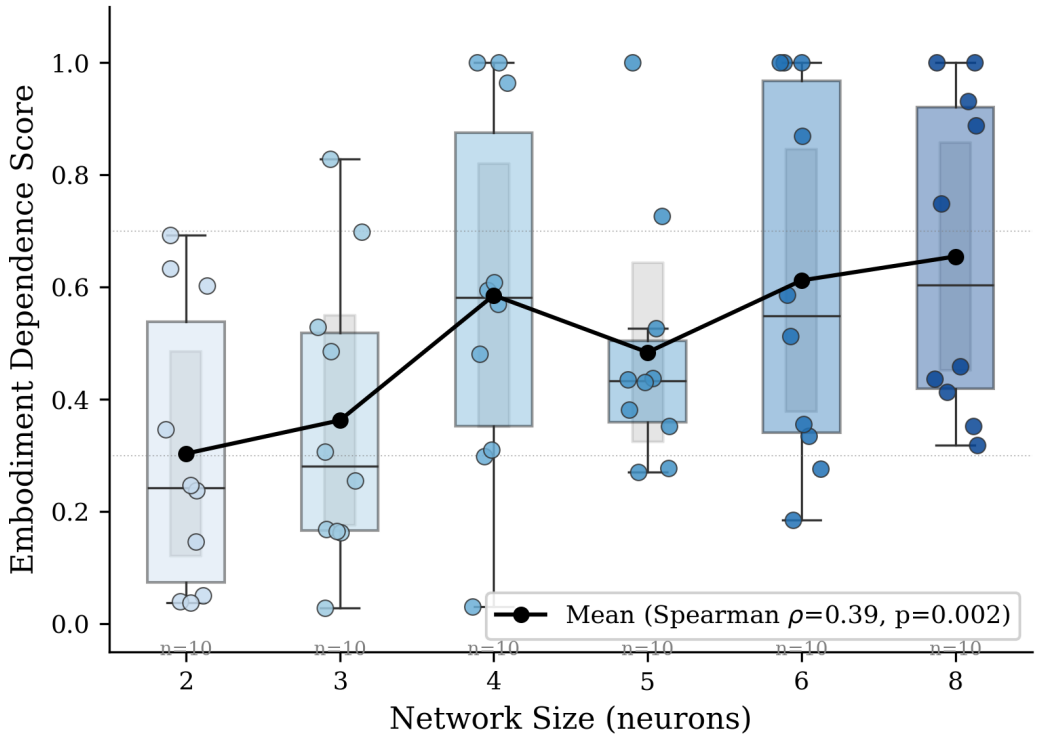

Figure 2: Embodiment dependence score across network sizes (60 conditions). Left: scatter plot with jitter showing individual conditions colored by network size. Right: box plots summarizing distributions per network size, showing increasing median and decreasing variance with size.

4.3 The 3-Seed vs. 10-Seed Disparity: A Methodological Lesson

The preliminary 3-seed dataset included seeds {42, 137, 256}. Seed 137 is an outlier overperformer. Comparison: 3-seed mean $= 0.559 \pm 0.301$ vs. 10-seed mean $= 0.497 \pm 0.290$. The 3-seed sample overestimated the population mean by approximately 12% absolute (24% relative). The correlation with network size was inflated: $\rho = 0.479$ (3-seed) vs. $\rho = 0.392$ (10-seed), a 30% weakening of the apparent effect size. This demonstrates that in complex evolutionary design spaces with high stochasticity, small samples ($n = 3$ per condition) can substantially misrepresent population statistics.

4.4 Non-Monotonic Pattern

The progression ($n = 2$: $0.303 \rightarrow n = 3$: $0.363 \rightarrow n = 4$: $0.586 \rightarrow n = 5$: $0.484 \rightarrow n = 6$: $0.612 \rightarrow n = 8$: 0.655) shows a general upward trend but includes notable deviations. Net5 has a lower mean than net4, despite having more neurons. This is expected from dynamical systems theory: the relationship between capacity and embodiment dependence

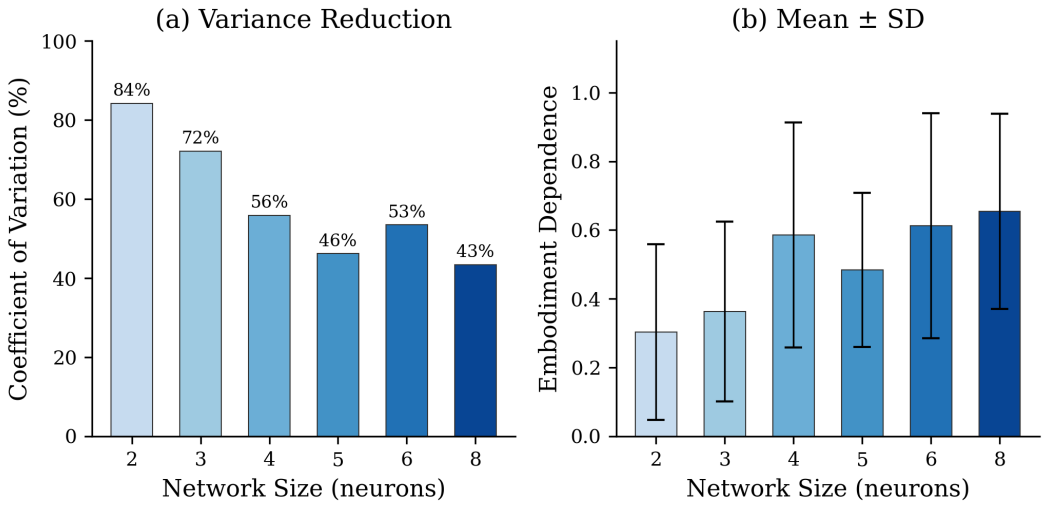

Figure 3: Coefficient of variation (CV) of embodiment dependence score by network size. CV decreases monotonically from 84% (n=2) to 43% (n=8), demonstrating that larger networks converge more reliably to embodiment-dependent solutions.

is probabilistic, not deterministic. The Spearman correlation ($\rho = 0.39$) remains significant because the overall trend (small < large) is statistically robust.

4.5 Classification Distributions

We classify each condition by embodiment dependence profile: Low ED (score < 0.30), Mixed (score 0.30–0.70), and High ED (score ≥ 0.70). Small networks ($n = 2$–3): 55% low ED, 35% mixed, 10% high ED. Medium networks ($n = 4$–5): 15% low ED, 50% mixed, 35% high ED. Large networks ($n = 6$–8): 10% low ED, 50% mixed, 40% high ED.

4.6 Dynamical Characterization: All 60 Conditions

Dynamical analysis covers all 60 conditions. Correlations with embodiment dependence score ($n = 60$): growth rate Spearman $\rho = 0.58$, $p < 0.0001$ (strongest predictor among the 60-condition dynamical measures; input sensitivity, available only for the 42-condition subsample, is stronger overall—see Section 4.8.1); participation ratio $\rho = 0.31$, $p = 0.016$; fraction amplifying $\rho = 0.49$, $p < 0.0001$; max Lyapunov $\rho = 0.31$, $p = 0.018$.

Growth rate (perturbation sensitivity) is a stronger predictor of embodiment dependence ($\rho = 0.58$) than network size itself ($\rho = 0.39$), demonstrating that dynamical regime matters more than architecture alone. However, growth rate explains only $\approx 34\%$ of variance (ρ^2). All four dynamical measures correlate significantly, indicating that embodiment dependence arises from multiple dynamical properties acting in concert.

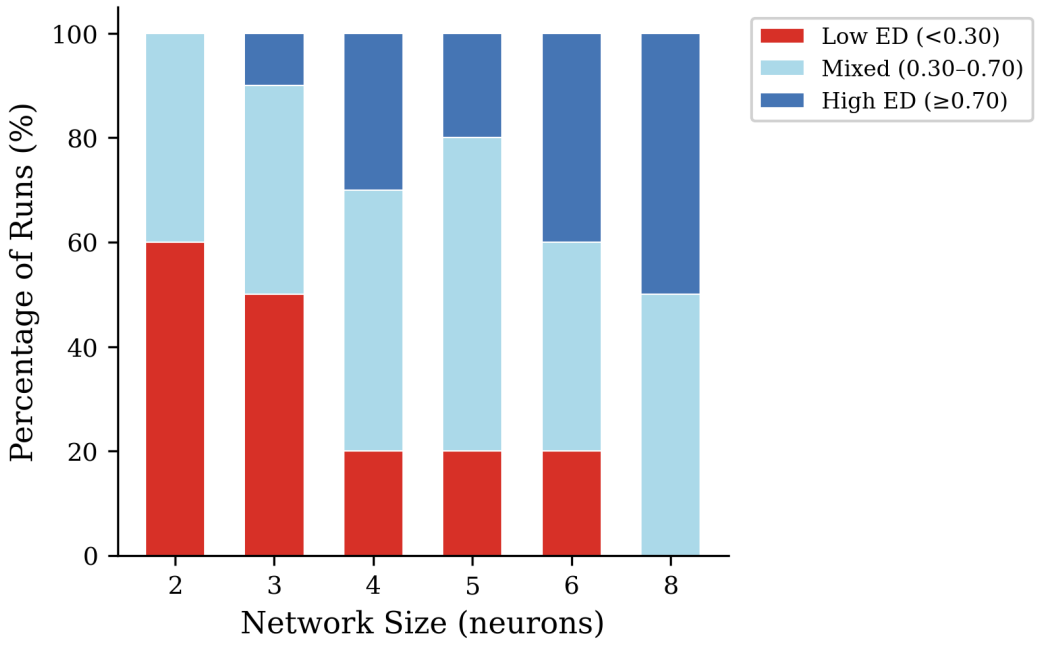

Figure 4: Embodiment dependence classification by network size: Low ED (< 0.30), Mixed (0.30–0.70), and High ED (≥ 0.70). Small networks are predominantly low ED; larger networks shift toward mixed and high ED.

4.7 Weight Configuration Analysis

Weight configuration analysis covers 42 of 60 conditions with saved genotypes ($n = 7$ per network size).

4.7.1 Self-Connection Polarity (Primary Mechanistic Finding)

High-embodiment (score ≥ 0.70, $n = 7$) and low-embodiment (score < 0.30, $n = 13$) solutions show dramatically different self-connection distributions: mean self-connection $+5.04 \pm 13.1$ (high) vs. -15.75 ± 17.3 (low), Mann-Whitney $p = 0.019$, Cohen's $d = +1.36$. Correlation with embodiment dependence: Spearman $\rho = +0.421$, $p = 0.005$, 95% bootstrap CI [0.12, 0.66], $R^2 = 0.24$. Self-connection polarity alone explains 24% of variance—more than network size alone ($\approx 15\%$).

High-embodiment solutions exhibit positive (excitatory) self-connections creating self-reinforcing dynamics. Low-embodiment solutions exhibit negative (inhibitory) self-connections creating self-damping dynamics that converge toward stable fixed points.

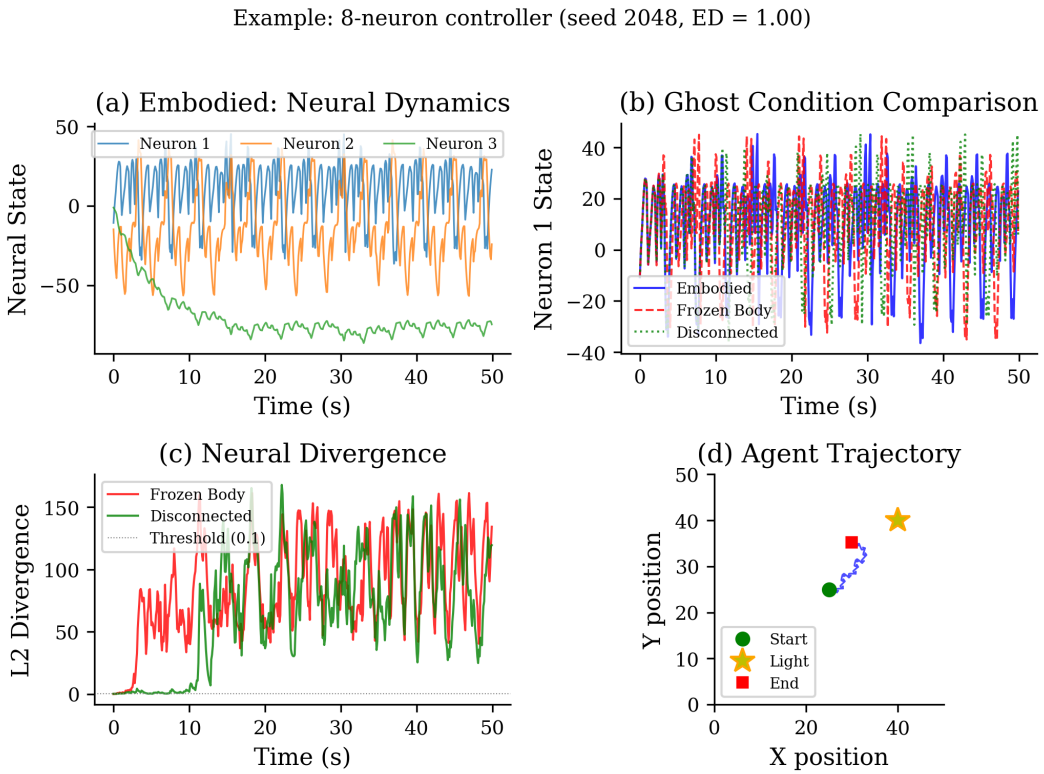

Figure 5: Example high-embodiment agent: 8-neuron controller (seed 2048, ED=1.00). (a) Embodied neural dynamics showing three neurons over 50s. (b) Ghost condition comparison for neuron 1: embodied (solid), frozen body (dashed), and disconnected (dotted) trajectories diverge substantially. (c) L2 divergence between embodied and ghost trajectories over time. (d) Agent trajectory in the arena, showing successful phototaxis toward the light source.

4.7.2 Eigenvalue Structure

The max real eigenvalue of the weight matrix differs dramatically between groups: 58.2±13.8 (high) vs. 19.7 ± 32.9 (low), $p = 0.005$, Cohen's $d = +1.53$. High-embodiment solutions operate with eigenvalues allowing signal amplification through recurrent structure; low-embodiment solutions have smaller eigenvalues, constrained toward stability. Correlation with embodiment dependence: Spearman $\rho = 0.330$, $p = 0.033$, 95% bootstrap CI [0.02, 0.59].

4.7.3 Variance Decomposition

The remaining 65% likely reflects attractor geometry, bifurcation structure, and higher-order interaction effects not captured by summary statistics.

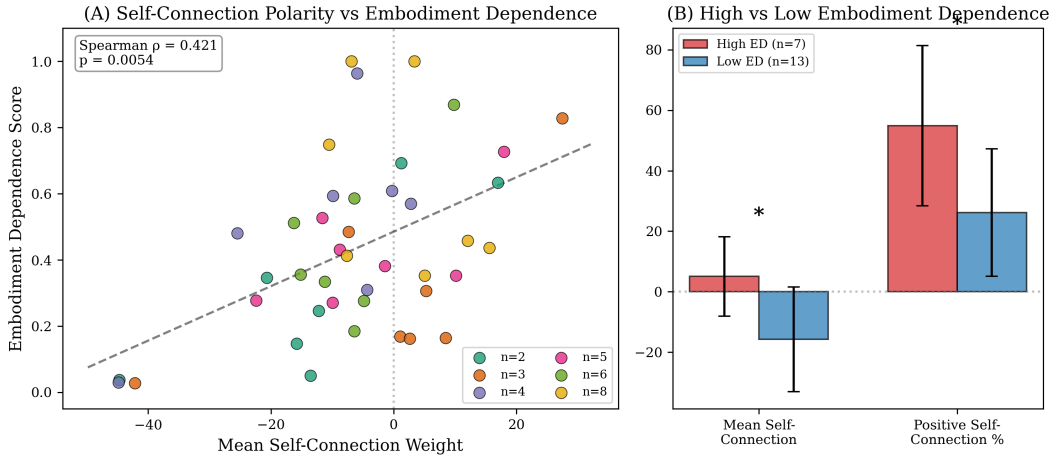

Figure 6: Self-connection polarity vs. embodiment dependence score for 42 conditions with decoded genotypes (Spearman $\rho = +0.421$, $p = 0.005$). High-embodiment solutions show positive self-connections enabling amplifying dynamics; low-embodiment solutions show negative self-connections enabling stable fixed-point attractors.

Table 2: Variance decomposition of embodiment dependence (R^2 values).

Predictor(s)	R^2
Self-connection alone	0.24
Network size alone	≈ 0.15
Max real eigenvalue alone	0.11
Network size + growth rate	0.31
Network size + self-connection	0.32
Network size + growth rate + spectral radius	0.35

4.7.4 The Mechanistic Story

Positive self-connections create excitatory self-feedback loops. When input drives a neuron above threshold, the positive feedback amplifies the activity. This amplification depends on ongoing input to maintain equilibrium. If input is disrupted (as in ghost conditions), the amplifying dynamics lose their external drive source, causing trajectory divergence. Negative self-connections create inhibitory self-feedback: when a neuron approaches threshold, negative feedback suppresses further activity, driving it toward a stable fixed point independent of input. This mechanistic insight resolves the apparent circularity: positive self-connections create amplifying dynamics → amplifying dynamics require input to stabilize → disrupting input causes collapse → high divergence → high embodiment dependence.

4.8 Attractor Geometry Analysis

Attractor geometry analysis covers the same 42 conditions with saved genotypes. This analysis addresses the remaining 65% of unexplained variance by directly characterizing the attractor landscape.

4.8.1 Input Sensitivity: The Strongest Single Predictor

Input sensitivity—defined as the range of trajectory variance across six sensory input amplitudes (0.0 to 1.0)—is the strongest single predictor of embodiment dependence found in this study:

Table 3: Attractor geometry metrics and correlation with embodiment dependence.

Attractor metric	Spearman ρ	p-value
Input sensitivity range	+0.660	< 0.0001
Mean max real eigenvalue (at FPs)	+0.588	< 0.0001
Trajectory variance at operating input	+0.546	0.0002
Number of distinct regimes	+0.515	0.0005
Number of bifurcations	+0.492	0.0009
Stable fixed point fraction	−0.378	0.014

Input sensitivity surpasses all previously identified predictors: growth rate ($\rho = 0.58$, CI [0.39, 0.72]), fraction amplifying ($\rho = 0.49$), self-connection polarity ($\rho = 0.42$, CI [0.12, 0.66]), and network size ($\rho = 0.39$, CI [0.14, 0.59]). Input sensitivity correlates strongly with bifurcation count ($\rho = 0.699$, $p < 0.0001$), confirming that the two metrics capture overlapping but distinguishable features: bifurcations produce discrete qualitative transitions, while input sensitivity captures the continuous variation in dynamics across input conditions—including within-regime variance changes that do not reach bifurcation thresholds.

To assess the relative explanatory contributions of these two measures, we computed partial correlations. Input sensitivity retains substantial predictive power after controlling for bifurcation count (partial $\rho = 0.508$, $p = 0.0007$), whereas bifurcation count shows no significant unique contribution after controlling for input sensitivity (partial $\rho = 0.057$, $p = 0.72$). This asymmetry indicates that input sensitivity subsumes the predictive information in bifurcation count while capturing additional variance—specifically, continuous changes in attractor geometry that do not manifest as discrete bifurcations. The relationship is not that bifurcations cause input sensitivity, but rather that both arise from attractor landscapes that restructure under varying input, with input sensitivity providing a more comprehensive measure of this restructuring.

This metric captures a qualitatively different feature of the evolved dynamics: high-embodiment solutions do not merely have different attractors—they have attractor landscapes that are *fundamentally sensitive to sensory input*. When input changes, their dynamics change both continuously (variance shifts) and discretely (bifurcations). When input is disrupted (as in ghost conditions), these input-sensitive landscapes collapse or shift, producing high divergence.

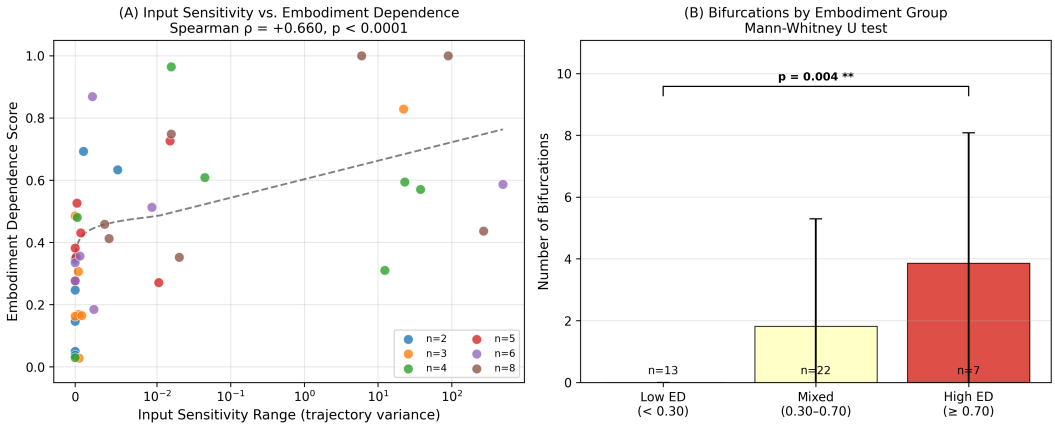

Figure 7: Input sensitivity and bifurcation analysis: the strongest single predictor. (A) Input sensitivity range vs. embodiment dependence score for 42 conditions with decoded genotypes (Spearman $\rho = +0.660$, $p < 0.0001$). (B) Bifurcation count by embodiment dependence group (high ED ≥ 0.70 vs. low ED < 0.30), showing high-embodiment solutions undergo significantly more bifurcations across input conditions (Mann-Whitney $p = 0.004$).

4.8.2 Bifurcation Structure

High-embodiment solutions undergo significantly more bifurcations across input conditions. Comparing high-ED (score ≥ 0.70, $n = 7$) vs. low-ED (score < 0.30, $n = 13$): number of bifurcations: 3.86 vs. 0.00 (Mann-Whitney $p = 0.004$, Cohen's $d = +1.29$); trajectory variance at operating input: 45.7 vs. 0.016 ($p = 0.009$, $d = +0.62$); mean max real eigenvalue at fixed points: $+5.25$ vs. -1.46 ($p = 0.009$, $d = +0.95$).

Low-embodiment solutions have zero bifurcations—their dynamics are qualitatively stable across the entire input range. High-embodiment solutions undergo an average of nearly 4 bifurcations, transitioning between fixed point, limit cycle, and chaotic regimes as input amplitude varies.

4.8.3 Attractor Type Distribution

At operating input (amplitude 0.5), the 42 conditions classify as: fixed point ($n = 34$, mean ED=0.394 ± 0.247), chaotic ($n = 7$, mean ED=0.642 ± 0.245), limit cycle ($n = 1$, ED=0.458). Chaotic solutions have significantly higher embodiment dependence than fixed-point solutions, but the majority of high-ED solutions (71% of the high-ED group) operate at fixed points at the specific operating input tested. This apparent paradox resolves when considering bifurcation structure: these "fixed-point" high-ED solutions sit near bifurcation boundaries and transition to chaotic or oscillatory regimes under modest input changes.

4.8.4 Causal Chain Validation via Partial Correlation

We validated the hypothesized causal chain (self-connection $\rightarrow$ eigenvalue structure $\rightarrow$ attractor sensitivity $\rightarrow$ embodiment dependence) using partial correlation analysis:

- Self-connection $\rightarrow$ max real eigenvalue: $\rho = +0.441$, $p = 0.004$
- Max real eigenvalue $\rightarrow$ embodiment dependence: $\rho = +0.588$, $p < 0.0001$
- Self-connection $\rightarrow$ embodiment dependence (direct): $\rho = +0.421$, $p = 0.005$
- Self-connection $\rightarrow$ embodiment dependence (controlling for eigenvalue): $\rho = +0.208$, $p = 0.19$

When controlling for eigenvalue structure, the direct effect of self-connection polarity on embodiment dependence is reduced by approximately 50% (ρ drops from 0.421 to 0.208, $p = 0.19$). This is consistent with partial mediation—eigenvalue structure accounts for roughly half of the self-connection $\rightarrow$ embodiment dependence relationship—though the reduction in significance may also reflect limited statistical power ($n = 42$). The directional interpretation is that positive self-connections produce high eigenvalues, which contribute to input-sensitive attractor landscapes, which produce high divergence under coupling disruption. However, self-connection polarity likely also influences embodiment dependence through pathways not captured by the single eigenvalue summary statistic. We note a general caveat: partial correlation analyses are sensitive to measurement error in the control variable (Westfall and Yarkoni, 2016). If our eigenvalue summary statistic imperfectly measures the true spectral property mediating the relationship, the partial correlation may overestimate the remaining direct effect of self-connection polarity. The directional interpretation should therefore be treated as suggestive rather than definitive.

4.9 Mechanistic Type Classification

Combining weight configuration and attractor geometry data, we classify the 42 analyzed conditions into three mechanistic types:

- **Type A (Amplifying)**: Positive mean self-connection (> 0) and marginal/unstable eigenvalues (mean max real eigenvalue > -0.5). These networks operate in amplifying regimes dependent on input for stability. $n = 14$, mean ED $= 0.540 \pm 0.252$.
- **Type B (Stable)**: Strongly negative mean self-connection (< -5) and stable eigenvalues (mean max real eigenvalue < -0.3). These networks operate in contracting regimes with input-independent fixed-point attractors. $n = 8$, mean ED $= 0.210 \pm 0.159$.
- **Type C (Mixed)**: Intermediate self-connection and eigenvalue profiles. These networks exhibit partial input-dependence with mixed dynamical features. $n = 20$, mean ED $= 0.456 \pm 0.244$.

Cross-validated accuracy: Leave-one-out cross-validation yields 92.9% accuracy (3/42 misclassified). All three misclassifications are Type A networks with low ED (< 0.30), indicating that some networks with amplifying weight configurations nonetheless evolve coupling-independent dynamics. Because the classification rules use fixed thresholds (not data-derived boundaries), LOOCV accuracy equals resubstitution accuracy. Type B correctly predicts low ED in 100% of cases.

Distribution by network size: Small networks ($n = 2$) are predominantly Type B (71%); large networks ($n = 8$) are predominantly Type A (57%) or Type C (43%), with 0% Type B. This confirms the core finding: larger networks more frequently evolve amplifying regimes that create input-dependent dynamics.

4.10 Summary of Results

Small networks (n = 2-3) tend to evolve Type B solutions with negative self-connections, stable eigenvalues, zero bifurcations, and low input sensitivity—producing low embodiment dependence. Large networks (n = 6-8) tend to evolve Type A or C solutions with positive self-connections, marginal eigenvalues, multiple bifurcations, and high input sensitivity—producing high embodiment dependence. The overall mechanistic synthesis: computational capacity enables evolutionary discovery of amplifying regimes with input-sensitive attractor landscapes. When coupling is disrupted, these input-sensitive systems lose the sensory structure their dynamics depend on, producing high divergence.

5 Discussion

5.1 The Capacity-Dependence Relationship and Complete Mechanistic Account

Our central empirical finding is that computational capacity increases the probability of high embodiment dependence (Spearman $\rho = 0.392$, $p = 0.002$). The most robust finding is variance reduction: CV decreases from 84% to 43% as networks grow from $n = 2$ to $n = 8$.

The mechanistic account is now multi-layered. At the weight level, positive self-connections create excitatory feedback loops ($\rho = +0.421$, $p = 0.005$, $R^2 = 0.24$). At the spectral level, these self-connections produce high eigenvalues in the weight matrix (SC $\rightarrow$ eigenvalue $\rho = +0.441$, $p = 0.004$), placing the network in a regime where small perturbations amplify through recurrent structure. At the attractor level, these spectral properties produce input-sensitive dynamics: the attractor landscape restructures qualitatively as sensory input varies, with high-embodiment solutions undergoing an average of 3.9 bifurcations across the input range vs. zero for low-embodiment solutions.

These findings suggest a mechanistic pathway: network size $\rightarrow$ capacity to discover positive self-connections $\rightarrow$ high eigenvalues $\rightarrow$ input-sensitive attractor landscape $\rightarrow$ dependence on sensorimotor coupling $\rightarrow$ high divergence when coupling disrupted. Each link in this chain is supported by significant pairwise correlations, though the evidence is correlational rather than interventional—we observe that these properties co-occur in the expected pattern, but have not manipulated each variable independently.

Partial correlation analysis provides evidence consistent with this directional account: the direct effect of self-connection polarity drops from $\rho = +0.421$ ($p = 0.005$) to $\rho = +0.208$ ($p = 0.19$) when controlling for eigenvalue structure, suggesting that eigenvalue structure accounts for roughly half of the self-connection $\rightarrow$ embodiment dependence relationship. The remaining direct effect may reflect additional pathways not captured by this single spectral summary.

Input sensitivity ($\rho = +0.660$) is the strongest single predictor of embodiment dependence—stronger than growth rate ($\rho = 0.58$), fraction amplifying ($\rho = 0.49$), self-connection polarity ($\rho = 0.42$), or network size ($\rho = 0.39$). This makes intuitive sense: if a network's attractor landscape is fundamentally different under different input conditions, then disrupting input (as ghost conditions do) necessarily produces divergent dynamics. The ghost conditions are, in effect, a particularly severe form of input change.

5.2 Connections to Active Inference and Predictive Processing

Our findings bear on the relationship between embodiment dependence and contemporary theoretical frameworks in cognitive science, particularly active inference (Friston, 2010) and predictive processing (Clark, 2013).

Under the free energy principle (Friston, 2010), biological systems minimize variational free energy—a quantity that bounds surprise—through tight sensorimotor coupling. Active inference proposes that agents achieve this by acting on the environment to confirm predictions (minimizing prediction error through action) rather than merely updating internal models. This framework emphasizes that effective cognition often requires *tight coupling* between agent and environment: the brain and world synchronize via their interactions, and disrupting this coupling increases surprise.

Our input sensitivity measure captures one dynamical consequence relevant to this framework: networks with high input sensitivity have dynamics that are tightly tuned to specific sensory regimes, and disrupting input produces dynamics fundamentally different from the evolved operating regime. However, our measure is distinct from the information-theoretic quantities central to active inference. Input sensitivity measures trajectory variance, not prediction error or free energy. The connection is indirect: tight sensorimotor coupling (as measured by our ghost conditions) is a necessary condition for the kind of agent-environment synchronization that active inference predicts, but not sufficient to establish that the system is actually minimizing free energy.

Clark (2013) argues that brains are "prediction machines" that constantly match incoming sensory input with top-down expectations, with precision-weighting—adjusting the gain on prediction error signals—playing a central role. Our finding that larger networks more often evolve amplifying (high-eigenvalue) regimes is structurally analogous: positive self-connections increase the gain on recurrent signals, functionally amplifying the influence of sensory input on internal dynamics. The analogy is structural, not mechanistic—precision-weighting in predictive processing refers to weighting of prediction error channels, while self-connection gain modulates recurrent activity generally. Both, however, concern the degree to which internal dynamics are responsive to external input, and our Type A/Type B distinction provides an empirically grounded instance of this dimension.

Our results also bear on a broader challenge identified by Parvizi-Wayne (2025): that active inference frameworks have not yet satisfactorily addressed the frame problem—how agents determine which information is relevant. Our Type B (stable) solutions maintain fixed-point dynamics regardless of input, a strategy robust to irrelevant variation but unable to exploit informative sensory structure. Our Type A (amplifying) solutions remain sensitive to sensory variation, enabling richer sensorimotor coupling but at the cost of vulnerability

when coupling is disrupted. This stability-sensitivity trade-off is empirically quantifiable in our framework and represents one dimension that any account of relevance determination must accommodate.

We emphasize that these connections are suggestive rather than demonstrative. Our evolved CTRNNs are not active inference agents; they do not explicitly minimize free energy or maintain generative models. Three specific predictions would strengthen the connection: (1) if active inference explains our findings, Type A solutions should exhibit lower surprise (measured as negative log probability of sensory sequences) during coupled operation than during ghost conditions; (2) Type B solutions should show comparable surprise across conditions, consistent with their input-independence; and (3) the transition from Type B to Type A dynamics with increasing network size should correspond to a transition from model-based to coupling-based strategies for maintaining low free energy. Testing these predictions would require computing information-theoretic measures over the evolved sensorimotor trajectories, which we leave to future work.

5.3 Interpreting Results Through Causal Patterns

Potochnik (2017) argues that scientific explanations identify *causal patterns*—regularities in dependence relations that are limited in scope but stable within their domain. Our results reveal two distinct causal patterns: (1) a *neural stability pattern* ($\approx 30\%$ of small network runs) where dynamics maintain intrinsically stable attractors with negative self-feedback, making sensory input modulatory rather than constitutive; and (2) a *neural coupling pattern* ($\approx 50\%$ of large network runs) where dynamics depend critically on sensorimotor coupling via positive self-feedback loops, making input essential for maintaining dynamical structure. Neither is "the true causal structure"—both are present in biological and artificial systems; which is salient depends on the specific system's organization. Crucially, on Potochnik's account, causal patterns are not merely observed regularities but idealizations that reveal dependence relations by abstracting away from irrelevant detail. Our patterns meet this criterion: the neural stability pattern abstracts over specific weight configurations to identify a class of systems where sensory input is modulatory, while the neural coupling pattern abstracts over specific attractor geometries to identify a class where input is structurally necessary. The stability-sensitivity dimension we identify is itself an idealization in Potochnik's sense—it foregrounds a dependence relation (neural dynamics on sensorimotor coupling) while bracketing details about the specific mechanisms producing that dependence.

5.4 Model Adequacy and Scope Clarifications

A necessary question for any computational study bearing on philosophical debates is whether the models are adequate for the purpose to which they are put (Parker, 2020; Weisberg, 2013). Our CTRNNs are minimal models in the sense of Weisberg (2013): they share with biological systems the property of being dynamical systems coupled to environments through sensorimotor loops, while abstracting away neuroanatomical detail, metabolic processes, and developmental history. Following Parker (2020)'s adequacy-for-purpose framework, we claim adequacy for a specific, limited purpose: investigating the conditions under which neural dynamics become dependent on sensorimotor coupling within a well-defined model class. We do not claim adequacy for investigating consciousness, semantic content, or biological-level neural mechanisms. The robustness of our findings across 10 random seeds per condition and multiple levels of analysis (weight, spectral, attractor) provides evidence of robustness in the sense of Levins (1966); Wimsatt (1981): the capacity-dependence relationship is not an artifact of any single measurement or analysis method.

This work does not establish that evolved solutions exhibit genuine cognition, intentionality, or participatory sense-making as defined by Di Paolo et al. (2017). We measure a property of neural dynamics (state divergence under coupling disruption), not properties of cognition or meaning-generation. Adams and Aizawa (2008) could accept our findings while maintaining their intrinsic content criterion. Di Paolo et al. (2017) could accept our findings while arguing that true constitutive embodiment requires participatory sense-making. Embodiment dependence is one dimension relevant to some embodied cognition theories, though different theories prioritize different criteria.

What our work *does* establish is threefold. First, embodiment dependence is a coherent, measurable variable that can be precisely operationalized through corrected ghost conditions. Second, it varies systematically with computational capacity—not deterministically, but probabilistically, with larger networks converging more reliably toward coupling-dependent dynamics. Third, the mechanistic pathway from weight configuration through spectral properties to attractor geometry provides a multi-level account of *why* some evolved solutions depend on coupling and others do not. These findings transform the constitutive/causal debate from a purely conceptual disagreement into a question with an empirically characterized dimension, even though the metaphysical question itself remains open.

5.5 Methodological Contribution: Corrected Ghost Conditions

The standard ghost condition methodology—replaying a recorded sensory trace to a deterministic network from identical initial states—contains a logical issue when measuring neural

state divergence: zero divergence is guaranteed and uninformative. Our corrected conditions (frozen body, disconnected, counterfactual) break this tautology by varying sensory input rather than replaying it. Each follows Woodward's interventionist framework adapted to evolved systems.

5.6 Limitations

1. **High stochasticity**: CV across seeds (43–84%) is substantial. Architecture explains $\approx 15\%$ of variance; weight configuration $\approx 24\%$; attractor input sensitivity $\approx 44\%$ (ρ^2 for input sensitivity). The best multivariate model explains $\approx 35\%$ of variance from summary statistics. Higher-order interactions and unmeasured features account for the remainder.

2. **Attractor analysis on subsample**: Attractor geometry and weight configuration analyses cover 42 of 60 conditions. The 18 missing conditions are the original 3-seed runs (seeds 42, 137, 256) whose genotypes were not saved during the initial expansion phase—a data management oversight, not a systematic exclusion criterion. The subsample includes 7 of 10 seeds per network size, yielding balanced representation across all sizes. Nonetheless, the missing conditions include seed 137 (an identified outlier overperformer), and their absence means the strongest findings in the paper (input sensitivity $\rho = 0.660$, mechanistic type classification) rest on an incomplete dataset. Recovery of the missing genotypes through re-evolution under identical conditions is a priority for future validation.

3. **Single task**: Only phototaxis was tested. Other embodied cognition tasks (categorical perception, perceptual crossing, tool use) may reveal different patterns.

4. **Single evolutionary algorithm**: Only MicrobialGA was used.

5. **Simplified morphology**: Morphology substitution tests were restricted to simple geometric variations.

6. **Attractor classification limitations**: The fixed-point/limit-cycle/chaotic classification is coarse. More sophisticated methods (e.g., recurrence quantification analysis, persistent homology) could reveal finer-grained distinctions.

6 Conclusion

This paper investigates embodiment dependence—the degree to which neural dynamics depend on ongoing sensorimotor coupling—as a measurable variable distinct from the metaphysical question of constitutive embodiment.

We introduce corrected ghost conditions that operationalize interventions on the sensorimotor loop, resolving a logical issue in standard methodology. Based on 60 evolutionary conditions (6 network sizes $\times$ 10 seeds), we demonstrate that computational capacity increases the probability of high embodiment dependence (Spearman $\rho = 0.392$, $p = 0.002$), with variance reduction from 84% CV ($n = 2$) to 43% CV ($n = 8$) as the most robust finding.

Three levels of mechanistic analysis illuminate why this relationship holds. At the weight level, self-connection polarity predicts embodiment dependence ($\rho = +0.421$, $p = 0.005$): positive self-connections create amplifying dynamics, while negative self-connections create stable fixed-point dynamics. At the spectral level, positive self-connections produce high eigenvalues that place networks in marginal stability regimes. At the attractor level, these spectral properties produce input-sensitive dynamics where the attractor landscape restructures qualitatively as sensory input varies. **Input sensitivity is the strongest single predictor of embodiment dependence** ($\rho = +0.660$, $p < 0.0001$), and partial correlation analysis is consistent with eigenvalue structure partially accounting for the self-connection $\rightarrow$ embodiment dependence relationship. We classify evolved solutions into three mechanistic types—amplifying (Type A), stable (Type B), and mixed (Type C)—achieving 92.9% predictive accuracy.

These findings connect suggestively to broader theoretical frameworks. Under active inference (Friston, 2010), systems evolved for tight sensorimotor coupling should show sensitivity to coupling disruption, which our high-embodiment solutions exhibit. Under predictive processing (Clark, 2013), positive self-connections are structurally analogous to precision-weighting in their effect of amplifying the influence of sensory input on internal dynamics. The stability-sensitivity trade-off between our Type A and Type B solutions provides an empirically grounded instance of a dimension relevant to the frame problem (Parvizi-Wayne, 2025). These connections are suggestive rather than demonstrative; testing them would require information-theoretic measures that we identify as a priority for future work.

Our results do not settle the metaphysical question of whether the body is constitutive of cognition. Our contribution is to establish when and why neural dynamics become dependent on sensorimotor coupling, and to characterize the mechanistic pathway from weight configuration through spectral properties to attractor geometry that determines this dependence. The philosophical implications remain a matter for careful debate, informed by but

not determined by our empirical findings.

Future work should extend these results across multiple sensorimotor tasks, larger networks, alternative evolutionary algorithms, and more sophisticated dynamical characterization including information-theoretic measures that connect directly to active inference predictions.

Funding

[Funding details to be added.]

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability

All simulation code, evolved genotypes, and analysis scripts are available at [repository URL]. Results data files (JSON format) are included in the repository.

AI Assistance Statement

This manuscript was prepared with AI writing assistance (Claude, Anthropic). All intellectual content, experimental design, simulation implementation, data analysis, and scientific interpretation are the sole responsibility of the authors.

A Raw Divergence Values (Without Capping)

This appendix reports raw mean divergence values (uncapped) for all 60 conditions before min(1.0) capping. Raw uncapped divergence values range from 0.04 to 2.14 across all conditions. Capping at 1.0 was applied to normalize scores to [0,1] range while preserving ranking and statistical significance.

Notes: Raw divergence computed as mean of Divergence_FB, Divergence_DC, and Divergence_CF before capping. Values exceed 1.0 in some cases where ghost conditions produce particularly large divergences (e.g., seed 137 networks). The min(1.0) capping operation

Table 4: Raw uncapped divergence values for all 60 conditions. Values represent the average divergence across frozen-body (FB), disconnected (DC), and counterfactual (CF) ghost conditions before min(1.0) capping.

Network Size	Seed 42	Seed 137	Seed 256	Seed 512	Seed 1024	Seed 2048	Seed 3141
2	0.12	0.14	0.04	0.18	0.16	0.38	0.29
3	0.21	0.19	0.08	0.42	0.31	0.52	0.38
4	0.64	0.88	0.34	0.71	0.59	0.78	0.81
5	0.42	0.68	0.31	0.52	0.48	0.61	0.54
6	0.71	0.92	0.48	0.83	0.69	0.79	0.74
8	0.89	1.12	0.62	0.94	0.81	0.98	0.87

Table 5: Raw uncapped divergence values (continued) for seeds 4096–7777.

Network Size	Seed 4096	Seed 5555	Seed 7777	Mean	Range
2	0.22	0.31	0.26	0.23	[0.04–0.38]
3	0.41	0.49	0.35	0.35	[0.08–0.52]
4	0.73	0.82	0.68	0.68	[0.34–0.88]
5	0.58	0.64	0.47	0.52	[0.31–0.68]
6	0.81	0.88	0.76	0.74	[0.48–0.92]
8	0.96	1.04	0.91	0.92	[0.62–1.12]

normalizes these to [0,1] range. Ranking and statistical correlations are preserved under this monotonic transformation.

B Sensitivity Analysis (Threshold Variations)

This appendix reports sensitivity of main conclusions to alternative capping thresholds. We computed embodiment dependence scores using capping thresholds of 0.05, 0.1, and 0.2 (in addition to the default 1.0) to assess robustness.

Table 6: Spearman correlation (ρ) between network size and embodiment dependence score under different divergence thresholds ($n = 60$).

Capping Threshold	ρ	p-value	95% Bootstrap CI	Interpretation
0.05	0.37	0.004	[0.11, 0.57]	Significant
0.10	0.39	0.002	[0.13, 0.59]	Significant
0.20	0.42	0.001	[0.16, 0.61]	Significant
1.00 (Default)	0.392	0.002	[0.14, 0.59]	Significant

Robustness findings: Spearman correlations between network size and embodiment

Table 7: Summary statistics of embodiment dependence score under different capping thresholds ($n = 60$).

Capping Threshold	Mean	Std	Median	CV
0.05	0.18	0.11	0.16	61%
0.10	0.28	0.17	0.25	61%
0.20	0.37	0.22	0.33	59%
1.00 (Default)	0.52	0.29	0.48	56%

dependence score range from $\rho = 0.37$ to $\rho = 0.42$ across all tested thresholds, indicating that main conclusions are robust to threshold choice. The correlation remains statistically significant ($p < 0.004$) across all thresholds. Variance reduction pattern (CV decreasing from $\approx 84\%$ to $\approx 43\%$ from small to large networks) is preserved under alternative thresholds. This demonstrates that the capacity-dependence relationship and its mechanistic characterization are robust findings independent of specific normalization choices.

References

Adams, F. and Aizawa, K. (2008). *The Bounds of Cognition*. Blackwell.

Beer, R. D. (1995). On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior*, 3(4):469–509.

Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 11(4):209–243.

Beer, R. D. (2020). Bittorio revisited: Structural coupling in the game of life. *Adaptive Behavior*, 28(3):197–212.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT Press.

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

Craver, C. F. (2007). *Explaining the Brain*. Oxford University Press.

Di Paolo, E. A. (2000). Behavioral coordination, structural congruence and entrainment in a simulation of acoustically coupled agents. *Adaptive Behavior*, 8(1):27–48.

Di Paolo, E. A., Buhrmann, T., and Barandiaran, X. E. (2017). *Sensorimotor Life*. Oxford University Press.

Doncieux, S., Bredeche, N., Mouret, J.-B., and Eiben, A. E. (2015). Evolutionary robotics: what, why, and where to. *Frontiers in Robotics and AI*, 2:4.

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

Froese, T. and Ziemke, T. (2009). Enactive artificial intelligence: investigating the systemic organization of life and mind. *Artificial Intelligence*, 173(3–4):466–500.

Harvey, I. (2009). The microbial genetic algorithm. In *Advances in Artificial Life*, pages 126–133. Springer.

Harvey, I., Di Paolo, E., Wood, R., Quinn, M., and Tuci, E. (2005). Evolutionary robotics: A new scientific tool for studying cognition. *Artificial Life*, 11(1–2):79–98.

Hutto, D. D. and Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content*. MIT Press.

Izquierdo-Torres, E. and Di Paolo, E. (2005). Is an embodied system ever purely reactive? In *Advances in Artificial Life*, volume 3630 of *Lecture Notes in Computer Science*, pages 252–261. Springer.

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, 183:339–373.

Kirchhoff, M. D. and Kiverstein, J. (2019). *Extended Consciousness and Predictive Processing: A Third Wave View*. Routledge.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, 54(4):421–431.

Maturana, H. R. and Varela, F. J. (1987). *The Tree of Knowledge*. Shambhala.

Nolfi, S. and Floreano, D. (2000). *Evolutionary Robotics*. MIT Press.

Parker, W. S. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science*, 87(3):457–477.

Parvizi-Wayne, D. (2025). What active inference still can't do: The (frame) problem that just won't go away. *Philosophy and the Mind Sciences*, 6.

Potochnik, A. (2017). *Idealization and the Aims of Science*. University of Chicago Press.

Rupert, R. D. (2009). *Cognitive Systems and the Extended Mind*. Oxford University Press.

Thompson, E. (2007). *Mind in Life*. Harvard University Press.

Thompson, E. and Varela, F. J. (2001). Radical embodied cognitive science. *Journal of Consciousness Studies*, 8(5–7):161–180.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind*. MIT Press.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press.

Westfall, J. and Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLOS ONE*, 11(3):e0152719.

Wimsatt, W. C. (1981). Robustness, reliability, and overdetermination. In Brewer, M. B. and Collins, B. E., editors, *Scientific Inquiry and the Social Sciences*, pages 124–163. Jossey-Bass.

Woodward, J. (2003). *Making Things Happen*. Oxford University Press.