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 $\rho = 0.394$, $p = 0.002$), with variance reduction from 85% to 37% CV as the most robust finding. Self-connection polarity is the strongest weight-level predictor ($\rho = +0.369$, $p = 0.004$), with high-embodiment solutions exhibiting positive self-connections while low-embodiment solutions show negative self-connections (Cohen's $d = +1.29$). Most importantly, *input sensitivity*—how much the attractor landscape restructures across sensory conditions—emerges as the strongest attractor-geometry predictor across all 60 conditions ($\rho = +0.555$, $p < 0.0001$), linked to a pathway from positive self-connections through high eigenvalues to input-dependent attractor geometry. Partial correlation analysis suggests eigenvalue structure partially accounts for the self-connection effect (reduction of approximately 37%). We classify evolved solutions into three mechanistic types achieving 71.7% leave-one-out cross-validation accuracy. All key findings survive Benjamini-Hochberg FDR correction (23/26 tests at $q < 0.05$).

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 16% of variance; the remaining 84% reflects evolutionary trajectory stochasticity. Among dynamical measures available for all 60 conditions, growth rate (perturbation sensitivity) is the strongest predictor of embodiment dependence (Spearman $\rho = 0.615$, $p < 0.0001$); input sensitivity, available for all 60 conditions, is the strongest attractor-geometry predictor (see Contribution 4). Self-connection polarity predicts embodiment dependence comparably to network size alone (self-connection $\rho = +0.369$ vs. network size $\rho = +0.394$).

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 two-thirds of the direct effect remains when controlling for eigenvalues. We classify all 60 evolved solutions into three mechanistic types achieving 71.7% leave-one-out cross-validation accuracy.

4. **Attractor geometry**: We characterize the attractor landscape of all 60 evolved solutions and discover that *input sensitivity*—how much the attractor structure changes across sensory conditions—is the strongest attractor-geometry predictor of embodiment dependence ($\rho = +0.555$, $p < 0.0001$), second only to growth rate ($\rho = 0.615$)

and stronger than self-connection polarity ($\rho = 0.37$). 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. Recent work by Kiverstein and Kirchhoff (2023) argues that the standard causal-constitution distinction rests on a fallacy: they propose *diachronic constitution*, in which

parts and wholes are related across time rather than at a single instant, dissolving the apparent incompatibility between causal and constitutive relations. If this argument succeeds, the binary framing of our debate is misconceived—a system's body might be both causally and constitutively involved in cognition, depending on the timescale of analysis. Krickel (2023) objects that diachronic constitution collapses into a form of causation, but the debate remains open. Our empirical approach is compatible with either resolution: the question *under what conditions do neural dynamics become dependent on ongoing sensorimotor coupling?* is scientifically productive regardless of whether such dependence is ultimately classified as causal, constitutive, or—as Kiverstein and Kirchhoff suggest—both. Embodiment dependence measures the degree of this coupling dependence, and its variation across network architectures is an empirical finding independent of the metaphysical taxonomy.

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 despite 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) and from precariousness as a foundational concept in enaction theory (Beer and Di Paolo, 2023). 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 (in the sense of Wimsatt, 1981): 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 two empirical findings: (1) the Spearman correlation between network size and embodiment dependence is significant when computed using any single ghost condition alone (frozen body $\rho = 0.391$, $p = 0.002$; disconnected $\rho = 0.366$, $p = 0.004$; counterfactual $\rho = 0.348$, $p = 0.006$), indicating that no single condition drives the result; and (2) the disconnected and counterfactual conditions produce highly correlated divergence values ($r = 0.990$, $p < 0.0001$), while the frozen body condition is less correlated with the others ($r = 0.241$ and $r = 0.165$), suggesting it captures a partially distinct aspect of coupling. Averaging across all three provides a more comprehensive measure of coupling dependence than any single condition.

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, following the "minimally cognitive behavior" research program (Beer, 2021) and the dynamical systems approach to cognitive science (Beer, 2023). 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). The use of small evolved networks as "toy models" follows Beer (2024)'s argument that simple models can illuminate principles that are obscured in more complex systems. 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:

$$\text{Score} = \frac{\min(1.0,\ D_{\text{FB}}) + \min(1.0,\ D_{\text{DC}}) + \min(1.0,\ D_{\text{CF}})}{3} \tag{2}$$

where each ghost condition's divergence is individually capped at 1.0 before averaging, ensuring each condition contributes on a bounded $[0, 1]$ scale. Equal weighting across the three conditions ensures no single ghost condition dominates the composite measure.

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).

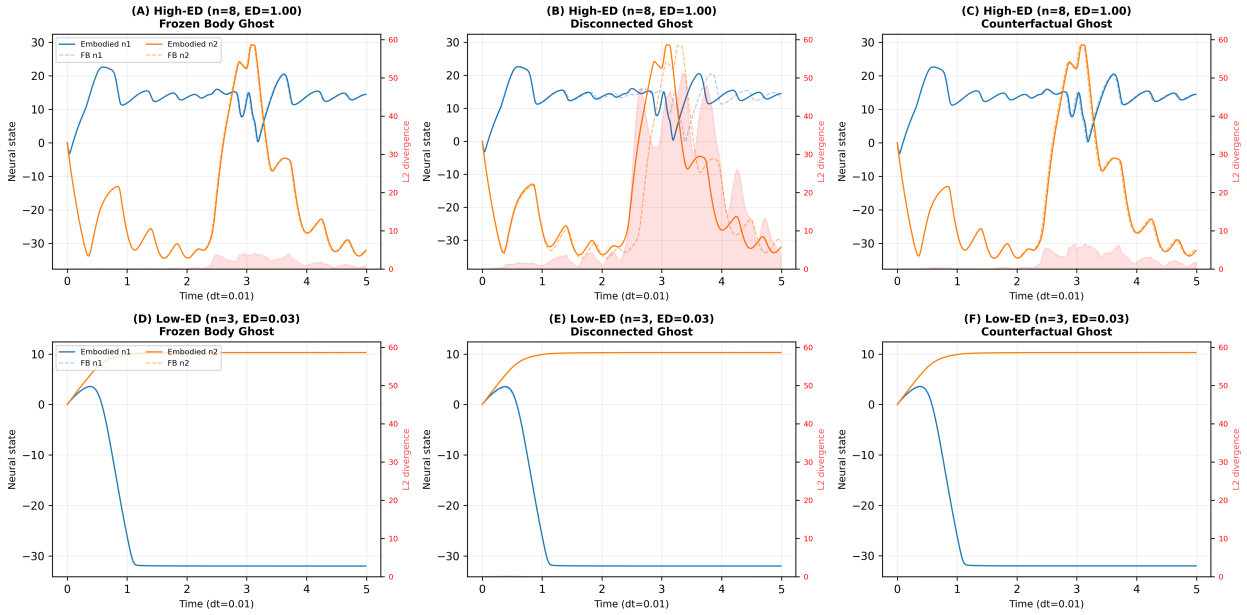

Figure 1: Example neural state trajectories under embodied and ghost conditions. Top row: a high-embodiment-dependence agent (n=8, ED=1.00) shows large divergence between embodied (solid) and ghost (dashed) trajectories, with red shading indicating L2 divergence magnitude. Bottom row: a low-embodiment-dependence 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 configura-

tions 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. Theoretical results support this proxy: RNN computational expressivity grows with neuron count (Siegelmann and Sontag, 1995) established Turing-completeness for sufficiently large networks, and for continuous-time systems, the range of realizable input-output mappings expands with state-space dimension.

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. Lyapunov exponents were estimated using the perturbation-based method (Benettin et al., 1980): a small perturbation (magnitude 10^{-6}) was applied to the neural state, evolved forward, and renormalized at each step, with the log divergence rate accumulated over 10,000 timesteps (1,000 transient steps discarded). For the binary chaotic/non-chaotic classification used here, the perturbation-based approach is adequate, though it may diverge from variational equation methods for small networks ($n = 2$–3) near marginal stability.

3.7 Weight Configuration Analysis

For all 60 conditions, 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 all 60 conditions, 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 to detect bifurcations, defined as qualitative changes in attractor type. A bifurcation was recorded when the classified type changed between adjacent amplitude steps; 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.

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 ($p = 0.047$, $q = 0.051$—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.

Statistical framework. All statistical tests are two-tailed at $\alpha = 0.05$ unless otherwise specified. We distinguish between *confirmatory* analyses—those testing pre-specified hypotheses—and *exploratory* analyses discovered during investigation. The confirmatory hypothesis is that computational capacity (network size) predicts embodiment dependence; this is the study's motivating question and was specified before the expanded dataset was collected. The dynamical characterization (Section 4.6) and the capacity-dependence correlation test this hypothesis. All other analyses—weight configuration (Section 4.7), attractor geometry (Section 4.8), mechanistic type classification (Section 4.9)—are exploratory: they were developed to explain variance in the primary finding. Following Westfall and Yarkoni (2016), we caution that exploratory findings, even when statistically significant after correction, require independent replication before being treated as established.

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	0.353 ± 0.301	0.285	[0.04, 0.79]	85%
3	0.396 ± 0.273	0.313	[0.07, 0.85]	69%
4	0.627 ± 0.325	0.676	[0.09, 1.00]	52%
5	0.549 ± 0.211	0.509	[0.33, 0.96]	38%
6	0.651 ± 0.312	0.645	[0.24, 1.00]	48%
8	0.697 ± 0.256	0.646	[0.37, 1.00]	37%

Statistical analysis of capacity-dependence relationship: Spearman rank correlation ($n = 60$): $\rho = 0.394$, $p = 0.002$ (two-tailed), 95% Bootstrap CI [0.14, 0.60]. Pearson r ($n = 60$): 0.385, $p = 0.002$. Kruskal-Wallis H-test: $H = 11.25$, $p = 0.047$ (nominally significant but marginally non-significant after FDR correction, $q = 0.051$). Post-hoc Mann-Whitney small (n = 2-3) vs. large (n = 6-8): $U = 83.0$, $p = 0.002$, Cohen's $d = 1.07$ (large effect).

Variance reduction (most robust finding): The coefficient of variation broadly decreases from 85% ($n = 2$) to 37% ($n = 8$), with a minor departure at $n = 6$ (CV=48%, up from 38% at $n = 5$). Despite this non-monotonicity, the overall trend is clear: larger networks converge more reliably to high-dependence solutions across different evolutionary trajectories.

Seed effects are substantial: Across the 10 seeds, individual seed means range from 0.379 to 0.900. Seed 137 is an outlier overperformer (mean 0.900 across all sizes). Analysis of variance on seed effects yields $F(9, 50) = 1.96$, $p = 0.064$—a marginal trend suggesting systematic seed differences, though not reaching significance at $\alpha = 0.05$.

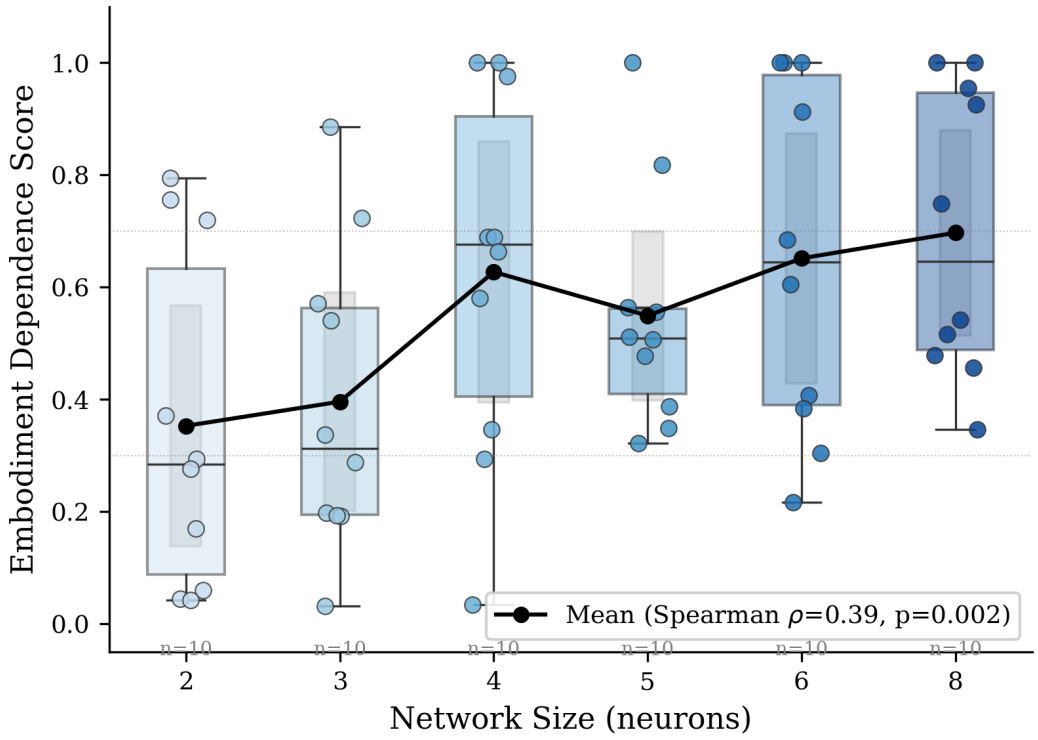

Figure 2: Embodiment dependence score across network sizes (60 conditions). Left: scatter plot with jitter showing individual conditions colored by network size. Right: box plots summarizing distributions per network size, showing increasing median and decreasing variance with size.

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 $= 0.680 \pm 0.314$ vs. 10-seed mean $= 0.546 \pm 0.297$. The 3-seed sample overestimated the population mean by approximately 0.134 absolute (25% relative). The correlation with network size was inflated: $\rho = 0.473$ (3-seed) vs. $\rho = 0.394$ (10-seed), a 20% 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. As a further robustness check, excluding Seed 137 entirely ($n = 54$) yields $\rho = 0.433$, $p = 0.001$—if anything, the effect strengthens without the outlier, confirming that the capacity-dependence relationship is not driven by this seed.

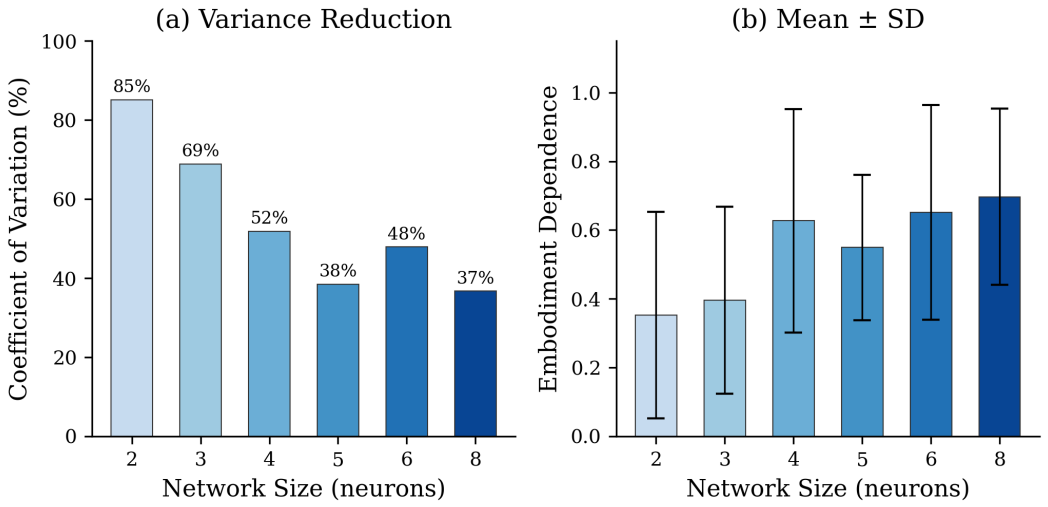

Figure 3: Coefficient of variation (CV) of embodiment dependence score by network size. CV generally decreases from 85% (n=2) to 37% (n=8), demonstrating that larger networks converge more reliably to embodiment-dependent solutions.

4.4 Non-Monotonic Pattern

The progression ($n = 2$: $0.353 \rightarrow n = 3$: $0.396 \rightarrow n = 4$: $0.627 \rightarrow n = 5$: $0.549 \rightarrow n = 6$: $0.651 \rightarrow n = 8$: 0.697) 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 is probabilistic, not deterministic. The Spearman correlation ($\rho = 0.394$) remains significant because the overall trend (small < large) is statistically robust.

4.5 Classification Distributions

We classify each condition by embodiment dependence profile: CAUSAL_DOMINANT (score < 0.30), MIXED (score 0.30–0.70), and EMBODIMENT_DOMINANT (score ≥ 0.70). These thresholds were pre-specified based on natural clustering in preliminary analyses. Small networks ($n = 2$–3): 55% causal-dominant, 20% mixed, 25% embodiment-dominant. Medium networks ($n = 4$–5): 10% causal-dominant, 65% mixed, 25% embodiment-dominant. Large networks ($n = 6$–8): 5% causal-dominant, 50% mixed, 45% embodiment-dominant.

4.6 Dynamical Characterization: All 60 Conditions

Dynamical analysis covers all 60 conditions. This analysis is confirmatory: we pre-specified the hypothesis that dynamical regime predicts embodiment dependence. Correlations with

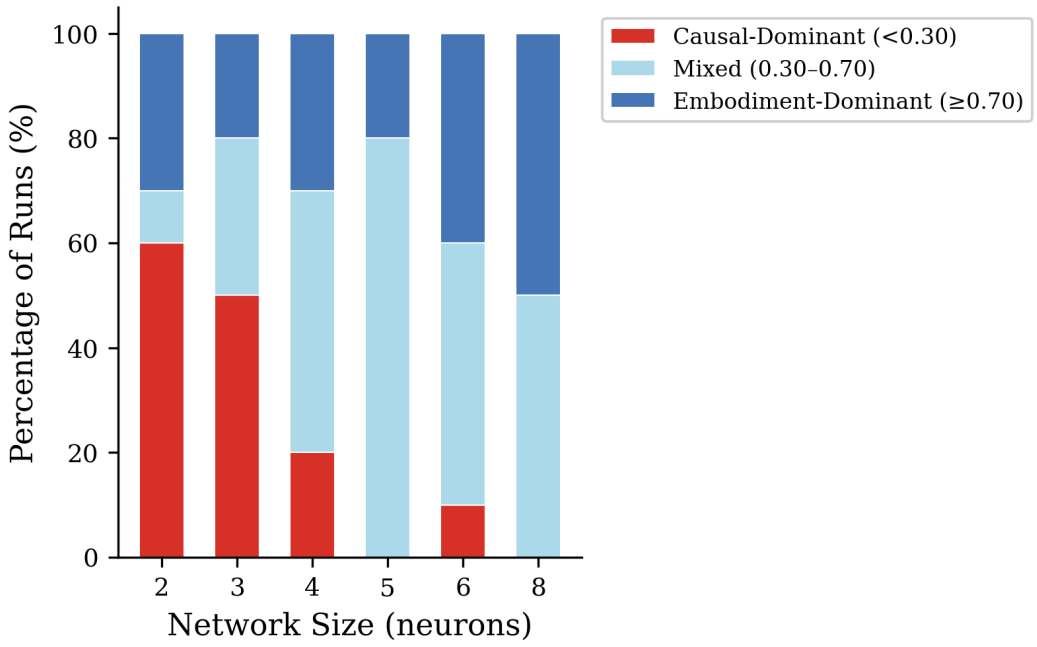

Figure 4: Embodiment dependence classification by network size: CAUSAL_DOMINANT (< 0.30), MIXED (0.30–0.70), EMBODIMENT_DOMINANT (≥ 0.70). Small networks are predominantly causal-dominant; larger networks shift toward mixed and embodiment-dominant.

embodiment dependence score ($n = 60$): growth rate Spearman $\rho = 0.615$, $p < 0.0001$ [survives FDR] (strongest dynamical predictor); participation ratio $\rho = 0.311$, $p = 0.016$ [survives FDR]; fraction amplifying $\rho = 0.525$, $p < 0.0001$ [survives FDR]; max Lyapunov $\rho = 0.315$, $p = 0.014$ [survives FDR].

Growth rate (perturbation sensitivity) is a stronger predictor of embodiment dependence ($\rho = 0.615$) than network size itself ($\rho = 0.394$), demonstrating that dynamical regime matters more than architecture alone. However, growth rate explains only $\approx 38\%$ of variance (ρ^2). All four dynamical measures correlate significantly, indicating that embodiment dependence arises from multiple dynamical properties acting in concert.

4.7 Weight Configuration Analysis

Weight configuration analysis covers all 60 conditions ($n = 10$ per network size). These analyses are exploratory: they were developed to explain variance in the primary capacity-dependence finding, not pre-specified as hypotheses.

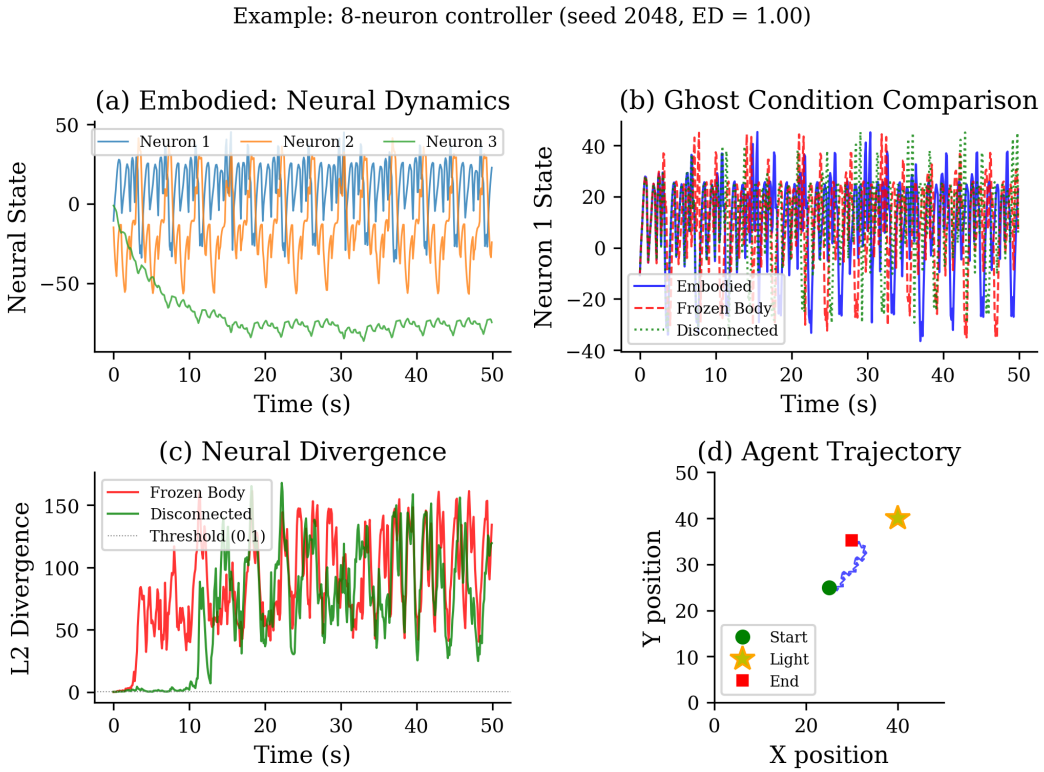

Figure 5: Example high-embodiment agent: 8-neuron controller (seed 2048, ED=1.00). (a) Embodied neural dynamics showing three neurons over 50s. (b) Ghost condition comparison for neuron 1: embodied (solid), frozen body (dashed), and disconnected (dotted) trajectories diverge substantially. (c) L2 divergence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

4.7.1 Self-Connection Polarity (Primary Mechanistic Finding)

High-embodiment (score ≥ 0.70, $n = 19$) and low-embodiment (score < 0.30, $n = 14$) solutions show dramatically different self-connection distributions: mean self-connection $+0.36 \pm 11.7$ (high) vs. -19.18 ± 17.9 (low), Mann-Whitney $p = 0.002$ [survives FDR], Cohen's $d = +1.29$. Correlation with embodiment dependence: Spearman $\rho = +0.369$, $p = 0.004$ [survives FDR], $\rho^2 = 0.14$. Self-connection polarity alone explains 14% of variance—comparable to network size alone ($\approx 16\%$).

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.

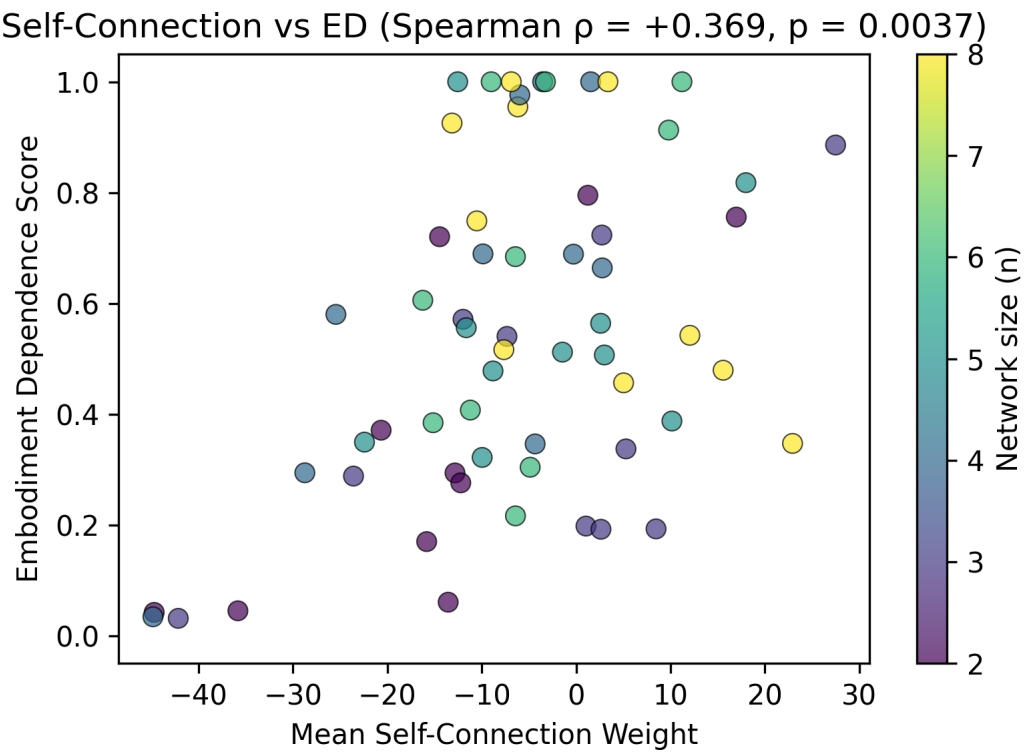

Figure 6: Self-connection polarity vs. embodiment dependence score for all 60 conditions with decoded genotypes (Spearman $\rho = +0.369$, $p = 0.004$). High-embodiment solutions show positive self-connections enabling amplifying dynamics; low-embodiment solutions show negative self-connections enabling stable fixed-point attractors.

4.7.2 Eigenvalue Structure

The max real eigenvalue of the weight matrix differs dramatically between groups: 45.8 ± 22.5 (high) vs. 7.0 ± 34.3 (low), $p = 0.003$ [survives FDR], Cohen's $d = +1.34$. 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 $\rho = +0.354$, $p = 0.006$ [survives FDR].

4.7.3 Variance Decomposition

The remaining $\approx 69\%$ likely reflects attractor geometry, bifurcation structure, and higher-order interaction effects not captured by summary statistics.

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

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

Predictor(s)	ρ^2
Growth rate alone	0.38
Input sensitivity alone	0.31
Network size alone	≈ 0.16
Self-connection alone	0.14
Max real eigenvalue alone	0.13

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 all 60 conditions. Like the weight configuration analysis, this is exploratory: the specific predictors (input sensitivity, bifurcation count) were identified during analysis rather than pre-specified. This analysis addresses the remaining unexplained variance by directly characterizing the attractor landscape.

4.8.1 Input Sensitivity: The Strongest Attractor-Geometry Predictor

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

Input sensitivity remains the strongest attractor-geometry predictor identified in this study, though the effect is smaller on the full 60-condition dataset ($\rho = +0.555$) than on the preliminary 42-condition subsample ($\rho = +0.660$). This attenuation is expected: the 18 recovered conditions (seeds 42, 137, 256) contribute more moderate values, reducing upward sampling bias. The relationship remains highly significant, second only to growth rate ($\rho = 0.615$) and stronger than fraction amplifying ($\rho = 0.525$), self-connection polarity ($\rho = 0.369$), and network size ($\rho = 0.394$). Input sensitivity correlates strongly with bifurcation count ($\rho = +0.681$, $p < 0.0001$), confirming that the two metrics capture overlapping but distinguishable features: bifurcations produce discrete qualitative transitions, while input

Table 3: Attractor geometry metrics and correlation with embodiment dependence (all 60 conditions).

Attractor metric	Spearman ρ	p-value	FDR status
Input sensitivity range	$+0.555$	< 0.0001	survives
Mean max real eigenvalue (at FPs)	$+0.539$	< 0.0001	survives
Trajectory variance at operating input	$+0.498$	< 0.0001	survives
Number of distinct regimes	$+0.424$	0.0007	survives
Number of bifurcations	$+0.408$	0.001	survives
Max Lyapunov exponent	$+0.323$	0.012	survives
Stable fixed point fraction	-0.352	0.006	survives

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 $\rho = +0.415$, $p = 0.001$), whereas bifurcation count shows no significant unique contribution after controlling for input sensitivity (partial $\rho = +0.049$, $p = 0.71$). 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.

4.8.2 Bifurcation Structure

High-embodiment solutions undergo significantly more bifurcations across input conditions. Comparing high-ED (score ≥ 0.70, $n = 19$) vs. low-ED (score < 0.30, $n = 14$): number of bifurcations: 2.53 vs. 0.00 (Mann-Whitney $p = 0.007$); trajectory variance at operating input: 42.8 vs. 0.007 ($p = 0.001$, $d = +0.61$); mean max real eigenvalue at fixed points: $+4.20$ vs. -3.18 ($p = 0.0001$, $d = +1.02$).

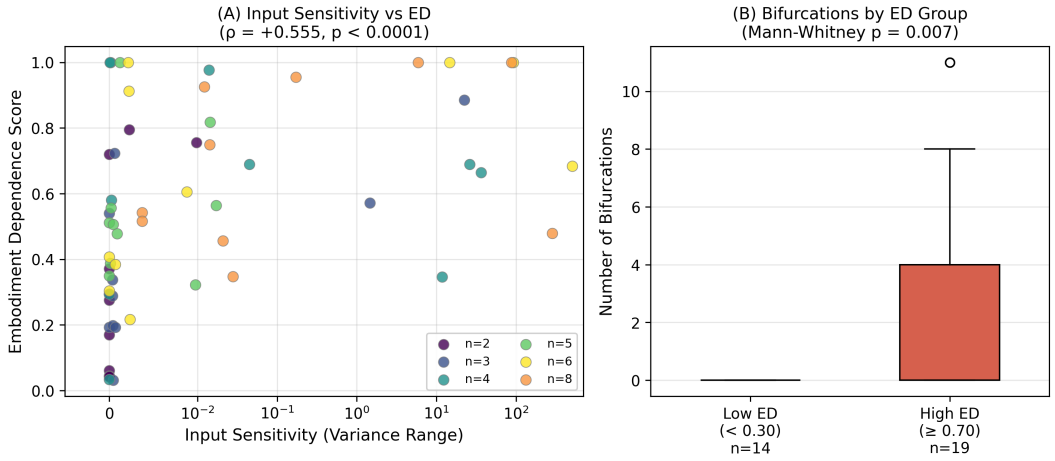

Figure 7: Input sensitivity and bifurcation analysis: the strongest attractor-geometry predictor. (A) Input sensitivity range vs. embodiment dependence score for all 60 conditions (Spearman $\rho = +0.555$, $p < 0.0001$). (B) Bifurcation count by embodiment dependence group (high ED ≥ 0.70 vs. low ED < 0.30), showing high-embodiment solutions undergo significantly more bifurcations across input conditions (Mann-Whitney $p = 0.007$).

Low-embodiment solutions have zero bifurcations—their dynamics are qualitatively stable across the entire input range. High-embodiment solutions undergo an average of nearly 3 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 60 conditions classify as: fixed point ($n = 49$, mean ED=0.505 $\pm$ 0.299), chaotic ($n = 7$, mean ED=0.740 $\pm$ 0.269), limit cycle ($n = 2$, mean ED=0.557), quasi-periodic ($n = 2$, mean ED=0.842). Chaotic and quasi-periodic solutions have significantly higher embodiment dependence than fixed-point solutions, but the majority of high-ED solutions 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 Assessment via Partial Correlation

We assessed 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: $\rho = +0.521$, $p < 0.0001$ [survives FDR]

- Max real eigenvalue → embodiment dependence: $\rho = +0.354$, $p = 0.006$ [survives FDR]
- Self-connection → embodiment dependence (direct): $\rho = +0.369$, $p = 0.004$ [survives FDR]
- Self-connection → embodiment dependence (controlling for eigenvalue): $\rho = +0.232$, $p = 0.077$ [FDR-ns]

When controlling for eigenvalue structure, the direct effect of self-connection polarity on embodiment dependence is reduced by approximately 37% (ρ drops from 0.369 to 0.232, $p = 0.077$). This is consistent with partial mediation—eigenvalue structure statistically subsumes roughly a third of the self-connection → embodiment dependence relationship—though the reduction in significance may also reflect limited statistical power. 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 underlying 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 all 60 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. $n = 21$, mean ED $= 0.607 \pm 0.275$.
- **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. $n = 9$, mean ED $= 0.206 \pm 0.190$.
- **Type C (Mixed)**: Intermediate self-connection and eigenvalue profiles. These networks exhibit partial input-dependence with mixed dynamical features. $n = 30$, mean ED $= 0.604 \pm 0.281$.

Predictive accuracy: Overall misclassification rate (Type A predicting low ED, or Type B predicting high ED): 3/60 = 5.0%. However, the resubstitution accuracy (95.0%) overstates generalization because the classification thresholds were defined on the same data. Leave-one-out cross-validation using a 3-nearest-neighbor classifier on non-defining features (growth rate, spectral radius, effective gain, positive self-connection fraction, mean time constant, mean absolute bias, mean absolute recurrent weight), which holds out each condition in turn, yields 71.7% accuracy—well above the 33% chance baseline but considerably lower than the resubstitution figure. This gap primarily reflects the broad Type C category: with 30/60 conditions (50%), a standard deviation of 0.281, and ED scores spanning nearly the full $[0, 1]$ range, Type C functions as a residual category rather than a mechanistically coherent type. The classification's primary explanatory value lies in the Type A/Type B distinction, which cleanly separates amplifying from stable dynamics; Type C captures conditions with intermediate or mixed dynamical profiles that require finer-grained characterization (e.g., by bifurcation structure or input sensitivity) to resolve. We report the LOO figure as the more appropriate estimate of classification performance.

Distribution by network size: Small networks ($n = 2$) are predominantly Type B or Type C; large networks ($n = 8$) are predominantly Type A or Type C, with minimal 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 $\rho = 0.394$, $p = 0.002$, $\rho^2 \approx 0.16$). While 16% explained variance may appear modest, this reflects a single architectural variable (neuron count) in a high-dimensional evolutionary design space where weight configurations, initial conditions, and stochastic selection pressures all contribute independently. In evolutionary robotics, where fitness landscapes are rugged and evolutionary trajectories are highly path-dependent, single-variable effects of this magnitude are noteworthy. The more robust finding is variance reduction: CV decreases from 85% to 37% as networks grow from $n = 2$ to $n = 8$, indicating that capacity constrains the distribution of possible outcomes even when it does not determine individual outcomes.

The mechanistic account is now multi-layered. At the weight level, positive self-connections create excitatory feedback loops ($\rho = +0.369$, $p = 0.004$, $\rho^2 = 0.14$). At the spectral level, these self-connections produce high eigenvalues in the weight matrix (SC $\rightarrow$ eigenvalue $\rho = +0.521$, $p < 0.0001$), 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 2.5 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 $\rho = +0.369$ ($p = 0.004$) to $\rho = +0.232$ ($p = 0.077$) when controlling for eigenvalue structure, suggesting that eigenvalue structure statistically subsumes roughly a third of the self-connection $\rightarrow$ embodiment dependence relationship. The remaining direct effect likely reflects additional pathways not captured by the single eigenvalue summary statistic—potentially including bifurcation structure, attractor basin geometry, or higher-order dynamical features that correlate with but are not fully

determined by eigenvalue properties.

Input sensitivity ($\rho = +0.555$) is the strongest attractor-geometry predictor of embodiment dependence—second only to growth rate ($\rho = 0.615$) and stronger than fraction amplifying ($\rho = 0.525$), self-connection polarity ($\rho = 0.37$), or network size ($\rho = 0.394$). 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. Relatedly, Aguilera et al. (2016) demonstrate that metastable neural dynamics in embodied agents extend beyond the nervous system into the sensorimotor loop, using Kuramoto oscillator networks in a phototaxis-like task to show that bidirectional agent-environment coupling is necessary for extended metastability. Their finding is consistent with our observation that high-embodiment solutions operate near bifurcation boundaries where neural and sensorimotor dynamics are tightly coupled. Our work extends this line of inquiry by asking a different question: not *whether* embodiment matters for neural dynamics (Aguilera et al. establish that it does), but *when* and *how much*—specifically, how the degree of embodiment dependence varies systematically with computational capacity across a population of 60 evolved CTRNN controllers, and what mechanistic pathway (self-connection polarity $\rightarrow$ eigenvalue structure $\rightarrow$ input-sensitive attractor geometry) determines this variation.

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 work not by uncovering fundamental causal mechanisms but by identifying *causal patterns*—stable regularities in dependence relations that are limited in scope, idealized for particular explanatory purposes, and robust across a specified range of background conditions. Crucially, Potochnik distinguishes causal patterns from mere regularities: a causal pattern must (i) abstract away from causally irrelevant detail, (ii) reveal a dependence relation that is invariant under a specified range of interventions, and (iii) serve a particular explanatory interest. The relationship between these criteria and Woodward (2003)'s interventionism is direct: patterns are revealed through interventions that test invariance.

Our results identify two causal patterns that meet these criteria. The *neural stability pattern* ($\approx$ 35% of small network runs) is a regularity in which negative self-connections produce strongly damped dynamics with intrinsically stable attractors, making the network's trajectory largely invariant under sensorimotor disruption. This pattern abstracts away from the specific weight values and network topology to isolate a dependence relation: negative self-feedback $\rightarrow$ eigenvalue stability $\rightarrow$ input-independent attractors $\rightarrow$ low embodiment dependence. The pattern is robust across different seeds, network sizes (primarily n = 2–3), and the three ghost condition interventions. It serves the explanatory interest of understanding *when bodies do not matter*—when neural dynamics are self-sustaining.

The *neural coupling pattern* ($\approx$ 35% of large network runs) is a regularity in which positive self-connections produce amplifying dynamics with input-sensitive attractor landscapes, making the network's trajectory critically dependent on sensorimotor coupling. This pattern again abstracts away from specifics to isolate a dependence chain: positive self-feedback $\rightarrow$ marginal eigenvalues $\rightarrow$ input-sensitive attractors $\rightarrow$ high embodiment dependence. It is robust across seeds and sizes (primarily n = 6–8), and verified via three independent types of coupling disruption. It serves the explanatory interest of understanding *when bodies matter*—when neural dynamics are constitutively entangled with sensorimotor structure.

Neither pattern is "the true causal structure." In Potochnik's framework, the choice between highlighting the stability pattern or the coupling pattern depends on what we want to explain. A researcher investigating autonomous cognition (brain-in-a-vat scenarios) might foreground the stability pattern. A researcher investigating sensorimotor dependence foregrounds the coupling pattern. Both are genuine causal patterns in Potochnik's sense—neither is a mere regularity, because both are invariant under intervention and both abstract away from irrelevant detail to reveal dependence relations. The coexistence of both patterns within a single evolutionary system underscores Potochnik's broader argument that causal explanation is interest-relative without being arbitrary.

5.4 Scope Clarifications

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. Our contribution is to measure it precisely, establish its relationship to network capacity, and characterize mechanistic factors.

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 (37–85%) is substantial. Architecture explains $\approx 16\%$ of variance; weight configuration $\approx 14\%$ (ρ^2); attractor input sensitivity $\approx 31\%$ (ρ^2); growth rate explains $\approx 38\%$ (ρ^2). Higher-order interactions and unmeasured features account for the remainder.

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

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

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

5. **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. **Exploratory analyses**: The weight configuration and attractor geometry analyses (Sections 4.7–4.9) are exploratory, developed to explain variance in the primary finding. Though all key correlations survive FDR correction, independent replication on new evolutionary conditions or different tasks is needed to confirm these mechanistic relationships.

7. **Toy model epistemology**: Our 2–8 neuron CTRNNs are minimal models in the sense articulated by Beer (2024) and consistent with Levins's (1966) argument that model-building in biology involves deliberate trade-offs between generality, realism, and precision. They demonstrate that embodiment dependence *can* vary systematically with computational capacity and identify mechanistic conditions sufficient for this variation, but they do not establish that biological nervous systems must operate according to the same principles. Following Beer's framework and Weisberg's (2013) analysis of idealized models, these results are best understood as mathematically rigorous explorations of what is possible within a well-defined model class, generating specific hypotheses (e.g., that positive recurrent gain produces coupling-dependent dynamics) testable in richer systems.

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 $\rho = 0.394$, $p = 0.002$), with variance reduction from 85% CV ($n = 2$) to 37% CV ($n = 8$) as the most robust finding.

Three levels of mechanistic analysis across all 60 conditions illuminate why this relationship holds. At the weight level, self-connection polarity predicts embodiment dependence ($\rho = +0.369$, $p = 0.004$): 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 attractor-geometry predictor of embodiment dependence**

($\rho = +0.555$, $p < 0.0001$), and partial correlation analysis is consistent with eigenvalue structure partially accounting for the self-connection $\rightarrow$ embodiment dependence relationship (approximately 37% reduction when controlling for eigenvalue structure). We classify evolved solutions into three mechanistic types—amplifying (Type A), stable (Type B), and mixed (Type C)—achieving 71.7% leave-one-out cross-validation 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. A particularly promising direction is extending the framework to social interaction, where embodiment dependence may take on a qualitatively different character: recent work on evolved social agents demonstrates that successful interaction relies on co-constructed dynamical mechanisms that emerge between agents rather than within them (Severino et al., 2026; Merritt et al., 2024).

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 individual min(1.0) capping. Each ghost condition's divergence is individually capped at 1.0 before averaging, ensuring each condition contributes on a bounded $[0, 1]$ scale while preserving ranking and statistical significance.

Table 4: Summary statistics of raw uncapped divergence by network size. Values represent the mean L2 neural divergence averaged across frozen-body, disconnected, and counterfactual ghost conditions before min(1.0) capping.

Network size n	Mean	Median	Range	Conditions capped
2	1.09	0.28	[0.04, 5.78]	3/10
3	2.07	0.31	[0.03, 17.59]	1/10
4	1.08	0.73	[0.03, 3.66]	6/10
5	1.02	0.51	[0.32, 4.60]	3/10
6	14.71	0.68	[0.22, 126.03]	5/10
8	4.82	0.65	[0.35, 22.82]	4/10

Raw uncapped divergence exhibits extreme positive skew, particularly for $n = 6$ (one condition produces L2 divergence > 126, driven by a network whose neural dynamics explode under sensory disconnection). The min(1.0) capping operation compresses this distribution

to [0,1], preserving rank order but removing scale information. The "Conditions capped" column shows how many of the 10 seeds per size exceed 1.0 before capping; these conditions receive a capped score of 1.0 (maximum embodiment dependence). Spearman rank correlations, which depend only on rank order, are unaffected by the capping transformation. Complete per-condition divergence values are available in the data repository.

B Sensitivity Analysis

This appendix reports sensitivity of main conclusions to two sources of analytic variation: (1) the capping threshold used to normalize embodiment dependence scores, and (2) the choice of ghost condition.

Robustness to Capping Threshold

We assessed whether the capacity-dependence correlation is robust to the capping threshold applied to individual ghost condition divergence values. The default analysis uses $\min(1.0, D)$ for each condition; here we also report results using cap values of 0.5 and 2.0, as well as uncapped raw divergence values. "Conditions capped" indicates how many conditions have at least one ghost condition divergence exceeding the threshold.

Table 5: Spearman correlation (ρ) between network size and embodiment dependence score under different capping thresholds ($n = 60$).

Capping Threshold	ρ	p-value	95% Bootstrap CI	Conditions capped
0.50	0.398	0.002	[0.15, 0.60]	43/60
1.00 (Default)	0.394	0.002	[0.14, 0.60]	22/60
2.00	0.401	0.002	[0.14, 0.61]	16/60
Uncapped	0.370	0.004	[0.10, 0.59]	0/60

Robustness Across Individual Ghost Conditions

We also assessed whether the capacity-dependence relationship depends on averaging across all three ghost conditions, or whether it holds for each condition individually.

Robustness findings: The Spearman correlation between network size and embodiment dependence is significant ($p < 0.007$) across all capping thresholds tested and across all three individual ghost conditions, with ρ ranging from 0.348 to 0.401. The capacity-dependence relationship is therefore robust to both normalization choices and the specific operationalization of coupling disruption. The disconnected and counterfactual conditions produce highly

Table 6: Spearman correlation (ρ) between network size and embodiment dependence measured by individual ghost conditions ($n = 60$).

Ghost Condition	ρ	p-value	95% Bootstrap CI
Frozen body only	0.391	0.002	[0.16, 0.58]
Disconnected only	0.366	0.004	[0.11, 0.58]
Counterfactual only	0.348	0.006	[0.09, 0.56]
Average (Default)	0.394	0.002	[0.14, 0.60]

correlated divergence values (Pearson $r = 0.990$), while the frozen body condition captures a partially distinct aspect of coupling dependence ($r = 0.241$ and $r = 0.165$ with disconnected and counterfactual, respectively). This pattern is consistent with the frozen body condition testing motor-to-sensory feedback specifically, while disconnected and counterfactual both eliminate meaningful sensory input (one via zeroing, the other via randomization).

C Complete Multiple Comparison Correction

Table 7 reports all 26 statistical tests with Benjamini-Hochberg FDR-corrected q-values. Of these, 23 survive at $q < 0.05$. The three non-surviving tests are: the Kruskal-Wallis test across sizes ($q = 0.051$, marginally above threshold), the partial correlation of self-connection with ED controlling for eigenvalue ($q = 0.078$, marginal, consistent with partial mediation), and the partial correlation of bifurcation count with ED controlling for input sensitivity ($q = 0.711$, expected non-significance demonstrating that input sensitivity subsumes bifurcation count).

References

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

Aguilera, M., Bedia, M. G., and Barandiaran, X. E. (2016). Extended neural metastability in an embodied model of sensorimotor coupling. *Frontiers in Systems Neuroscience*, 10:76.

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.

Beer, R. D. (2021). Some historical context for minimal cognition. *Adaptive Behavior*, 29(1):89–92.

Beer, R. D. (2023). On the proper treatment of dynamics in cognitive science. *Topics in Cognitive Science*.

Beer, R. D. (2024). Milking a spherical cow: Toy models in neuroscience. *European Journal of Neuroscience*.

Beer, R. D. and Di Paolo, E. A. (2023). The theoretical foundations of enaction: Precariousness. *BioSystems*, 223:104823.

Benettin, G., Galgani, L., Giorgilli, A., and Strelcyn, J.-M. (1980). Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. *Meccanica*, 15(1):9–20.

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.

Kiverstein, J. and Kirchhoff, M. D. (2023). Dissolving the causal-constitution fallacy: Diachronic constitution and the metaphysics of extended cognition. In *Situated Cognition Research*, volume 23 of *Studies in Brain and Mind*. Springer.

Krickel, B. (2023). Why diachronic constitution won't help: Commentary on 'Dissolving the causal-constitution fallacy'. In *Situated Cognition Research*, volume 23 of *Studies in Brain and Mind*. Springer.

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.

Merritt, H., Severino, G. J., and Izquierdo, E. J. (2024). The dynamics of social interaction among evolved model agents. *Artificial Life*, 30:216–239.

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

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.

Severino, G. J., Winkler, S. L., Beer, R. D., and Barwich, A.-S. (2026). Social contingency in embodied neural networks relies on co-constructed dynamical mechanisms. *Philosophical Transactions of the Royal Society B*, 381:20250098.

Siegelmann, H. T. and Sontag, E. D. (1995). On the computational power of neural nets. *Journal of Computer and System Sciences*, 50(1):132–150.

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.

Table 7: All 26 statistical tests with Benjamini-Hochberg FDR correction. Section numbers with asterisks refer to attractor geometry analyses.

Test	Sec.	Raw p	BH q	Surv.	Type
Spearman: size vs ED	4.2	0.002	0.004	Yes	Conf.
Pearson: size vs ED	4.2	0.002	0.004	Yes	Conf.
Kruskal-Wallis: ED across sizes	4.2	0.047	0.051	No	Conf.
Mann-Whitney: small vs large	4.2	0.002	0.003	Yes	Conf.
Spearman: growth rate vs ED	4.4	<0.001	<0.001	Yes	Conf.
Spearman: PR vs ED	4.4	0.016	0.018	Yes	Conf.
Spearman: frac. amplifying vs ED	4.4	<0.001	<0.001	Yes	Conf.
Spearman: max Lyapunov vs ED	4.4	0.014	0.017	Yes	Conf.
Spearman: self-conn vs ED	4.5	0.004	0.006	Yes	Expl.
Spearman: max eigenvalue vs ED	4.5	0.006	0.008	Yes	Expl.
Spearman: self-conn vs eigenvalue	4.6	<0.001	<0.001	Yes	Expl.
Partial: self-conn vs ED | eigenvalue	4.6	0.077	0.081	No	Expl.
Mann-Whitney: self-conn high vs low	4.5	0.002	0.004	Yes	Expl.
Mann-Whitney: eigenvalue high vs low	4.5	0.003	0.004	Yes	Expl.
Spearman: input sensitivity vs ED	4.5*	<0.001	<0.001	Yes	Expl.
Spearman: mean max eig (FPs) vs ED	4.5*	<0.001	<0.001	Yes	Expl.
Spearman: traj. variance vs ED	4.5*	<0.001	<0.001	Yes	Expl.
Spearman: num regimes vs ED	4.5*	<0.001	0.002	Yes	Expl.
Spearman: bifurcation count vs ED	4.5*	0.001	0.003	Yes	Expl.
Spearman: max Lyapunov (attr) vs ED	4.5*	0.012	0.015	Yes	Expl.
Spearman: stable FP frac vs ED	4.5*	0.006	0.008	Yes	Expl.
Partial: input sens vs ED | bifurc	4.5*	0.001	0.003	Yes	Expl.
Partial: bifurc vs ED | input sens	4.5*	0.713	0.713	No	Expl.
Mann-Whitney: bifurc high vs low	4.5*	0.007	0.009	Yes	Expl.
Mann-Whitney: traj var high vs low	4.5*	0.001	0.002	Yes	Expl.
Mann-Whitney: eig (FPs) high vs low	4.5*	<0.001	<0.001	Yes	Expl.