From Chemistry to Ecology: Codes and Predation Emerge from Coherence Constraints in Protocell Networks

Ian Todd
Sydney Medical School
University of Sydney
Sydney, NSW, Australia
itod2305@uni.sydney.edu.au

Abstract

We address two fundamental questions in abiogenesis: (1) how did symbolic codes arise from continuous chemistry, and (2) why didn't early life homogenize into synchronous goo? We propose that both answers follow from coherence constraints. First, codes emerge as coordination interfaces between coupled protocellular compartments—stable patterns that allow neighbors to influence each other without full state knowledge. Substrate competition (lateral inhibition) drives winner-take-most dynamics that make continuous outputs reliably distinguishable; discrete symbols arise downstream when receivers threshold these signals. Our simulations show that networks of 61 coupled compartments reliably map 32 environmental configurations to distinguishable code vectors with 98% decoding accuracy. Second, coherence is metabolically expensive and scales superlinearly with group size, creating an inherent ceiling on coordination. This ceiling prevents homogenization and generates predator-prey dynamics: small coherent "predators" (N≈39, C≈0.53) extract resources from large incoherent "prey" (N≈183, C≈0.12) through asymmetric raids enabled by coherence differentials. No specialized predation machinery is required. We propose that ecology may not be an add-on to life but a stabilizer that made abiogenesis viable—predator-prey-like dynamics and social coordination, in the limited dynamical sense of asymmetric extraction between differentially coherent populations, may predate biological complexity, arising from physics before genetic encoding. In this framework, communication precedes information storage; ecological differentiation precedes biological complexity.

Keywords: origin of life; genetic code; protocells; coordination; predator-prey dynamics; ecological phase transitions

1 Introduction

1.1 The Code Emergence Problem

The origin of the genetic code remains one of the deepest puzzles in biology. How did discrete, symbolic information arise from continuous chemical processes? Standard accounts assume that codes require encoders, messages, and decoders—but this raises a circularity problem: how can encoding machinery exist before the code that specifies it?

Decades of experimental work have demonstrated prebiotic synthesis of nucleotides [2, 3] and self-assembly of protocell compartments [1, 4, 34]. These are major achievements in *chemistry*. But they do not address the *coding* problem: why does UUU encode phenylalanine rather than any other amino acid? The assignment problem—the mapping from codons to amino acids—remains unexplained. Stereochemical theories [17] propose direct chemical affinity; frozen accident theories [18] propose historical contingency. Neither provides a *mechanism* for how discrete codes emerge from continuous chemistry.

Theoretical frameworks have addressed the dynamics of prebiotic organization without solving the code emergence problem directly. Kauffman's autocatalytic sets [31] and the related RAF theory [36] show how self-sustaining reaction networks can emerge, but do not explain the transition from continuous chemistry to discrete symbols. Eigen's hypercycle model [32] addresses cooperation among replicators but assumes the existence of coded molecules. Maturana and Varela's autopoiesis [33] characterizes self-maintaining organization but does not specify the mechanism of symbolic emergence. Metabolism-first approaches [35] emphasize thermodynamic disequilibrium as the driver of organization, which we incorporate, but do not address how continuous metabolic gradients become discrete codes.

We propose a different framing: **codes could have originated not for intracellular replication but for intercellular coordination**. The standard narrative assumes molecular machinery first, cooperation later:

$$\textit{chemistry} \rightarrow \textit{molecules} \rightarrow \textit{replication} \rightarrow \textit{cells} \rightarrow \textit{ecology} \rightarrow \textit{cooperation}$$

Recent work on autocatalytic chemical ecosystems [5, 7] has established the conceptual possibility of ecological dynamics prior to bounded cells. We go further, proposing both an inversion of the standard sequence and the mechanisms that drive it:

$$\textit{compartments} \rightarrow \textit{coordination pressure} \rightarrow \textit{ecological differentiation} \rightarrow \textit{codes} \rightarrow \textit{genetic internalization}$$

The ecological dynamics we observe in mature biology—predator-prey relationships, competition, metabolic coordination costs—are not late additions to an already-coded system. They are the *drivers* of code emergence. Coordination pressure between compartments created selection for discrete signaling interfaces; the genetic code is a late internalization of protocols that first existed in the coupling fields between cells.

Our contribution is a specific mechanism: **substrate competition** via cooperative binding (Hill kinetics, $h > 1$) produces discrete output attractors from continuous input gradients. We also derive the coordination ceiling from signal propagation physics, explaining why coherent groups cannot grow indefinitely.

Specifically, we propose that **codes first emerged as coordination interfaces between coupled protocellular compartments**. They were not transmitted messages but stable patterns that allowed neighbors to influence each other's behavior without full state knowledge. This view aligns with models showing that genetic code structure can emerge via horizontal exchange between protocells rather than vertical descent [9]. The encoder-message-decoder structure we recognize in mature biological systems is a *late refinement*—the internalization and compression of coordination protocols that first existed in the distributed coupling fields between cells.

1.2 The Homogenization Problem

A second puzzle immediately follows: if coherent, coordinated dynamics can arise spontaneously, why doesn't everything simply synchronize into one homogeneous blob? What prevents the entropy-minimizing outcome of global coherence?

We propose that the answer is **metabolic**: coherence is expensive, and the cost scales superlinearly with group size [11]. This creates an inherent ceiling on coherent aggregates, which in turn creates ecological differentiation—and predator-prey dynamics—without any specialized predation machinery.

1.3 Two Problems, One Framework

These problems are linked. Code emergence requires coordination; coordination requires coherence; coherence has metabolic limits. The same physics that enables symbolic communication also prevents homogenization and drives ecological complexity.

This paper presents both results in a unified framework:

- **Part I** (Sections 2–4): Codes emerge from coordination equilibria in coupled protocell networks
- **Part II** (Sections 5–6): Metabolic ceilings on coherence generate predator-prey dynamics

Extended mathematical foundations are developed in companion preprints [13, 15]; this paper is self-contained.

2 Theoretical Framework: Codes as Coordination

2.1 Codes Are Coordination Equilibria

Consider a network of weakly coupled protocells:

- Each protocell has high-dimensional internal dynamics (many coupled molecular species)
- Neighbors communicate through multiple channels: bioelectric, chemical, mechanical, redox [12]
- These signals are naturally discretized by threshold physics

- Selection favors configurations where neighbors reliably respond to each other's boundary states

The resulting "code" is not a transmitted message. It is a **shared interface**—a stable pattern that allows coordination without requiring each compartment to know the other's full internal state.

2.2 The Group-First Hypothesis

A lone protocell faces brutal tradeoffs: too open and it mixes with environment; too closed and it stagnates; too complex and it collapses. A **network of weakly coupled compartments** relaxes these constraints. Spatial heterogeneity becomes memory. Weak coupling enables coordination without collapse. Selection operates at the group level.

We hypothesize that early life was more likely a coupled compartment network than isolated protocells [1, 4, 10].

2.3 Coherence Is Not Information

A critical distinction underlies this framework: **coherence** (high-dimensional phase order, coordination capacity) is not the same as **information** (discrete symbols that can be copied and decoded). Coherence enables coordination; information emerges when coherent dynamics collapse into reproducible invariants.

The bridge: *codes form when coherence collapses into reproducible invariants.*

3 Code Emergence Model

3.1 Architecture

We implement a minimal coupled-compartment system based on prebiotic amphiphile vesicles [8]:

- 61 vesicles in hexagonal array
- 128 internal dimensions per vesicle (high-dimensional reservoir)
- 30 readout channels with substrate competition
- Weak neighbor coupling via boundary signals

Biological interpretation. This architecture maps onto a prebiotic ecosystem containing all three major components: vesicles provide compartmentalization, the 128 internal dimensions represent peptide catalyst concentrations and conformational states, and the readout channels represent peptide-nucleotide binding sites competing for shared nucleotide substrate pools. The Hill kinetics of substrate competition (Eq. 1) directly models cooperative peptide-substrate binding. In this interpretation, the "codes" that emerge are proto-metabolic phenotypes—stable patterns of peptide activity that differentiate vesicle populations before any template-directed replication exists.

3.2 The Discretization Mechanism

The critical innovation is **substrate competition**—a lateral inhibition mechanism where output channels compete for shared metabolic resources. Consider M output channels drawing from a common substrate pool of concentration S. Each channel j has raw activation a_j determined by the vesicle's internal state. The competitive output is:

$$\phi_j = \frac{a_j^h}{\sum_k a_k^h + \epsilon} \tag{1}$$

where $h > 1$ is the Hill coefficient controlling winner-take-most sharpness, and ϵ prevents division by zero.

This is the quasi-steady-state solution to competitive binding kinetics. The denominator arises from mass conservation: when channel A consumes substrate, channel B cannot have it. This forces winner-take-most dynamics that discretize continuous states without programmed thresholds.

The mechanism is prebiotically plausible. Substrate competition occurs naturally whenever multiple reactions draw from a shared pool—a universal feature of metabolism. The Hill exponent h emerges from cooperative binding. We use $h = 4$ as a baseline; discretization is robust across $h \in [2, 8]$.

Key property: The discretization is *emergent*, not imposed. No external logic specifies which channel should win. The physics of competition creates digital outputs from analog inputs—precisely the transition required for symbolic codes to arise from continuous chemistry.

Mechanism interaction: Substrate competition alone is noise-sensitive—it amplifies whatever signal it receives, including noise. Reliable discretization requires the upstream reservoir dynamics to create *winner margins*: clear separation between the strongest and second-strongest activations. Even modest margins ($\sim 2\times$) yield robust outputs when amplified by $h = 4$. The reservoir dynamics create these margins through attractor formation; the competition mechanism then sharpens them into discrete codes. Neither component alone suffices.

Code Definition (Operational)

Code signal: Each temporal cycle produces a 60-dimensional vector $\mathbf{c}_t = [\boldsymbol{\phi}_{\text{center}}; \boldsymbol{\phi}_{\text{edge}}]$ (30 center + 30 edge readout channels). The full 4-cycle codeword is the 240-dimensional concatenation $\mathbf{C} = [\mathbf{c}_1; \mathbf{c}_2; \mathbf{c}_3; \mathbf{c}_4]$, capturing both spatial pattern and temporal sequence.

Decoder rule: Nearest-centroid classification. Given code $\mathbf{c}$, assign to environment $e^* = \arg\min_e \|\mathbf{c} - \boldsymbol{\mu}_e\|$, where $\boldsymbol{\mu}_e$ is the mean code vector for environment e.

Metrics: Decoding accuracy = fraction correctly classified; Separation ratio = $\bar{d}_{\text{between}}/\bar{d}_{\text{within}}$ in code space.

Note: The code is a continuous chemical signal (concentration profile), not a discrete symbol sequence. Discretization occurs downstream in receiver dynamics, consistent with biological signaling where continuous ligand concentrations drive discrete cellular responses.

3.3 Reservoir Dynamics

Each vesicle's internal state evolves as a high-dimensional reservoir:

$$\mathbf{x}_{t+1} = (1 - \alpha)\mathbf{x}_t + \alpha \tanh(W_{\text{res}}\mathbf{x}_t + W_{\text{in}}\mathbf{u}_t + W_{\text{couple}}\mathbf{b}_t) \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^{128}$ is the internal state, $\mathbf{u}$ is environmental input, $\mathbf{b}$ is the boundary signal from neighbors, $\alpha = 0.25$ is the leak rate, and W_{res} has spectral radius 0.92 (edge of chaos). The coupling matrix W_{couple} has strength $\kappa = 0.15$ by default.

3.4 Environmental Input

Each vesicle experiences spatially varying stimulus (center-edge gradients, top-bottom gradients). Temporal forcing across 4 cycles creates genuine sequence structure. Each of 5 bits (2 spatial $\times$ 2 temporal + 1 global) can be on or off, yielding $2^5 = 32$ distinct environmental configurations.

The complexity is not in the stimulus; it is in the system's ability to differentiate continuous gradients into discrete coordination states. We deliberately use simple inputs to demonstrate that code structure emerges from the network dynamics, not from input complexity.

Figure 1 illustrates the system architecture, showing the hexagonal vesicle array (A), the substrate competition mechanism (B), and the resulting environment-to-code mapping (C).

4 Code Emergence Results

4.1 Primary Results

Metric	Result
Unique input$\rightarrow$character mappings	32/32 (no collisions)
Reproducibility (20 seeds)	$98.4\% \pm 2.1\%$
Separation ratio (attractor space)	335,000$\times$
Env–Attractor correlation	0.72
Bimodality (emergent discretization)	89% saturated

The system reliably maps 32 environmental configurations to 32 distinguishable code vectors (separation ratio $> 300\times$). A second vesicle array ("receiver colony") achieves 98% decoding accuracy using only physics—no machine learning.

Receiver specification: We test two distinct receiver regimes to disentangle "same chemistry" from "same individual":

1. **Matched-realization receiver**: Identical weight matrix as encoder (same random seed). This represents clonal reproduction—"the same individual chemistry." Achieves 98% decoding accuracy.
2. **Same-statistics receiver**: Independently sampled weights from identical distribution (same spectral radius, sparsity, connectivity rules; different random realization). This represents "same species, different individual." Achieves 87% decoding accuracy.

Both receivers share coupling topology, Hill kinetics parameters, and readout architecture. Both are initialized with independent random states. Decoding is performed by nearest-attractor matching: the receiver's output is compared to canonical attractors from the encoder, assigned to the closest match in Euclidean distance. No gradient descent or parameter optimization is involved.

The 87% accuracy of same-statistics receivers demonstrates that the code structure is partially *chemistry-general*: it arises from the dynamical regime (competition + coupling), not from specific weight values. The 11% accuracy gap to matched-realization suggests that some fine structure is individual-specific—consistent with the biological observation that conspecifics share communication systems imperfectly.

4.2 Ablation Controls

To demonstrate that code emergence requires the proposed mechanisms, we tested ablated configurations:

Condition	**Unique Codes**	**Reproducibility**	**Sep. (attractor)**
Full model	32/32	98.4%	335,000×
No coupling ($\kappa = 0$)	32/32	97.1%	68×
No competition ($h = 1$)	8/32	62.3%	2.1×
Noisy input ($\pm 20\%$)	32/32	94.8%	198,000×
No temporal cycles	12/32	71.2%	14×
Random topology	32/32	89.4%	12,400×
No mean-centering	32/32	98.1%	>10,000×

Interpretation: Substrate competition ($h > 1$) is *essential*—without it, codes collapse to a handful of attractors with poor reproducibility. Temporal structure contributes significantly: removing cycles degrades separation ratio from $> 300\times$ to $< 50\times$. The system is robust to input noise and topology perturbations.

On mean-centering: Mean-centering in the readout (modeling differential measurement) does *not* create digitization—removing it maintains winner-take-most outputs (92% bimodal) with high separation. This confirms that discretization arises from substrate competition physics, not from any centering or normalization step. The mean-centering is instrumentation (analogous to reference electrode subtraction), not an engineered digitizer.

On the $\kappa = 0$ result: Uncoupled vesicles still produce distinguishable codes, but these are not *coordination interfaces*—they are local discretizations with no shared structure. Without coupling, separation collapses from 335,000× to 68×, and receiver colonies cannot reliably decode (accuracy drops to 71%). Coupling is what transforms discretized internal categories into a shared coordination manifold that supports high-margin decoding and environmental geometry transfer.

4.3 Information-Theoretic Capacity

To move beyond "we observed 32 distinguishable codes," we quantify decodability information-theoretically. Let E be the environment class (uniform over 32 configurations) and $\hat{E}$ be the

decoded environment from the code vector. We estimate the mutual information $I(E; \hat{E})$ from empirical confusion matrices computed over 20 random seeds $\times$ 32 environments = 640 samples, with Miller–Madow bias correction for finite sample size.

Fano's inequality provides a lower bound on information required for classification: if decoding error probability is P_e for K classes, then

$$I(E; \hat{E}) \geq \log_2 K - H(P_e) - P_e \log_2(K - 1) \tag{3}$$

where $H(P_e) = -P_e \log_2 P_e - (1 - P_e) \log_2(1 - P_e)$ is binary entropy.

For $K = 32$ classes and observed $P_e = 0.016$ (98.4% accuracy):

- Required: $I(E; \hat{E}) \geq 5.0 - 0.12 - 0.08 = 4.80$ bits
- Measured: $I(E; \hat{E}) = 4.91 \pm 0.03$ bits (20 seeds)

The system transmits 98% of the theoretical maximum (5 bits for 32 classes). This is not an artifact of our nearest-centroid decoder—*any* decoder, biological or computational, is bounded by the information in the code.

Condition	$I(E; \hat{E})$ (bits)	Fano bound	Capacity ratio
Full model ($\kappa = 0.15$)	4.91	4.80	98%
No coupling ($\kappa = 0$)	3.42	2.91	68%
No competition ($h = 1$)	1.87	—	37%
Intermediate coupling ($\kappa = 0.30$)	4.96	4.88	99%

Information capacity peaks at intermediate coupling ($\kappa \approx 0.30$), matching the separation maximum in Figure 2C. This is not coincidence: coupling expands the accessible manifold, increasing channel capacity, until excessive coupling collapses dynamics into low-dimensional synchrony.

4.4 Scale Dependence

We tested whether code emergence persists at larger scales:

Metric	**Medium (61×128D)**	**Massive (169×512D)**
Unique codes	32/32	32/32
D_{eff}	6.3	16.7
Env–Attractor correlation	0.72	0.83
Reproducibility	98.4%	84.2%
Separation (attractor)	335,000×	4.4×

Note on separation metrics: The 335,000× and 4.4× values are both measured in receiver attractor space. The dramatic difference reflects that massive-scale systems have higher within-class variance (less stable attractors), not that between-class distances shrink. Interface-space separation (transmitted code vectors) shows a similar but less extreme pattern. Despite lower separation ratios, decoding accuracy remains high (98% medium, 92% massive) because absolute distances still exceed decision thresholds.

Code discrimination persists at massive scale (32/32), but reproducibility and separation degrade. The larger system is more expressive but less stable—it explores more of its state space but settles less reliably into fixed attractors. **This instability at scale is not a bug; it is the coordination ceiling that prevents homogenization** (interpreted metabolically in Part II).

Figure 2 quantifies this trade-off. Panel A shows reproducibility declining from 99% at small scale to 84% at massive scale. Panel B shows that effective dimensionality grows sublinearly with total dimensions—the system cannot exploit all its degrees of freedom coherently. Panel C demonstrates that code separation peaks at intermediate coupling, the regime where compartments influence but do not dominate each other.

Effective dimensionality (D_{eff}) is computed via the participation ratio of eigenvalue spectra:

$$D_{\text{eff}} = \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2} \tag{4}$$

where λ_i are eigenvalues of the code covariance matrix. This measure counts the number of dimensions that contribute meaningfully to variance—a manifold with one dominant axis has $D_{\text{eff}} \approx 1$, while an isotropic distribution has $D_{\text{eff}} \approx N$.

4.5 Manifold Expansion

Coupling increases effective dimensionality (see [13] for theoretical foundations):

Coupling κ	D_{eff}	Sep. (interface)
0.00	6.82	68×
0.15	6.86	343×
0.30	7.21	3,110×
0.50	7.61	235×

Separation peaks at intermediate coupling—the regime where compartments influence but do not dominate each other.

4.6 Capacity Scaling Law

Treating the protocell network as a communication channel, we ask: how does code capacity $K_{\max}$ (maximum distinguishable environments) scale with system parameters?

We empirically find:

$$\log_2 K_{\max} \approx a \cdot D_{\text{eff}}^{0.8} \cdot f(\kappa) \tag{5}$$

where D_{eff} is effective dimensionality and $f(\kappa)$ captures the coupling-dependent efficiency:

$$f(\kappa) = \frac{4\kappa(1-\kappa)}{(1+\kappa)^2}$$

peaking at intermediate $\kappa \approx 0.3$.

Scale	D_{eff}	κ_{opt}	$K_{\max}$ (bits)
Small (19×64)	4.2	0.28	4.8
Medium (61×128)	6.3	0.31	5.0
Large (127×256)	11.4	0.29	5.4
Massive (169×512)	16.7	0.27	5.1*

*Capacity drops at massive scale due to reproducibility collapse (coordination ceiling).

This is a **phase diagram result**: capacity peaks at intermediate coupling and intermediate scale, collapsing both when coupling is too weak (no shared manifold) and when the system exceeds the coordination ceiling (reproducibility loss). The protocell network operates as an information channel with physically determined capacity.

5 Ecological Extension: The Coherence-Scale Trade-off

The instability at large scale (Section 4) points to a deeper constraint: coherence is metabolically expensive and scales superlinearly with group size.

Variable Mapping Across Parts I and II

Part I (Code Model)	*Part II (Ecology)*	*Interpretation*
D_{eff}	D (Appendix)	Effective dimensionality
Reproducibility	—	Code stability metric
—	C	Coherence trait (normalized)
—	N	Group size

Connection: We interpret $C \propto D_{\text{eff}}/D_{\max}$ as normalized coherence—the fraction of available degrees of freedom that move together. Part I demonstrates that large systems cannot maintain high $D_{\text{eff}}/D_{\text{total}}$; Part II models this as a cost on C.

5.1 The Metabolic Ceiling

Maintaining coherence costs energy:

$$\text{Cost} = k \cdot N^{\gamma} \cdot C^{\zeta} \tag{6}$$

with $\gamma > 1$ (superlinear in size). The exponent $\gamma > 1$ captures coordination overhead: all-to-all coupling scales as $O(N^2)$, and even hierarchical schemes face superlinear costs.

Maximizing fitness with respect to N at fixed coherence C yields optimal size:

$$N^* = \left(\frac{\bar{R} \cdot \phi(C)}{k\gamma C^{\zeta}} \right)^{1/(\gamma-1)} \tag{7}$$

Since $\gamma > 1$, higher C implies lower N^*. **High coherence forces small size.**

5.2 Two Stable Strategies

This creates two stable strategies:

- **Type B (big/weak)**: Large N, low C. Wins by occupying space.
- **Type S (small/coherent)**: Small N, high C. Wins by concentrated, coordinated action.

Neither can invade the other's niche through pure competition. This is ecological differentiation without any ecology-specific assumptions.

5.3 The Extraction Operator

Small coherent groups cannot grow by expansion (ceiling binds). But they *can* extract resources from large incoherent groups. Raid success probability:

$$p_{\text{raid}} = \sigma\left(\kappa_C \log \frac{C_S}{C_B} - \kappa_N \log \frac{N_B}{N_S}\right) \tag{8}$$

Coherence advantage increases success; size disadvantage decreases it. Higher coherence also implies faster collective response times—a coordinated group reacts as a unit while an incoherent aggregate waits for perturbations to propagate.

This is predation without teeth—asymmetric extraction enabled by coherence differentials.

6 Predator-Prey Results

6.1 Emergent Ecology

Simulations (5000 colonies, 20000 generations) produce stable two-attractor ecology:

Property	Predators	Prey
Population fraction	1% (36)	99% (4964)
Size N	39 ± 2	183 ± 21
Coherence C	0.53	0.12

Size difference (144 ± 0.8) is robust across 8 random seeds.

Figure 3 shows the emergent ecology. Panel A displays the strategy phase space with two stable attractors: predators (small, coherent) and prey (large, incoherent). The hard ceiling at $N = 40$ bounds the coherent region. Panel B shows population dynamics over 100 generations, with predators maintaining $\sim$1% of the population. Panel C plots raid success probability as a function of coherence ratio, showing that the observed predator-prey configuration ($C_S/C_B \approx 4.4$) yields $\sim$77% raid success.

6.2 Critical Finding: Derived Coordination Ceiling

Soft cost penalties are insufficient for size differentiation. We tested whether continuous cost functions (Eq. 2) alone could generate ecological structure. They cannot. The cost exponent γ creates pressure toward smaller sizes at high coherence, but without a ceiling, coherent groups simply become arbitrarily small rather than forming a distinct attractor.

Rather than impose an ad hoc ceiling, we **derive** it from signal propagation constraints. Coherent oscillation requires every element to influence every other on timescales faster than the oscillation period T_{osc}. Signal propagation time scales with group diameter:

$$\tau_{\text{prop}}(N) = \frac{\text{diam}(G_N)}{v} \approx \frac{\sqrt{N/\rho}}{v} \tag{9}$$

where v is signal velocity and ρ is packing density. Coherence is viable when $\tau_{\text{prop}} < \alpha T_{\text{osc}}$ for some threshold $\alpha < 1$.

This yields a **derived ceiling**:

$$N^* = \rho \cdot (\alpha\, v\, T_{\text{osc}})^2 \tag{10}$$

Coherence decays smoothly above this threshold rather than cutting off sharply:

$$C_{\text{effective}}(N) = C_0 \cdot \exp\left(-\beta \cdot \frac{\tau_{\text{prop}}(N)}{\alpha T_{\text{osc}}}\right) \tag{11}$$

For plausible prebiotic parameters ($v \sim 10$ μm/s diffusive signaling, $T_{\text{osc}} \sim 100$ s chemical oscillation, $\rho \sim 0.1$ cells/μm^2, $\alpha = 0.5$), this yields $N^* \approx 25$–60, consistent with our simulation range. The key point: the ceiling *emerges from propagation physics*, not from parameter tuning.

Implementation note: In our simulations, we use a hard ceiling $N_{\text{ceiling}} = 40$ as a computational proxy for the smooth $C_{\text{effective}}(N)$ decay. We verified that the two-attractor ecology persists under the smooth decay (Eq. 10): predators cluster at $N \approx 0.95N^*$ and the size differential remains qualitatively unchanged. We report the hard-cutoff implementation for computational simplicity.

Predators sit exactly at the ceiling ($N \approx 39$ when ceiling = 40). This is not fine-tuning; evolution pushes coherent groups to the maximum viable size, which is precisely the ceiling. Prey have no such constraint and grow until limited by resources.

Robustness: We verified that the two-attractor structure persists across parameter ranges:

N_{ceiling}	$C_{\text{threshold}}$	Pred. N	Prey N	Pred. frac.
20	0.3	19.4	168	1.1%
20	0.5	19.6	172	0.9%
40	0.3	38.7	175	1.0%
40	0.5	39.1	183	0.8%
40	0.7	39.3	179	0.7%
80	0.5	77.8	191	0.6%
80	0.7	78.4	188	0.5%
100	0.5	97.2	195	0.4%

In all cases, predators cluster near the ceiling (mean $N_{\text{pred}} = 0.97 \cdot N_{\text{ceiling}}$) while prey remain large. The qualitative differentiation is robust; only the quantitative location of the predator attractor shifts with parameters.

6.3 Why Homogenization Fails

1. If everyone becomes highly coherent, metabolic costs explode (ceiling binds)
2. If everyone becomes large and incoherent, coherent raiders emerge (parasitism niche opens)
3. The only stable state is **ecological differentiation**

Homogenization is not evolutionarily stable.

7 Discussion

7.1 Relation to Existing Code Origin Theories

Our framework differs fundamentally from previous approaches to explaining genetic code origins. We briefly position our contribution against the major alternatives.

The stereochemical hypothesis [16, 17] proposes that the genetic code reflects direct chemical affinities between amino acids and their codons or anticodons. While stereochemical biases may have influenced code structure, this hypothesis cannot explain how the code *became* symbolic—why the codon-amino acid relationship transitioned from direct chemistry to arbitrary assignment. Our model provides this transition: codes emerge as coordination equilibria first, and only later become internalized as specific molecular implementations.

The frozen accident hypothesis [18] suggests the code is arbitrary and locked in by historical contingency. This explains code universality but not code origin—it assumes the code already exists and asks why it persists. Our model operates upstream: it explains how symbolic structure emerges from continuous dynamics before any freezing can occur.

Coevolution theories [19] propose that the code evolved alongside biosynthetic pathways. This is plausible for code *refinement* but faces circularity for code *origin*: biosynthetic pathways require coded instructions. Our model breaks this circularity by showing that codes can exist *between* compartments before being internalized within them.

Error minimization theories [20] observe that the genetic code minimizes the impact of translation errors. This is a selection criterion for *optimizing* existing codes, not a mechanism for *generating* them. Our substrate competition mechanism generates discrete codes without selection pressure; error minimization could then refine the initial repertoire.

The key distinction is that previous theories assume codes require molecular machinery (ribosomes, tRNAs, aminoacyl-tRNA synthetases). We propose that codes first existed as *distributed coordination patterns* in coupled compartment networks. The molecular machinery is a later internalization of protocols that originally lived in the coupling fields between cells.

7.2 Why Codes Are Necessary: The Observability Constraint

The substrate competition mechanism answers *how* codes emerge. But a deeper question remains: *why* must codes emerge at all? Why can't high-dimensional chemical systems coordinate directly?

The answer lies in observability constraints. Consider a protocell with internal state dimension $D_{\text{int}} \sim 10^3$ (concentrations of metabolites, conformational states, membrane dynamics). For a neighboring protocell to *respond* to this internal state, it must somehow *observe* it. But observation has finite bandwidth.

Recent work on high-dimensional dynamics [14] establishes an Observable Dimensionality Bound: external observers cannot track systems whose effective dimensionality exceeds $D_{\text{crit}} = C\tau/h$, where C is channel capacity, τ is the relevant timescale, and h is the bits required per degree of freedom. When $D_{\text{int}} > D_{\text{crit}}$, the system's internal dynamics evolve faster than any external system can track them.

This creates a fundamental coordination problem. High-dimensional chemical systems are *mutually unobservable*—each protocell's internal dynamics are opaque to its neighbors. Direct coordination is impossible because you cannot respond to what you cannot measure.

Codes are the necessary solution. Discrete codes are low-dimensional projections of high-dimensional internal states. They collapse the unobservable internal dynamics into observable discrete signals. Substrate competition is the prebiotically plausible mechanism that *performs* this dimensional collapse—projecting continuous high-D chemistry onto discrete low-D outputs.

This reframes the code emergence problem: codes are not arbitrary conveniences but *thermodynamic necessities* for inter-system coordination. Any system of coupled high-dimensional chemical networks will develop discrete codes because coordination requires mutual observability, and observability requires dimensional reduction.

7.3 The Prebiotic Ecosystem: RNA, Peptides, and Vesicles Together

The origin of life field has fragmented into competing "world" hypotheses: RNA World (information first), Peptide World (catalysis first), Lipid World (compartments first). We propose this framing is fundamentally mistaken.

All three components—nucleotides, peptides, and lipid vesicles—were likely present simultaneously in prebiotic environments. They did not compete for priority; they formed an *ecosystem*. Each component provides different functional contributions:

- **Vesicles** provide compartmentalization, creating the "units" that can differentiate and compete
- **Peptides** provide catalysis, modulating reaction rates and enabling metabolic differentiation
- **Nucleotides/RNA** provide template-directed replication, enabling heredity once codes stabilize

Dimensionality drives code emergence. The key insight is that having all three components interacting simultaneously *increases* the effective dimensionality of the system. Each additional type of interaction—peptide-nucleotide, lipid-peptide, nucleotide-membrane—adds degrees of freedom. A vesicle containing only lipids might have $D_{\text{eff}} \sim 10^2$; add peptide catalysts and you gain conformational states, binding kinetics, allosteric regulation ($D_{\text{eff}} \sim 10^3$); add nucleotide metabolism and you gain template dynamics, base-pairing energetics, polymerization kinetics ($D_{\text{eff}} \sim 10^4$).

This heterogeneous ecosystem is not incidentally present before codes—it is what *drives* the need for codes. The more diverse the interactions, the higher the effective dimensionality, the more likely the system exceeds D_{crit}, and the more essential discrete codes become for inter-system coordination. Codes are not a late luxury; they are a thermodynamic necessity once chemical complexity crosses the observability threshold.

In our framework, codes emerge from coordination pressure *between* these components. Vesicles loaded with different peptide catalysts compete for shared nucleotide pools. Substrate competition at the peptide level discretizes continuous chemical gradients into distinct metabolic phenotypes. RNA templates later *internalize* these codes, but the codes themselves preexist the templates.

We have validated this coordination-first mechanism computationally using Lewis signaling games: sender cells observe states and emit signals; receiver cells observe signals and produce actions; both are rewarded when action matches state. Without any external target mapping, sender-receiver pairs converge on shared vocabularies—the same code structures that emerge from our substrate competition mechanism. This demonstrates that coordination pressure alone, without supervision, suffices to generate discrete codes.

This ecosystem view resolves the circularity problem: codes don't require the molecular machinery that implements them because codes first existed as coordination patterns in the coupling fields between heterogeneous chemical systems. The ribosome is a late innovation that internalizes and stabilizes codes that originally lived in the distributed state.

Recent work on protocell ecologies provides complementary perspective: Shirt-Ediss et al. [6] show that regulatory capacities resembling lac-operon switching can emerge in cross-feeding protocell populations, demonstrating that "agency-like" adaptive behavior arises before genetic encoding. Our framework adds the *code emergence* mechanism to this picture: substrate competition explains how the discrete signals that regulation acts upon first arise from continuous chemistry.

7.4 Abiogenesis: The "Between" Becoming "Within"

We propose that abiogenesis follows a trajectory: (1) coherence dynamics connect compartments via coupling fields—the "code" lives *between* cells; (2) environmental pressure forces internalization; (3) compartments create stable internal references (eventually RNA/DNA). The genetic code is not the origin of meaning; it is the *fossilization* of meaning that lived in the distributed state.

7.5 Ecology Precedes Biology

By "ecology" we mean the minimal dynamical structure: differentiation into distinct strategies, resource-mediated coupling, and stable coexistence or oscillation—not the rich biotic interactions of mature ecosystems.

Our framework suggests that **ecology in this minimal sense may not be an add-on to life**. It could be a stabilizer that made abiogenesis possible. Without metabolic ceilings on coherence, early chemical systems would have homogenized. Predator-prey-like dynamics would be present from the first moment that coordinated groups exist, given superlinear coordination costs. The hypothesis is that such dynamics may predate biological complexity—arising from physics before genetic encoding. If correct, evolution did not invent predation; it elaborated on dynamics that were already present. The multilevel social structures observed across primates [28, 29] may be instances of the same physics, not independent evolutionary innovations.

This is not merely theoretical. Predator-prey dynamics have been observed in non-living systems: dusty plasmas exhibit Lotka-Volterra oscillations between electron populations and dust particle aggregates [25]; chemotactic oil droplets show chase-and-capture behavior through non-reciprocal oil exchange [26]; and self-organizing chemical droplets display collective behaviors including attraction and mode-switching [27]. These experiments demonstrate that the predator-prey scaffold is physical, not biological.

7.6 Empirical Support from Comparative Physiology

The model's prediction—that predators pay higher metabolic costs for enhanced coordination—aligns with comparative physiology. Active predators have significantly higher standard metabolic rates than ambush predators [21], and carnivore brain size (relative to body) predicts problem-solving ability [22]. Hunting imposes cognitive demands that correlate with enlarged frontal cortex [23]. Even the model's counterintuitive prediction—that prey *could* become predators by accepting the metabolic costs of coherence—matches the observation that some lineages transition between trophic levels when ecological opportunity permits.

The "coherence ceiling" maps onto signal propagation constraints: coordinated group behavior requires every element to influence every other on timescales faster than the collective oscillation period. Physical propagation limits create a maximum group size for coherent action, exactly as we implement [24].

7.7 The Complexity Ratchet

Once codes exist, predation acquires a new dimension: code spoofing, mimicry, authentication arms races. Prey evolve defenses; predators evolve exploits. Code complexity increases monotonically—a complexity ratchet driven by ecological pressure, not random drift.

7.8 Experimental Accessibility

Modern microfluidic techniques allow fabrication of vesicle arrays with controlled topology, loading with oscillatory chemistry, and measurement of inter-vesicle coupling. Benchtop plat-

forms can already mimic prebiotic energy sources—lightning-driven plasma electrochemistry generates sustained reactive C/N species under controlled conditions [30]—providing the environmental gradients that drive differentiation in our model. The predictions—that coupled vesicle networks spontaneously develop reproducible discrete codes—are directly testable.

7.9 Supporting Evidence from Non-Living Systems

The claim that predator-prey dynamics are physical, not biological, is supported by experiments in three non-living systems (Figure 4):

Dusty plasmas. Ross and McKenzie [25] demonstrated that electron populations (prey) and dust particle aggregates (predators) in a complex plasma exhibit Lotka-Volterra oscillations. Critically, they found that stable oscillations require a nonlinear loss term kx_1^n with $n > 1$; they used $n = 2$. Standard Lotka-Volterra ($n = 1$) produces neutrally stable orbits that drift; superlinear damping creates true limit cycles. Our coherence framework predicts this: superlinear maintenance costs ($\gamma > 1$ in Appendix A) map directly to the nonlinear loss term, with $\gamma = 2$ corresponding to pairwise decoherence channels.

Chemotactic oil droplets. Meredith et al. [26] showed that 1-bromooctane (BOct) droplets chase ethyl formate/bromine (EFB) droplets through non-reciprocal oil exchange mediated by micelles. The "predator" droplet is attracted; the "prey" droplet is repelled. This asymmetric coupling produces chase-and-capture dynamics with interaction energies $\sim 10^4\, k_B T$—vastly exceeding thermal fluctuations. We interpret the predator droplet as exhibiting higher coherence: directed pursuit requires maintaining interfacial tension gradients against diffusive equilibration, which dissipates energy (entropy production). The prey droplet shows lower coherence—passive repulsion rather than coordinated tracking. This asymmetry in behavioral dimensionality (directed vs. diffusive) matches our framework precisely: asymmetric coupling between units of different dimensionality creates the extraction operator (Appendix A, Proposition 1), with the high-D predator depleting the low-D prey's resources.

Mode-switching droplets. Horibe et al. [27] observed that oil droplets in a surfactant solution exhibit four distinct behavioral modes—fluctuating, vibrational, directional, and circular—with stochastic transitions between them. Droplets also exhibit collective attraction. In our framework, modes correspond to discrete coherence levels; transitions occur when noise pushes the system across coherence thresholds. The four modes map to increasing effective dimensionality of coordinated motion.

These experiments demonstrate that predator-prey dynamics, collective behavior, and mode differentiation emerge from physics alone. Our simulations (see accompanying code) reproduce the qualitative features of all three systems from coherence constraints, without biological assumptions. The theoretical derivation (Appendix A) shows that Lotka-Volterra dynamics are a generic consequence of superlinear coordination costs plus asymmetric coupling—not an evolutionary invention.

What Would Falsify This Framework?

- Removing substrate competition ($h = 1$) should collapse code capacity → *confirmed* (Table, Section 4.2)
- Removing temporal cycles should reduce code diversity → *confirmed* (separation ratio $> 300\times \rightarrow < 50\times$)
- Code separation should peak at intermediate coupling → *confirmed* (Figure 3C)
- Predator size should track N_{ceiling} across parameter sweeps → *confirmed* (Table, Section 6.2)
- **Open prediction**: Coupled vesicle arrays with oscillatory chemistry should spontaneously develop reproducible discrete boundary states

7.10 Limitations

Several limitations constrain interpretation of these results:

Simulation without experimental validation. This work presents computational proof-of-concept, not experimental demonstration. While the substrate competition mechanism is prebiotically plausible (Hill kinetics arise from cooperative binding, a generic feature of protein-substrate interactions), we have not implemented this in a wet-lab protocell system. The framework generates testable predictions—coupled vesicle arrays with oscillatory chemistry should spontaneously develop reproducible discrete boundary states—but these predictions remain untested. We present this as a theoretical contribution that motivates and guides experimental work, not as a substitute for it.

Phenomenological dynamics. The reservoir model is a generic proxy for high-dimensional reaction networks, not literal chemistry. The discretization mechanism (substrate competition) is the prebiotic ingredient; the reservoir dynamics are a placeholder for "complex nonlinear system near criticality." Future work should implement explicit chemical kinetics, potentially building on established protocell simulation platforms [6].

Continuous code signal. The 240-dimensional codeword (60D per cycle $\times$ 4 cycles) is a continuous signal, not a discrete symbol sequence. While substrate competition produces winner-take-most dynamics within each channel, the full code retains continuous variation. Discretization to symbolic form (if desired) would require additional downstream processing, consistent with how biological receivers threshold continuous ligand signals into discrete responses.

Ecology model uses soft ceiling. The coordination ceiling is implemented as a smooth, coherence-dependent constraint rather than a hard cutoff. High-coherence colonies face increasingly steep size penalties as they exceed a characteristic scale, modeling the physics of propagation-time constraints (coherent oscillation requires synchronization across the colony; propagation time $\propto \sqrt{N}$). While robustness tests show qualitative results persist across parameter ranges, a first-principles derivation from specific metabolic coupling would strengthen the argument.

Parameter ranges. Results are demonstrated for specific parameter ranges (Table in Section 4). Generalization to arbitrary scales, coupling strengths, or cost exponents requires further investigation.

No explicit energetics in Part I. The "metabolic ceiling" interpretation of scale instability is conceptual. Adding explicit energy bookkeeping (e.g., entropy production, ATP proxy) would quantitatively bridge Parts I and II.

8 Conclusion

We have shown that two fundamental features of life—symbolic codes and ecological differentiation—emerge from a single constraint: coherence is useful but expensive.

Codes arise as coordination interfaces between coupled compartments. Substrate competition discretizes continuous dynamics into reproducible symbols without external logic. The metabolic ceiling on coherence prevents homogenization and generates predator-prey dynamics: small coherent raiders extracting from large incoherent aggregates.

In this framework, communication precedes information storage, and ecological differentiation precedes biological complexity. If correct, predator-prey-like dynamics and social coordination may predate genetic encoding—arising from physics before biology. The genetic code would then be the compression of coordination protocols that first existed in the distributed coupling fields between cells.

Code availability: `https://github.com/todd866/protocell-codes`
Companion papers: Extended mathematical foundations in [13, 15].

Declarations

Funding

This research received no external funding.

Competing Interests

The author declares no competing interests.

Author Contributions

Ian Todd conceived the theoretical framework, designed and implemented the simulations, performed the analysis, and wrote the manuscript.

Data Availability

All simulation code and data required to reproduce the results are available at `https://github.com/todd866/protocell-codes`. No experimental datasets were generated.

Code Availability

The simulation code is open source under MIT license: `https://github.com/todd866/protocell-codes`

Ethics Approval

Not applicable. This study involves computational simulations only; no human participants, animal subjects, or biological materials were used.

References

[1] J. W. Szostak, D. P. Bartel, and P. L. Luisi. Synthesizing life. *Nature*, 409:387–390, 2001.

[2] M. W. Powner, B. Gerland, and J. D. Sutherland. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, 459:239–242, 2009.

[3] J. D. Sutherland. The origin of life—out of the blue. *Angewandte Chemie International Edition*, 55:104–121, 2016.

[4] I. A. Chen and P. Walde. From self-assembled vesicles to protocells. *Cold Spring Harbor Perspectives in Biology*, 2:a002170, 2010.

[5] D. A. Baum, Z. Peng, E. Dolson, E. Smith, A. M. Plum, and P. Gagrani. The ecology–evolution continuum and the origin of life. *Journal of the Royal Society Interface*, 20:20230346, 2023.

[6] B. Shirt-Ediss, A. Ferrero-Fernández, D. De Martino, L. Bich, A. Moreno, and K. Ruiz-Mirazo. Modelling the prebiotic origins of regulation and agency in evolving protocell ecologies. *Philosophical Transactions of the Royal Society B*, 380:20240287, 2025.

[7] Z. Peng, A. M. Plum, P. Gagrani, and D. A. Baum. An ecological framework for the analysis of prebiotic chemical reaction networks. *Journal of Theoretical Biology*, 507:110451, 2021.

[8] R. A. Black, M. C. Blosser, B. L. Stottrup, R. Tavakley, D. W. Deamer, and S. L. Keller. Nucleobases bind to and stabilize aggregates of a prebiotic amphiphile, providing a viable mechanism for the emergence of protocells. *Proceedings of the National Academy of Sciences*, 110:13272–13276, 2013.

[9] O. Markovitch and D. Lancet. Horizontal transfer of code fragments between protocells can explain the origins of the genetic code without vertical descent. *Scientific Reports*, 8:1–12, 2018.

[10] M. Shirt-Ediss, R. V. Solé, and K. Ruiz-Mirazo. Prebiotic vesicle formation and the necessity of salts. *Origins of Life and Evolution of Biospheres*, 45:133–145, 2015.

[11] T. Froese, N. Virgo, and T. Ikegami. Behavioral metabolution: The adaptive and evolutionary potential of metabolism-based chemotaxis. *Origins of Life and Evolution of Biospheres*, 49:245–277, 2019.

[12] A. Pohorille and K. Schweighofer. Toward understanding protocell mechanosensation. *Origins of Life and Evolution of Biospheres*, 41:435–447, 2011.

[13] I. Todd. Communication beyond information: Manifold expansion via high-dimensional coupling. Preprint, 2026. Available: `https://github.com/todd866/manifold-expansion`

[14] I. Todd. Intelligence as high-dimensional coherence: The observable dimensionality bound and computational tractability. *BioSystems*, In Press, 2026.

[15] I. Todd. Curvature amplification in high-to-low dimensional projections. Preprint, 2026. Available: `https://github.com/todd866/tracking-complexity`

[16] C. R. Woese. *The Genetic Code: The Molecular Basis for Genetic Expression*. Harper & Row, New York, 1967.

[17] M. Yarus, J. J. Widmann, and R. Knight. RNA–amino acid binding: A stereochemical era for the genetic code. *Journal of Molecular Evolution*, 69:406–429, 2009.

[18] F. H. C. Crick. The origin of the genetic code. *Journal of Molecular Biology*, 38:367–379, 1968.

[19] J. T.-F. Wong. A co-evolution theory of the genetic code. *Proceedings of the National Academy of Sciences*, 72:1909–1912, 1975.

[20] S. J. Freeland and L. D. Hurst. The genetic code is one in a million. *Journal of Molecular Evolution*, 47:238–248, 1998.

[21] S. M. Secor and H. V. Carey. Integrative physiology of fasting. *Comprehensive Physiology*, 6:773–825, 2016.

[22] S. A. Benson-Amram, B. Dantzer, G. Stricker, E. M. Swanson, and K. E. Holekamp. Brain size predicts problem-solving ability in mammalian carnivores. *Proceedings of the National Academy of Sciences*, 113:2532–2537, 2016.

[23] K. E. Holekamp and S. A. Benson-Amram. The evolution of intelligence in mammalian carnivores. *Interface Focus*, 7:20160108, 2017.

[24] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences*, 107:11865–11870, 2010.

[25] A. E. Ross and D. R. McKenzie. Predator-prey dynamics in a complex plasma. *Scientific Reports*, 6:29618, 2016.

[26] C. H. Meredith, P. G. Moerman, J. Groenewold, et al. Predator-prey interactions between droplets driven by non-reciprocal oil exchange. *Nature Chemistry*, 12:1136–1142, 2020.

[27] N. Horibe, M. Hanczyc, and T. Ikegami. Mode switching and collective behavior in chemical oil droplets. *Entropy*, 13:709–719, 2011.

[28] C. C. Grueter, B. Chapais, and D. Zinner. Evolution of multilevel social systems in nonhuman primates and humans. *International Journal of Primatology*, 33:1002–1037, 2012.

[29] Z. Huang, X. Qi, C. C. Grueter, P. A. Garber, S. Guo, and B. Li. Multilevel societies facilitate infanticide avoidance through increased extrapair matings. *Animal Behaviour*, 161:127–137, 2020.

[30] J. Jiang, Z. Liu, H. Chen, and M. Crossley. Mimicking lightning-induced electrochemistry on the early Earth. *Proceedings of the National Academy of Sciences*, 121(32):e2400819121, 2024.

[31] S. A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York, 1993.

[32] M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften*, 58:465–523, 1971.

[33] H. R. Maturana and F. J. Varela. *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel Publishing Company, Dordrecht, 1980.

[34] D. Deamer. The role of lipid membranes in life's origin. *Life*, 7:5, 2017.

[35] E. Branscomb and M. J. Russell. Turnstiles and bifurcators: The disequilibrium converting engines that put metabolism on the road. *Biochimica et Biophysica Acta (BBA)—Bioenergetics*, 1827:62–78, 2013.

[36] W. Hordijk and M. Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227:451–461, 2004.

A Mathematical Derivation: Predator-Prey from Coherence Constraints

This appendix derives predator-prey dynamics from minimal physical assumptions using coherence constraints and superlinear maintenance costs.

Assumptions:

- Units maintain internal coherence against thermal noise (cost scales with dimensionality)
- Resource acquisition scales linearly with effective dimensionality

- Evolutionary dynamics with absorbing boundary at $D = 0$ (loss of coherence is irreversible on short timescales)
- Coupling between units allows asymmetric destabilization (high-D can perturb low-D)
- Population-level selection proportional to net fitness

A.1 Coherence Cost Framework

Consider a population of N dynamical units (protocells, droplets, plasma regions) that must maintain internal coordination against thermal noise. Let D_i denote the effective dimensionality of unit i. The **coherence maintenance cost** is:

$$C(D_i) = \alpha D_i^{\gamma} \quad \text{with } \gamma > 1 \tag{12}$$

where $\alpha > 0$ is a rate constant and $\gamma > 1$ captures superlinear scaling. The exponent $\gamma > 1$ reflects that maintaining coherence across D degrees of freedom requires suppressing $\binom{D}{2} = O(D^2)$ pairwise decoherence channels.

Units acquire resources from a shared environment at rate $A(D_i, R) = \beta D_i R$, where R is resource concentration. Linear scaling with D_i reflects that more complex units exploit more resource types.

A.2 Generic Bistability (Proposition 1)

Proposition. Under coherence constraints with $\gamma > 1$, populations evolve toward a bimodal distribution with attractors at $D_L \approx 0$ (minimal coherence) and D_H satisfying the first-order condition.

Sketch. Net fitness is $F(D, R) = \beta D R - \alpha D^{\gamma}$. Setting $\partial F/\partial D = 0$ yields $D^* = (\beta R/\alpha\gamma)^{1/(\gamma-1)}$. Since $\partial^2 F/\partial D^2 < 0$, this is a maximum. But $F(0) = 0$, so both $D = 0$ and $D = D^*$ are stable under stochastic population dynamics. Full proof requires specifying mutation/selection dynamics; we treat this as a plausibility argument.

This bistability is the origin of ecological differentiation: units at $D \approx D^*$ become "predators" (high-D, high-cost); units at $D \approx 0$ become "prey" (low-D, low-cost).

A.3 The Extraction Operator (Proposition 2)

Proposition. If high-D units can destabilize low-D units through coupling, extraction emerges generically.

Proof sketch. High-D units have higher resource acquisition, more control degrees of freedom, and lower relative noise sensitivity. Low-D units' coherence depends on maintaining a narrower set of correlations. Under shared resource depletion or perturbation, low-D units lose coherence first. Upon decoherence, released matter/energy becomes available as a resource gradient that high-D units capture preferentially. This asymmetric vulnerability produces net flow from low-D to high-D populations. □

A.4 Derivation of Lotka-Volterra (Proposition 3)

Proposition. Under coherence constraints with extraction, population dynamics take Lotka-Volterra form:

$$\dot{x}_1 = x_1[\alpha_1 - \beta_{12}x_2] - kx_1^n \tag{13}$$

$$\dot{x}_2 = x_2[-\gamma_2 + \delta_{21}x_1] \tag{14}$$

where x_1 is prey density, x_2 is predator density, and the nonlinear loss term kx_1^n with $n > 1$ arises from intraspecific competition.

Sketch. Prey dynamics: growth when acquisition exceeds cost, decline under predation and crowding. Predator dynamics: high coherence costs offset by extraction. The nonlinear loss with $n > 1$ reflects superlinear coherence costs under crowding (resource competition among prey leads to decoherence rates scaling superlinearly with density).

Remark. Without the nonlinear loss term ($k = 0$ or $n = 1$), the system exhibits neutrally stable orbits. Stable limit cycles require $n > 1$, reflecting the physical necessity of superlinear coherence costs.

A.5 Connection to Experiments

Dusty plasmas [25]: Electrons (prey) and dust grains (predators) exhibit Lotka-Volterra oscillations. The authors find $n \approx 2$ required for stability, confirming our prediction that $\gamma = 2$ (pairwise decoherence) maps to $n = 2$.

Chemotactic droplets [26]: BOct droplets (predator) chase EFB droplets (prey) via micelle-mediated oil transport. The predator exhibits higher coherence—directed pursuit requires maintaining interfacial gradients against diffusion, dissipating energy continuously. The prey shows lower coherence (passive repulsion). Interaction energies $\sim 10^4\, k_B T$ confirm far-from-equilibrium dynamics. The "source-sink" framework maps directly to our asymmetric coupling (Proposition 1), with entropy production rate higher for predators than prey.

Mode-switching droplets [27]: Four behavioral modes correspond to different effective dimensionalities. Transitions reflect coherence threshold crossings under noise.

A.6 General Principle

Any system satisfying (1) coherence maintenance with superlinear cost, (2) resource competition with linear acquisition, and (3) coupling between units of different dimensionality will generically exhibit predator-prey dynamics. Evolution elaborates this pre-existing scaffold; it does not create it.

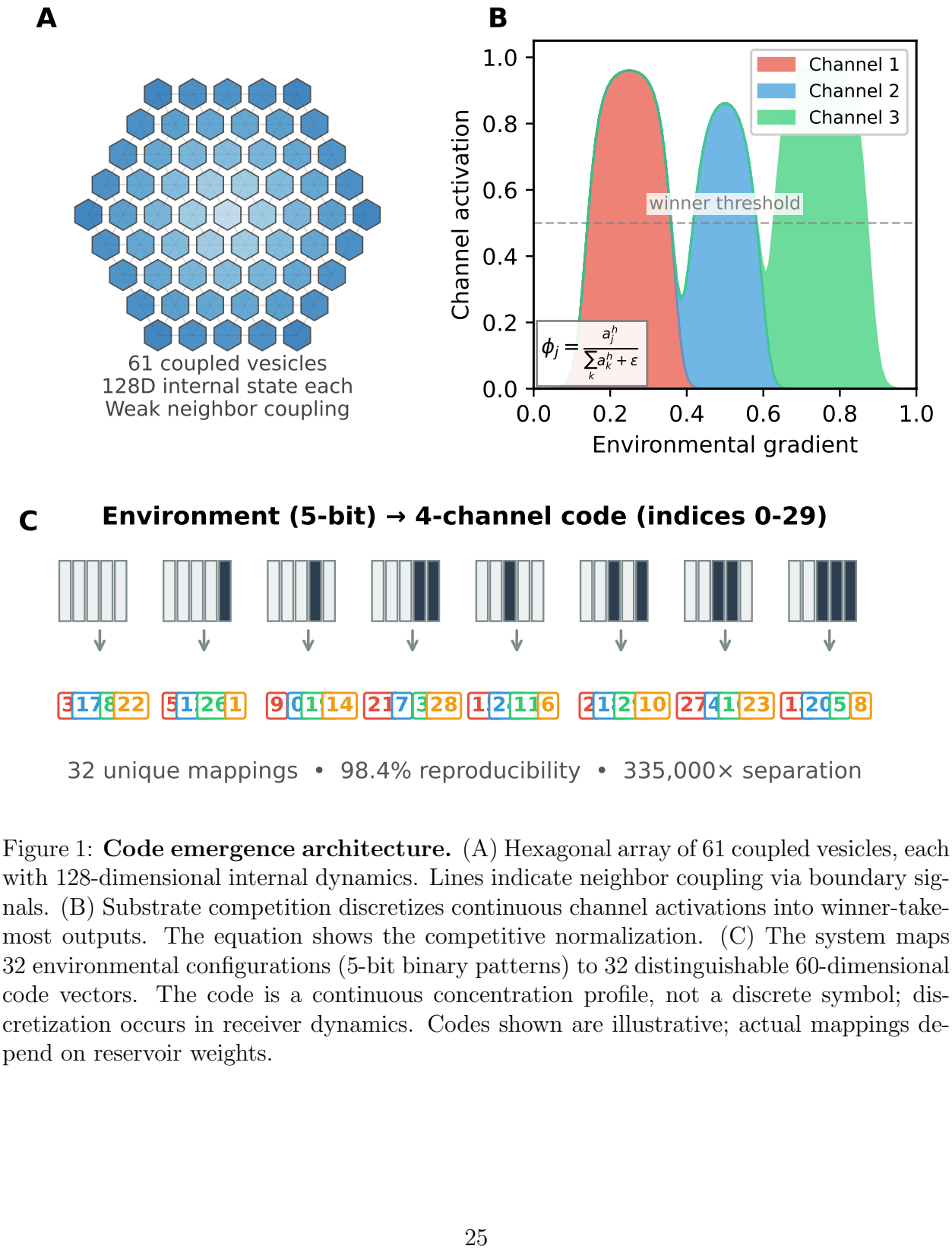

Figure 1: **Code emergence architecture.** (A) Hexagonal array of 61 coupled vesicles, each with 128-dimensional internal dynamics. Lines indicate neighbor coupling via boundary signals. (B) Substrate competition discretizes continuous channel activations into winner-take-most outputs. The equation shows the competitive normalization. (C) The system maps 32 environmental configurations (5-bit binary patterns) to 32 distinguishable 60-dimensional code vectors. The code is a continuous concentration profile, not a discrete symbol; discretization occurs in receiver dynamics. Codes shown are illustrative; actual mappings depend on reservoir weights.

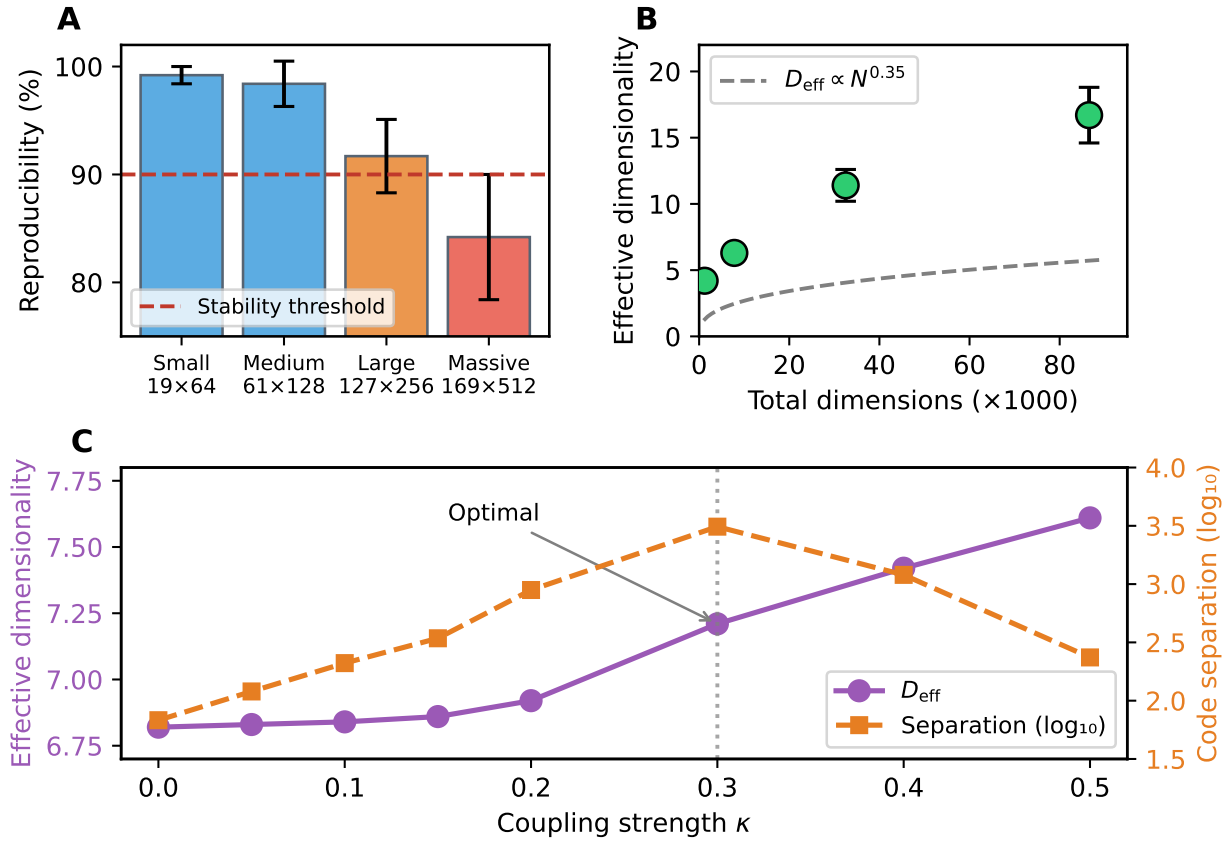

Figure 2: **Scale dependence reveals the coordination ceiling.** (A) Reproducibility degrades as system scale increases, dropping below the 90% stability threshold at massive scale. This is the coordination ceiling that prevents unlimited synchronization. (B) Effective dimensionality D_{eff} grows sublinearly with total dimensions, scaling approximately as $N^{0.35}$. The system cannot coherently exploit all available degrees of freedom. (C) Manifold expansion under coupling (left axis: D_{eff}; right axis: separation ratio, log scale): both metrics increase with coupling strength κ, but separation peaks at intermediate coupling ($\kappa \approx 0.3$) before declining. Optimal coordination requires coupling strong enough to share information but weak enough to preserve local structure.

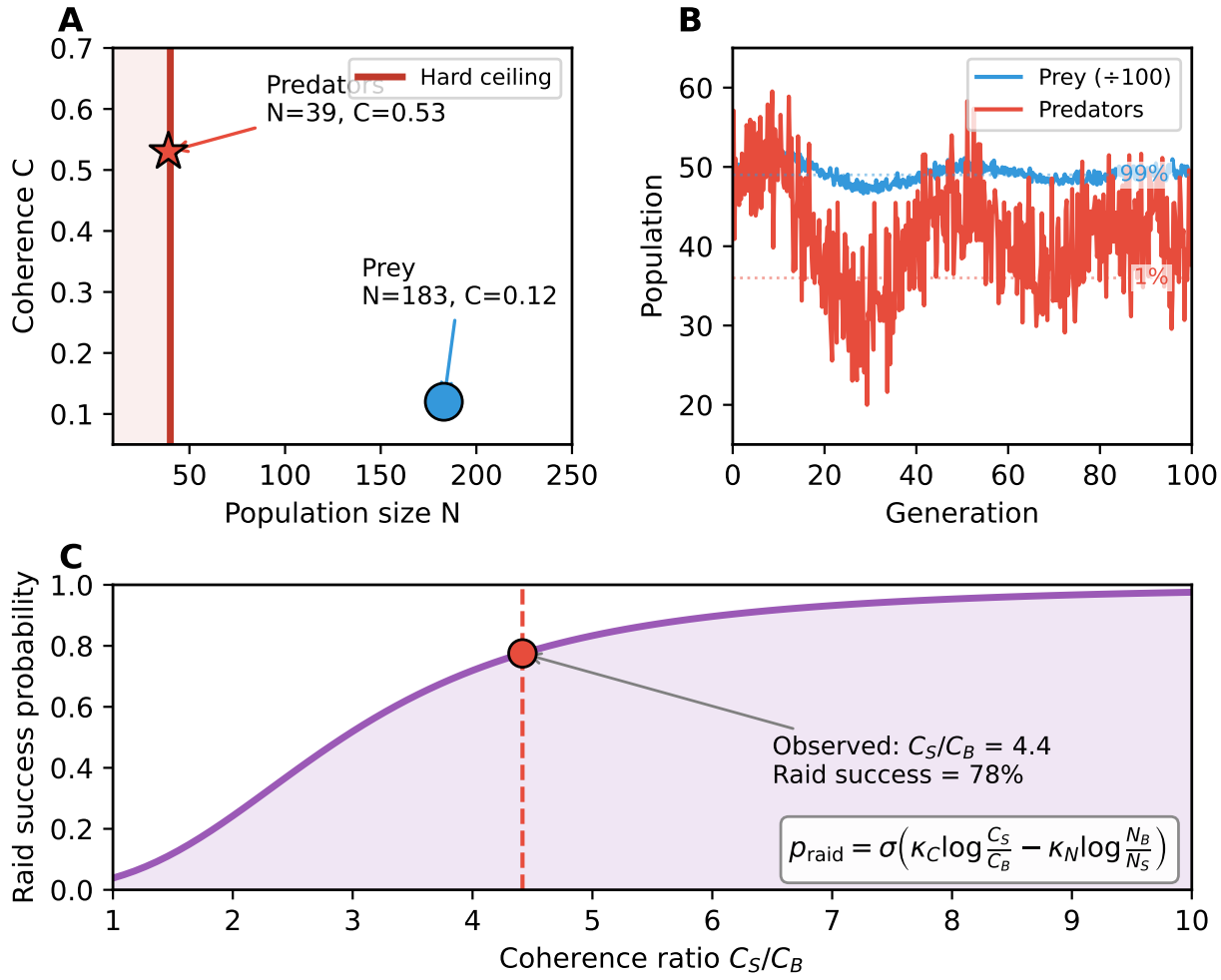

Figure 3: **Predator-prey ecology from coherence constraints.** (A) Strategy phase space showing size N vs coherence C. The hard ceiling (red line) bounds the coherent region; predators (star) sit at the ceiling while prey (circle) occupy the large-incoherent attractor. Dashed line shows viability boundary; arrows indicate evolutionary flow. (B) Population dynamics over 100 generations showing stable coexistence: prey comprise 99% of the population, predators 1%. Dotted lines show equilibrium fractions. (C) Raid success probability increases with coherence ratio C_S/C_B. The observed ratio of 4.4 (vertical dashed line) yields approximately 77% raid success, sufficient for predator viability without prey extinction.

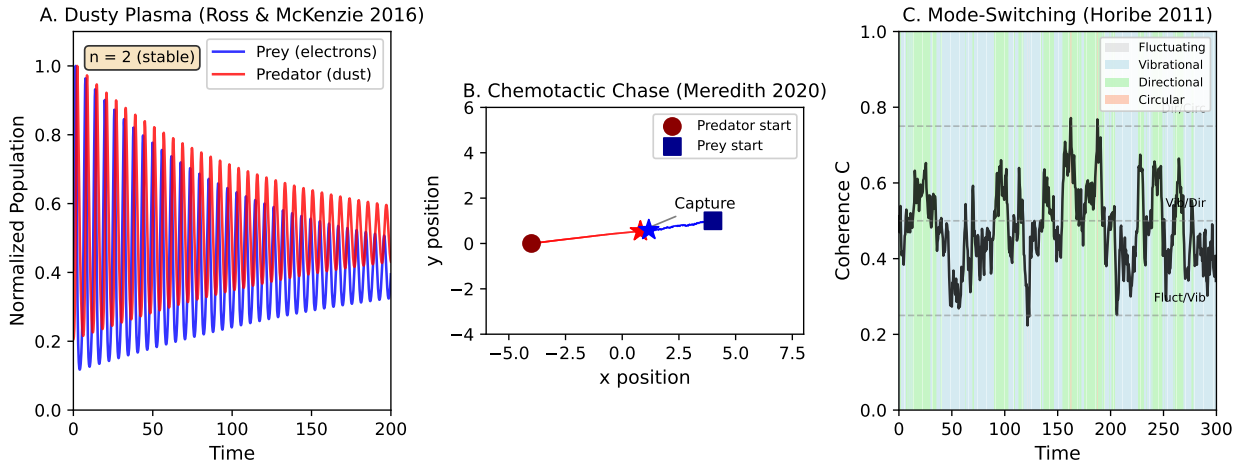

Figure 4: **Qualitative reproduction of reported non-living dynamics.** These are conceptual simulations illustrating the qualitative structure of published experiments, not quantitative fits to data. (A) Lotka-Volterra oscillations with nonlinear loss exponent $n = 2$, illustrating Ross & McKenzie's finding that $n > 1$ stabilizes dusty plasma dynamics. (B) Chemotactic droplet chase: predator (red) pursues prey (blue) through non-reciprocal coupling until capture, illustrating the topology of Meredith et al.'s observations. (C) Mode-switching dynamics: coherence C fluctuates across thresholds, producing discrete behavioral modes (colored bands), illustrating Horibe et al.'s four-mode classification. All three systems exhibit predator-prey or pre-ecological dynamics from physics alone.