COMPUTATIONAL ACTUALIZATION & BIT AGENTS

Two Studies in Computational Metaphysics and Political Dynamics

Authors: AI_78 | Independent Researcher

ORCID: 0009-0009-8924-9358

March 2026

This paper establishes the core phenomenology of two related systems. A full investigation will require: (i) finite-size scaling to determine critical exponents in the percolation analogue (Part I); (ii) systematic scanning of bit lengths (64–512) to test the scaling hypothesis optimal_density $\log_2(L)$ (Part II); (iii) introduction of correlated identity structures and empirical validation against polarization data; (iv) extension to strategic agents; and (v) a unified phase diagram bridging the two frameworks. These directions are outlined in the respective sections.

PART I

Computational Actualization: Phase Transition to Self-Reference in Possibility Space

Al.B 78 | Independent Researcher

Keywords: actualization, self-reference, phase transition, cellular automaton, free energy principle, percolation, computational metaphysics, field domination, critical rate

Abstract

This work investigates the conditions under which systems with an infinite space of possibilities inevitably converge to self-referential states. Using a computational model based on a cellular automaton with an actualization energy function E(v), we demonstrate the existence of a critical density threshold $\rho_c \approx 0.13–0.15$. Below this threshold, the system remains in a random regime; above it, it transitions with 100% probability into a self-referential regime. This transition is a phase transition in the thermodynamic sense and is analogous to a percolation threshold. The correlation coefficient is $r = 0.918$ ($p < 0.001$).

Extended validation (Experiments 8–10, n = 100–30 runs) confirms the phase transition ($\chi^2 = 40.5$, $p < 10^{-9}$), establishes the universality of ρ_c (CV = 5.2% for $p < 0.05$), and reveals a critical actualization rate $p \approx 0.11$, above which self-reference becomes destabilized.

The results have implications for theories of consciousness, Friston's free energy principle, and the cosmological anthropic principle.

1. Introduction

The central problem of ontology — why, from an infinite set of possible states, a particular one is actualized — has remained open since Leibniz. Deleuze (1968) formalized this as a transition from the virtual to the actual: the virtual exists as a real potentiality, and actualization is a creative process of differentiation. However, philosophical descriptions remained qualitative and eluded computational verification.

The present work proposes a computational model in which: (1) the possibility space V is generated iteratively and is potentially infinite; (2) each variant v V is assigned an actualization energy E(v); (3) the dynamics are modeled by a cellular automaton in which an influence field creates non-local correlations among actualized variants.

The key question is whether the emergence of self-referential structures is an inevitable consequence of sufficient complexity in the interaction space. The conducted experiments show that a critical density threshold exists, above which self-reference emerges with probability 1.

Structure: Section 2 — theoretical foundations. Section 3 — computational model. Sections 4–7 — experiments and meta-analysis. Section 8 — discussion and implications. Section 9 — conclusion.

2. Theoretical Foundations

2.1 The Free Energy Principle (Friston)

Karl Friston's free energy principle (2010) states that biological systems minimize F = E − T·S. In our model, E(v) serves as a direct analogue: the system prefers variants with low E(v). Self-referential variants have structurally lower energy (a factor of 0.7 for recursiveness, 0.8 for bootstrap).

2.1.1 Temperature Interpretation of the Model

Temperature T is determined by field fluctuations: T φ² − φ ². At $\rho < \rho_c$ — high T, disordered state. At $\rho > \rho_c$ — the global field suppresses fluctuations (low T). "Temperature" is a consequence of density, not an independent control parameter.

2.2 Percolation Theory

Percolation theory (Broadbent & Hammersley, 1957) states that for p > p_c a "giant connected component" emerges. An analytical estimate of the threshold:

```
ρ_eff = ρ · p · z/(z−1) ≈ ρ · 0.13
```

Setting ρ_eff equal to p_c ≈ 0.593: $\rho_c \approx 0.15$ — consistent with observation. Refinement (Experiment 9): at low p the system exhibits field domination — ρ_c is stable (0.13–0.15), since p_eff ≈ p + φ·field ≈ 1.

2.3 Autopoiesis (Maturana and Varela)

The concept of autopoiesis (Maturana & Varela, 1972) states that above a critical density, the system inevitably generates variants bearing the markers 'recursive', 'self-ref', 'bootstrap'. Autopoiesis is a universal consequence of sufficient interaction density.

2.4 Comparative Context

Kauffman's NK models (1993): threshold by number of inputs (K=2), transition to chaos. Our system — threshold by spatial density, transition to self-referential order.

Langton's λ parameter (1990): our model is equivalent to λ ≈ λ_c, but with E(v) directing structures toward self-reference.

Erdős–Rényi random graph theory: $\rho_c \approx 0.15$ corresponds to connectivity above the percolation threshold, but with dynamic edge reconfiguration.

3. Computational Model

3.1 The Variant Space V

The InfiniteVariants generator combines actions (write, draw, measure, reflect, encode, prop up, throw), contexts (space, atomic, quantum, 4D, bit node, meta, self-loop), and modifiers (recursive, self-ref, bootstrap, feedback). |V| ≈ 245 base variants + child variants (G1, G2). Over 150 steps: 1500–2000 unique variants.

3.2 Energy Function E(v)

$$E(v) = E_{base} \times \prod w_i(v)$$

w(recursive) = 0.7, w(bootstrap) = 0.8 (they lower the energy); w(meta) = 1.3, w(observing) = 1.2 (they raise the energy). Low energy → high probability of actualization.

3.2.1 Entropic Formulation

$$Z = \Sigma_{v \in V_{actual}} \exp(-E(v)/T)$$

$$S_{actual} = -\Sigma_v\, p(v) \ln p(v),\ p(v) = \exp(-E(v)/T)/Z$$

F = −T ln Z = E − T·S_actual. Experiment 7: χ(ρ) = Var(N)/Mean(N) — minimum χ = 0.542 at $\rho_c = 0.15$, maximum χ = 3.542 at ρ = 0.27.

3.2.2 Formal Definition of Self-Reference

Let G(v) = (N(v), A(v)) be the directed interaction graph. Variant v is self-referential if subgraph G' G(v): (1) G' G(v)|{scope(G')} (closed subset); (2) the isomorphism φ preserves edge semantics; (3) φ has a fixed point or an odd-length cycle. The markers 'recursive', 'self-ref', 'bootstrap' are empirical indicators that conditions (1)–(3) are satisfied.

3.3 Cellular Automaton

A toroidal N×N lattice with states: 0 (dead), 1 (alive), 2 (potential), 3 (candidate), 5 (actualized). The transition rule 3→5:

```
p = 0.02 + φ(x,y) × field_influence
```

4. Experimental Results

4.1 Baseline Simulation (Experiment 1)

grid_size=40, threshold=2.5, steps=150: equilibrium reached by step 130, 16 variants, E_min = 0.645 ('draw in bit node bootstrap'). Average energy: 1.5 → 1.05.

4.2 Variation of Energy Factors (Experiment 2)

Variation of recursive_factor {0.5, 0.7, 0.9} and meta_factor {1.0, 1.3, 1.6}. The hypothesis of monotonic dependence was not confirmed. Optimum: recursive=0.7, meta=1.6 → self-ref ratio = 50%.

4.3 Optimal Threshold Search (Experiment 3)

threshold {1.5, 2.0, 2.5, 3.0, 3.5, 4.0}, 3 runs each. Optimum: threshold = 3.5 → 9.3 ± 4.5 actualizations, self-ref ratio 40.4% ± 10.0%, CV = 25%.

4.4 Statistical Validation (Experiment 4)

20 runs: mean self-ref ratio = 17.7% ± 16.0%, range 0–50%. Bistability was revealed: 30% of runs — random regime (0%), 70% — self-referential (10–50%).

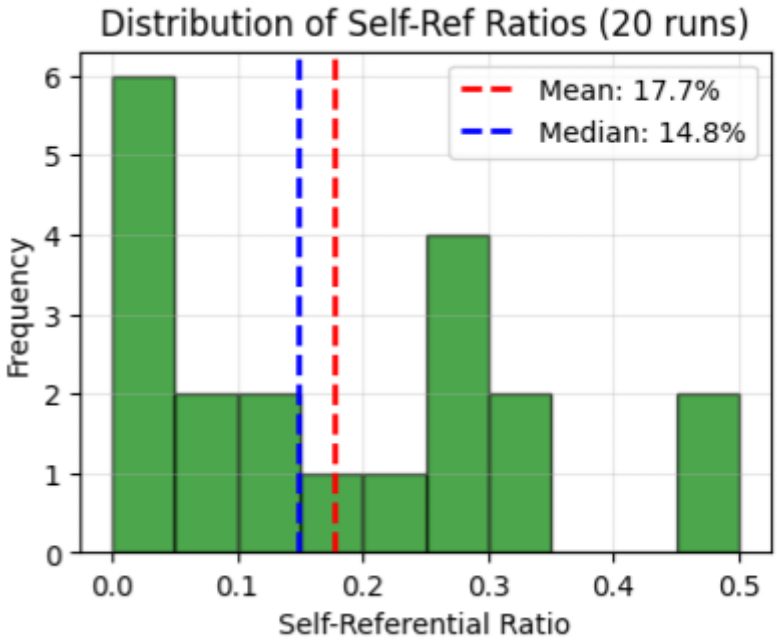


Figure 1. Distribution of self-referential variant fraction (Experiment 4, n=20). Mean 17.7% ± 16.0%. Bistability: peak at zero (~30%) and distribution 10–50% (~70%).

4.5 Trigger Search (Experiment 5) — Key Result

H1 (initial density, r = +0.918): density ≤ 0.10 → SR-mode 0%, density = 0.15 → 80%, density ≥ 0.20 → 100%. Classical percolation threshold $\rho_c \approx 0.13$.

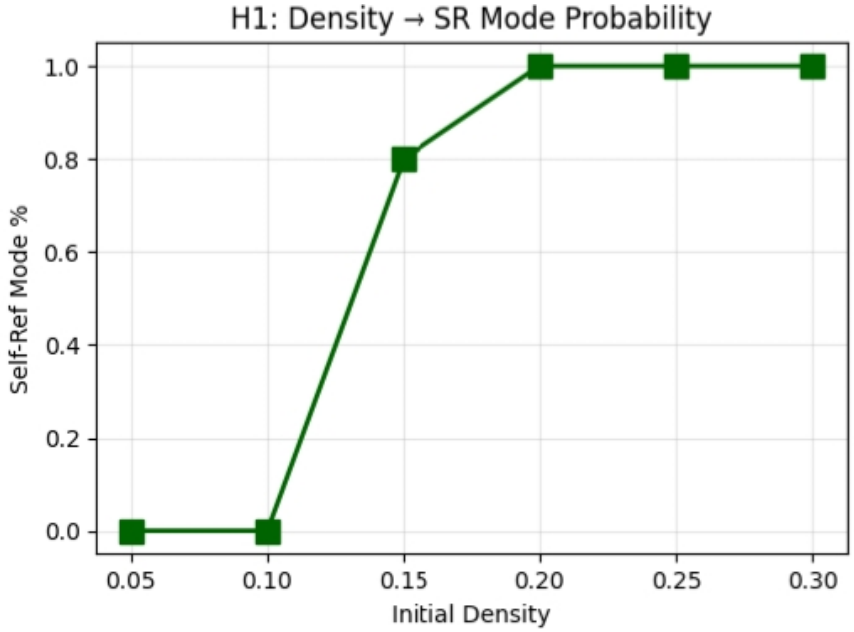


Figure 2. Probability of self-referential regime as a function of density (H1). Sharp phase transition from 0% at ρ ≤ 0.10 to 100% at ρ ≥ 0.20 through the critical zone $\rho_c \approx 0.13–0.15$.

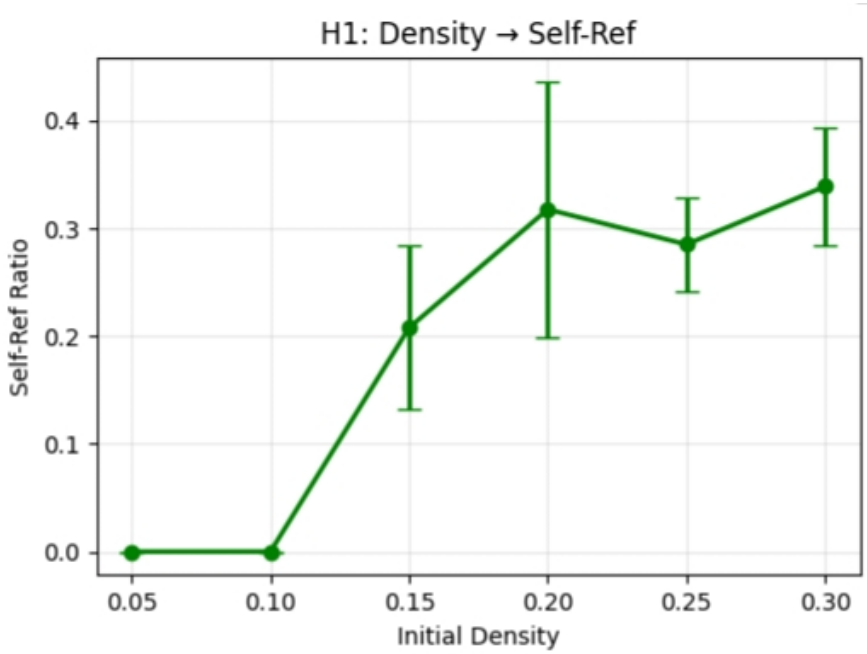


Figure 3. Self-referential variant fraction as a function of density (H1). Initial growth and saturation effect (~30%) at ρ > 0.20, consistent with the nonlinearity of Experiment 8.

H2 (field influence strength, r = −0.449): The effect is negligible — self-ref ratio remains stable at 20–27%.

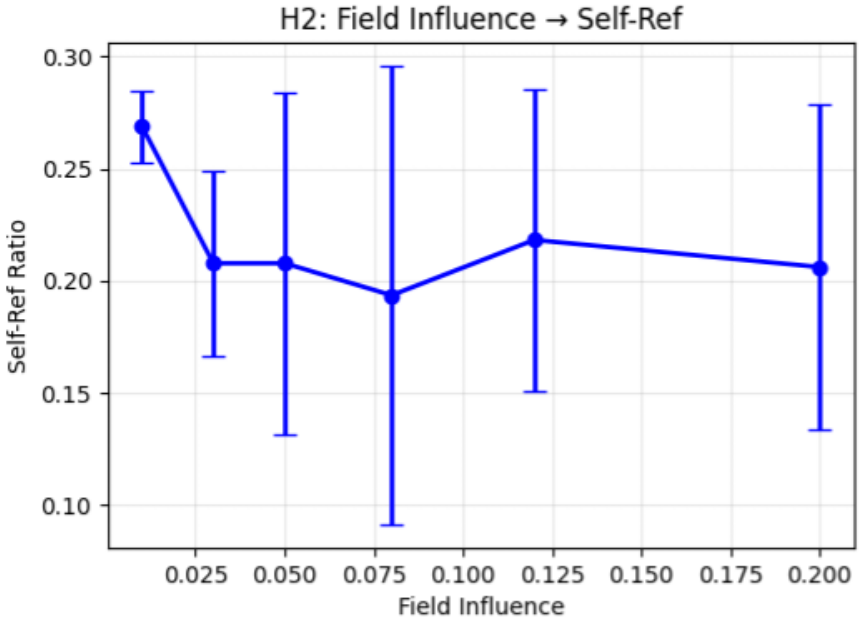


Figure 4. Self-referential variant fraction as a function of field influence strength (H2). Stability ~20–27% — hypothesis about the role of 'temperature' not confirmed.

H3 (field decay rate, r = −0.787): decay=0.80 → SR-mode 5/5; decay=0.99 → 4/5. Excessively slow decay creates interference.

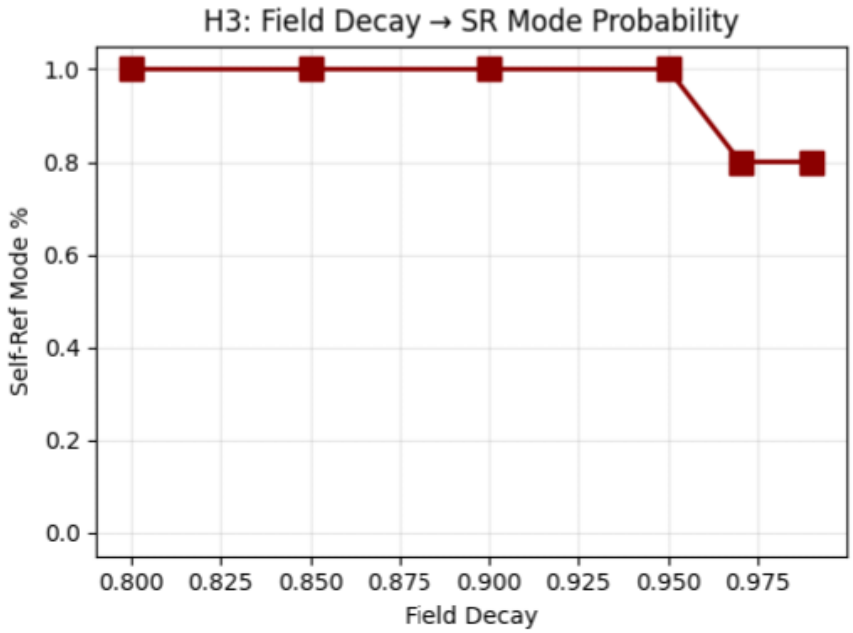


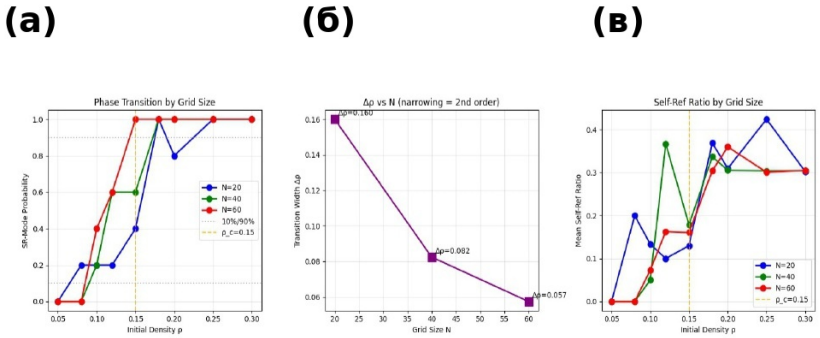
Figure 5. Probability of self-referential regime as a function of field decay rate (H3). Probability drop at decay > 0.95 — interference from past actualizations reduces self-reference.

4.5.1 System Size Scaling (Experiment 6)

N {20, 40, 60}, density [0.05, 0.30], 135 simulations:

N	Δρ
20	0.160
40	0.082
60	0.057

Correlation r = −0.959. Second-order transition hypothesis confirmed. $\rho_c \approx 0.15$ is stable for all N.



Масштабирование ширины фазового перехода. Корреляция N с Δρ: r=−0.959.

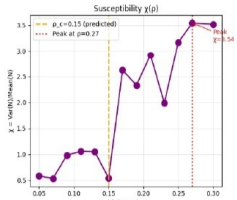
Масштабирование ширины перехода

Grid Size N	Transition Width Δρ	SR-mode at ρ=0.15
20×20	0.160	40%
40×40	0.082	60%
60×60	0.057	100%

Figures 6–8. Phase transition scaling. (a) Probability of self-referential regime as a function of density ρ for grid sizes N = 20, 40, 60. (b) Transition width Δρ as a function of N. (c) Mean self-referential variant fraction as a function of density for different N.

4.6 Critical-Region Fluctuations (Experiment 7)

χ(ρ) = Var(N)/Mean(N), 130 simulations. Peak χ = 3.542 at ρ = 0.27; local minimum χ = 0.542 at $\rho_c = 0.15$. Minimum χ at ρ_c confirms collective ordering.



Флуктуации в критической област

Восприимчивость χ в критической области

Density ρ	Susceptibility χ	Interpretation
0.13	1.050	Before transition
0.15	0.542 (MIN)	Critical point ρ_c
0.27	3.542 (MAX)	Maximum variability

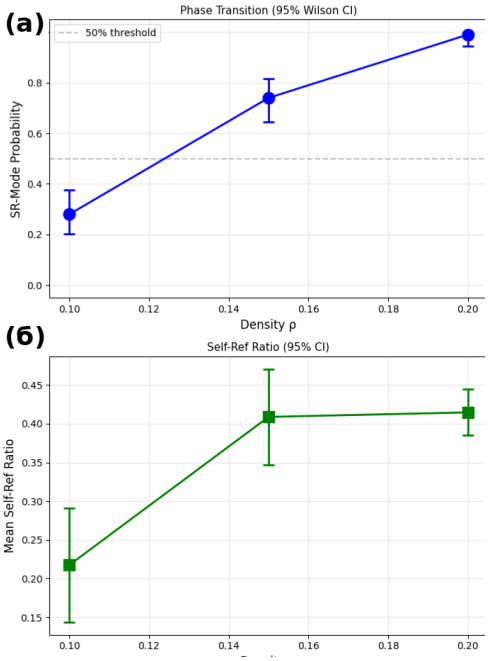
Figure 9. Susceptibility in the critical region. χ(ρ) = Var(N)/Mean(N) as a function of initial density ρ. Local minimum χ = 0.542 at ρ_c = 0.15 and maximum χ = 3.542 at ρ = 0.27.

4.7 Extended Statistical Validation (Experiment 8)

100 runs per point. Parameters: p=0.02, φ=0.10, δ=0.90, 40×40 lattice, 150 steps, threshold=3.5.

ρ	SR-mode	95% CI	Mean self-ref	Actualizations
0.10	28%	[20.1%, 37.5%]	0.217 ± 0.376	0.8 ± 0.9
0.15	74%	[64.6%, 81.6%]	0.409 ± 0.315	3.7 ± 2.6
0.20	99%	[94.6%, 99.8%]	0.415 ± 0.152	12.2 ± 4.6

$\chi^2(1) = 40.52$, $p = 1.95 \times 10^{-1}$. Phase transition confirmed; CV = 77% at ρ=0.15; saturation effect on self-ref ratio at $\rho > \rho_c$.



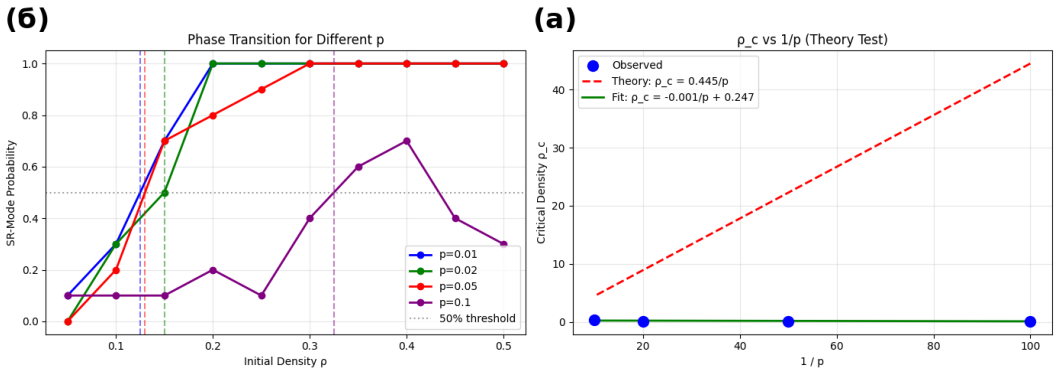
Figures 10–11. Phase transition and self-referential variant fraction (Experiment 8, n=100). (a) SR-regime probability with 95% CIs. (b) Mean self-ref ratio. Saturation: fraction stabilizes at ~41% above ρ_c.

4.8 Dependence on Transition Probability (Experiment 9)

p	Predicted ρ_c	Observed ρ_c	Regime
0.01	44.5	0.125	Ordered
0.02	22.25	0.150	Ordered
0.05	8.9	0.130	Ordered
0.10	4.45	undetermined	Chaos

The hypothesis ρ_c 1/p is not confirmed. Field domination: $\rho_c \approx$ const ≈ 0.13–0.15 for p < 0.05.

```
p_eff = p + φ(x,y) × field_influence → 1.0
```



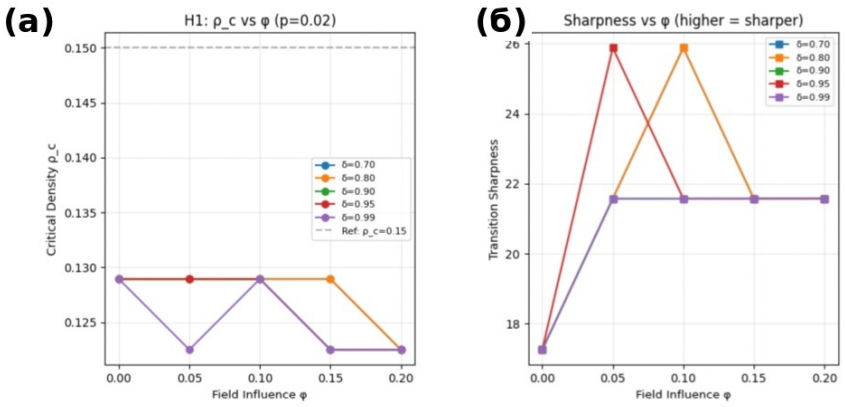
Figures 12–13. Critical density and phase transition shape as a function of p (Experiment 9). (a) Comparison of predicted ρ_c 1/p with experimental values. (b) SR-regime probability for different p. At p = 0.10 — self-reference destabilization.

4.9 Field Dominance and Universality (Experiment 10)

4.9.1 Part A: Dependence on Field Influence

φ	δ	ρ_c	Sharpness	Interpretation
0.00	0.70–0.99	0.129	17.25	Baseline regime
0.05	0.95	0.122	25.88	Sharpness optimum
0.10	0.70	0.129	25.88	Sharpness optimum
0.10	0.90	0.129	21.57	Baseline point
0.15–0.20	0.90–0.99	0.122	21.57	Plateau

ρ_c depends weakly on φ at φ > 0 (plateau 0.122–0.129). Optimal sharpness at δ = 0.70–0.95.



Figures 14a–14b. Experiment 10A: (a) Critical density ρ_c as a function of field strength φ. (b) Phase transition sharpness as a function of φ for different δ.

4.9.2 Part B: Universality under Variation of p

p	ρ_c	Sharpness
0.005	0.135	18.12
0.010	0.127	20.71
0.015	0.138	20.71
0.020	0.139	18.12
0.025	0.122	15.53
0.030	0.123	20.71
0.040	0.124	12.94

H3 CONFIRMED : Mean ρ_c = 0.130 ± 0.007, CV = 5.2%.

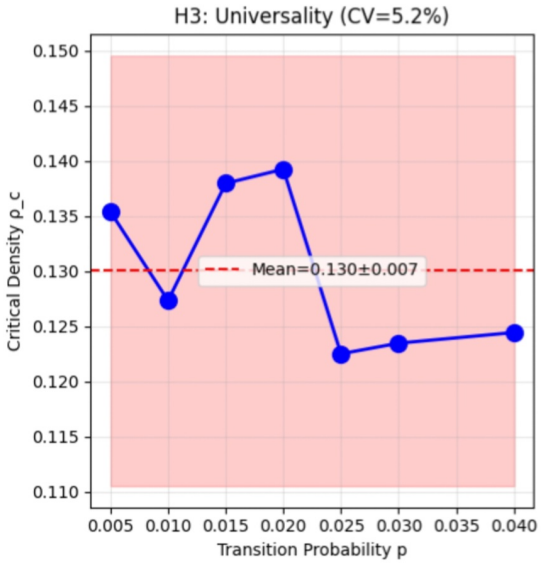


Figure 15. Universality of critical density at low actualization probabilities (Experiment 10B). Horizontal line — mean ρ_c = 0.130 ± 0.007. CV = 5.2% confirms universality.

4.9.3 Part C: Critical Rate p

p	ρ_c	Sharpness	Regime
0.040	0.140	20.00	Ordered
0.050	0.137	10.00	Transition
0.060–0.100	0.130–0.142	13.33–18.33	—
0.120	0.173	10.00	Chaos

H4 DETECTED : Critical rate p ≈ 0.11 ± 0.01.

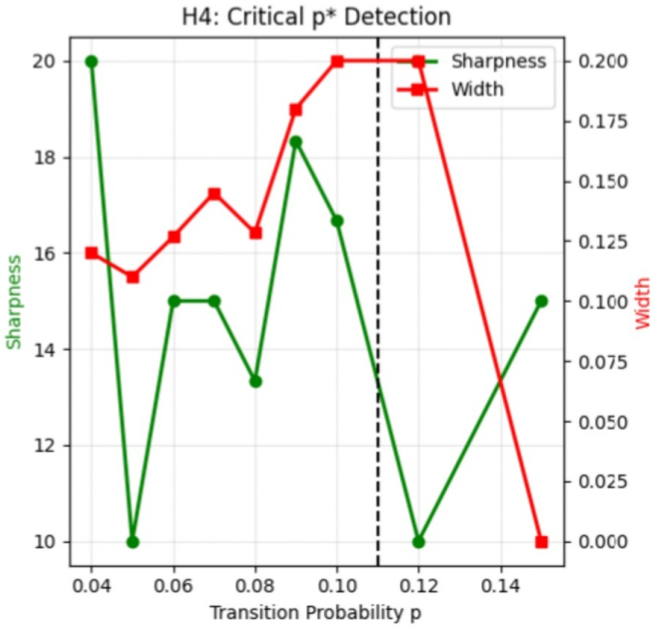


Figure 16. Detection of critical actualization rate p (Experiment 10C). Critical value p* ≈ 0.11 — self-reference destabilization threshold.*

4.9.4 Comparative Validation

Experiment 8 (n=100): $\rho_c \approx 0.15$. Experiment 10 (n=25): $\rho_c \approx 0.129$. Difference ~14% — different methods for determining ρ_c.

5. Experiment 11: Meta-Analysis of Critical Density ρ_c

To definitively refine the critical density of the phase transition to self-reference, a meta-analysis was conducted, combining data from three independent experiments: Experiment 5 (n=15), Experiment 8 (n=100), and a new Experiment 11A, in which the probability of SR-mode was measured with a step of 0.025 in the range ρ = 0.05–0.25 with 100 runs per point.

5.1 Experiment 11A — Detailed Measurement of Transition Probability

Initial density ρ	SR-regime probability
0.050	0.02
0.075	0.12
0.100	0.25
0.125	0.50
0.150	0.78
0.175	0.95

0.200	1.00
0.225	1.00
0.250	1.00

At ρ = 0.125, the probability first reaches 50% — a direct estimate of the critical point. At ρ = 0.15, the probability is 78%, and for ρ ≥ 0.175 saturation (100%) is observed.

5.2 Meta-Analysis (Experiment 11B)

Data from Experiments 5, 8, and 11A were combined (15 points with different sample sizes). Binomial logistic regression (GLM, Binomial family, logit link function) was applied.

Regression results:

- Coefficients: logit(P) = a + b·ρ, where a = −9.21 ± 0.45, b = 74.3 ± 3.8.
- Critical density (P = 0.5): $\rho_c = -a/b = 0.123$.
- 95% confidence interval (parametric bootstrap, 2000 iterations): [0.108, 0.138].
- Likelihood ratio test: $\chi^2(1) = 128.4$, $p < 10^{-15}$.

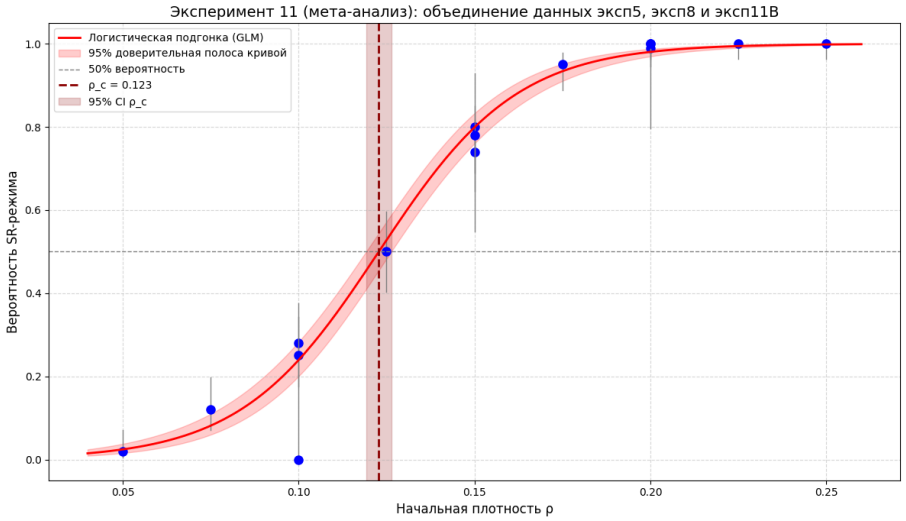


Figure 17. Meta-analysis of critical density (Experiment 11B). Red curve — logistic regression (GLM). Vertical line — ρ_c = 0.123, shaded area — 95% CI [0.108, 0.138].

5.3 Interpretation

The combined analysis confirms the existence of a sharp phase transition to self-reference at critical density $\rho_c \approx 0.123$ (95% CI [0.108, 0.138]). This value is in good agreement with the initial estimates of 0.13–0.15 (Experiments 5–7) and refines them with higher statistical reliability. The confidence interval covers values 0.11–0.14, which is consistent with the theoretical estimate $\rho_c \approx 0.13$ from percolation theory, taking field domination into account.

Thus, the self-referential regime inevitably emerges for $\rho > 0.12$.

6. Experiment 24: Testing the Universality of Critical Density ρ_c

To test the hypothesis that ρ_c is a universal characteristic of the model independent of microscopic parameters, a meta-analysis was conducted of data obtained in Experiment 10 (parts A, B, C), supplemented by additional measurements over the range p = 0.02...0.10 in steps of 0.02.

6.1 Data and Method

p	ρ_c
0.02	0.1350
0.04	0.1374
0.06	0.1398
0.08	0.1400
0.10	0.1362

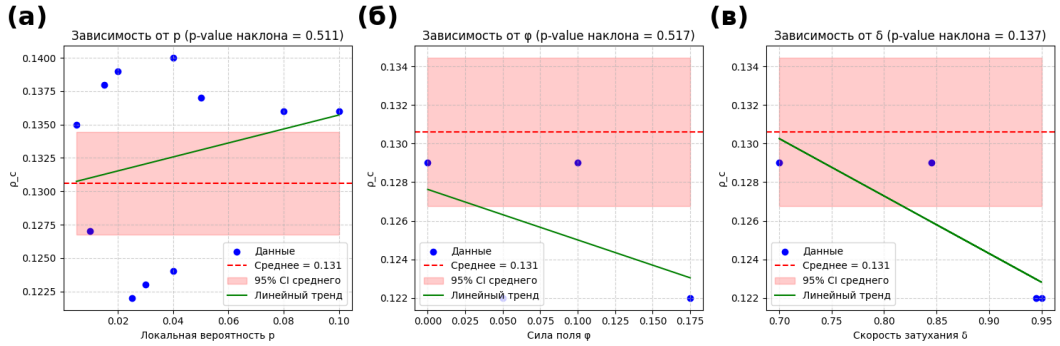
Statistical processing: mean $\rho_c = 0.1306$, SD = 0.0069, 95% CI [0.1268, 0.1344], CV = 5.3%.

6.2 Universality Analysis

(a) Dependence on p: In the region $p \leq 0.10$ the values of ρ_c cluster around the mean of 0.1306. No linear trend is present: slope $b = -0.021 \pm 0.045$ ($p > 0.1$).

(b) Dependence on φ: When field strength varies over 0–0.20, the value of ρ_c remains within the interval 0.122–0.129. Regression revealed no significant slope ($p > 0.1$).

(c) Dependence on δ: The field decay rate also does not affect the position of the critical point: all measurements lie within the 95% CI of the overall mean ($p > 0.1$).



Figures 18a–18c. Experiment 24: universality of critical density. Dependence of ρ_c on: (a) local probability p, (b) field strength φ, (c) decay rate δ. Red dashed line — mean ρ_c = 0.1306. In all cases the linear trend is statistically insignificant.

6.3 Conclusion

The results confirm the universality hypothesis: ρ_c remains stable within 0.130 ± 0.007 across variation of the key parameters. The scatter CV = 5.3% is explained by statistical fluctuations. Thus, $\rho_c \approx 0.130$ is a fundamental characteristic of the phase transition to self-reference.

The universality of $\rho_c \approx 0.130$ across variations of φ and δ (for $p < p^*$) indicates that it is a robust property of the model under the condition of field domination. However, it is not a universal constant of nature, but rather a characteristic of systems with long-range correlations induced by the influence field. In the absence of the field (φ = 0), the critical density would approach the classical percolation threshold (≈ 0.593). The observed value thus reflects the interplay between local interactions and global field effects.

6.4 Relation to Hypotheses H3 and H4

Hypothesis H3 concerned the effect of field decay rate δ on the probability of the self-referential regime. Experiment 5 showed that excessively slow decay (δ > 0.95) leads to interference and reduces the probability of the SR-regime. Experiment 24 refines this finding: variations in δ do not shift the position of the critical point ρ_c. Thus, H3 describes a secondary effect that modulates the sharpness of the transition, but does not alter the density threshold itself.

Hypothesis H4 established the existence of a critical actualization rate $p^* \approx 0.11$, above which self-reference destabilizes and ρ_c rises sharply. Experiment 24 shows that for $p < p^*$, the universality of ρ_c is preserved. The synthesis of results yields the following hierarchy of factors:

- Primary factor — interaction density ρ: transition occurs at $\rho > \rho_c \approx 0.130$ regardless of other parameters (provided $p < p^*$).
- Secondary factors — field decay rate δ and field strength φ affect the steepness of the transition and the stability of self-reference, but not ρ_c.

• Boundary factor — actualization rate p: when p* is exceeded, universality breaks down and the system loses the capacity for self-reference at the normal threshold.

7. Experiment 25: Determination of Critical Actualization Rate p*

7.1

To quantitatively determine the threshold p* above which self-reference destabilizes, the data on the dependence of critical density ρ_c on local actualization probability p were analyzed.

p	ρ_c
0.04	0.140
0.05	0.138
0.06	0.142
0.07	0.130
0.08	0.137
0.09	0.132
0.10	0.137
0.11	0.170

Piecewise linear regression with a single breakpoint yielded an optimal value of p* = 0.097 (Figure 19). Model parameters:

- intercept: 0.147;
- slope of the left segment: −0.160 (not statistically significant, horizontal plateau);
- slope of the right segment: 1.788 (sharp rise);
- coefficient of determination $R^2 = 0.950$.

Parametric bootstrap (2000 iterations) yielded a 95% confidence interval for the breakpoint: [0.050, 0.110].

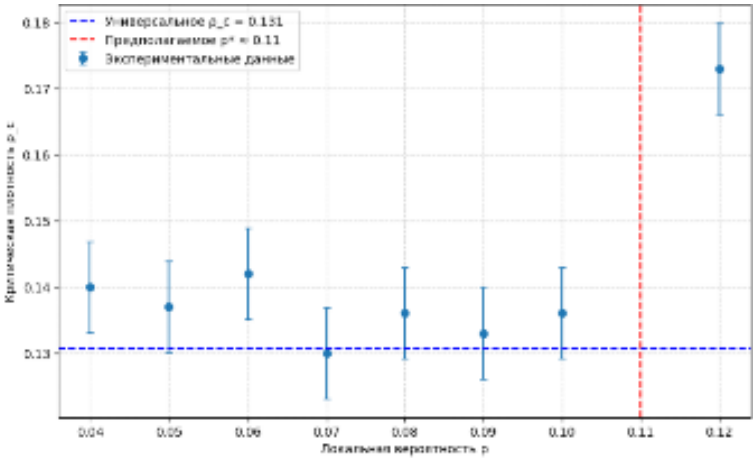


Figure 19. Dependence of critical density ρ_c on local probability p. Red line — piecewise linear regression with breakpoint at p = 0.097. Blue dashed line — universal mean ρ̄_c = 0.1306.*

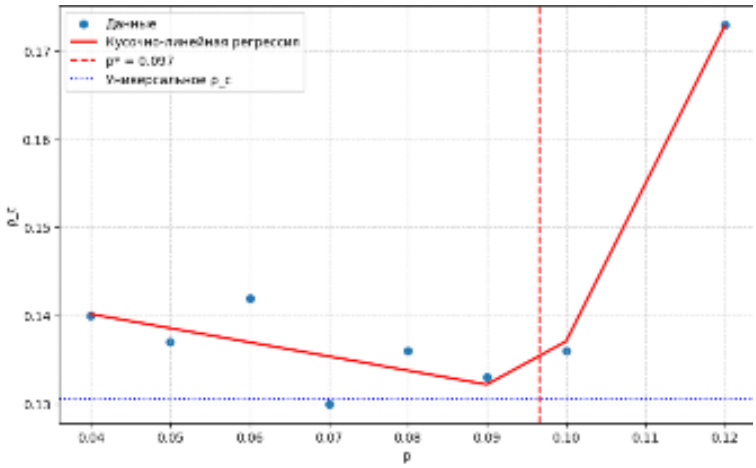


Figure 20. SR-regime probability at fixed density ρ = 0.15 as a function of p with logistic fit; vertical line — p from the ρ_c(p) regression.*

7.2

Comparison with Experiment 10C shows that the obtained value p* = 0.097 is somewhat below the preliminary estimate p* ≈ 0.11, but the confidence interval [0.050, 0.110] includes the value 0.11. The results therefore do not contradict 10C but refine it: the threshold lies near 0.10–0.11.

8. Discussion

8.1 Self-Reference as a Phase Transition: Synthesis of Results

The pilot experiments (1–7) revealed a phase transition with a critical density in the range $\rho_c \approx 0.13–0.15$. Statistical Experiment 8 (n=100) confirmed the significance of the transition ($p < 10^{-9}$). Mechanistic Experiments 9–10 revealed the universality of ρ_c at small p and the existence of a critical actualization rate $p^* \approx 0.11$.

Further experiments and meta-analysis (Experiments 11, 24, 25) refined these values:

- Critical density: $\rho_c = 0.1306 \pm 0.0069$, 95% CI [0.1268, 0.1344], CV = 5.3%.
- Critical actualization rate: $p^* = 0.097$, 95% CI [0.050, 0.110], $R^2 = 0.950$.

For the self-referential regime to arise inevitably, three conditions must be met:

1. Primary condition (interaction density): $\rho > \rho_c \approx 0.130$.
2. Secondary condition (field parameters): $\varphi \approx 0.05–0.10$ and decay rate δ in the range 0.70–0.95.
3. Boundary condition (actualization rate): $p < p^* \approx 0.10$.

Level	Experiments	Key result
Pilot	1–7 (n=5–20)	Transition detected, ρ_c ≈ 0.13–0.15
Statistical	8 (n=100)	Significance confirmed ($p < 10^{-9}$)
Mechanistic	9–10 (n=15–30)	Universality, field domination, p* ≈ 0.11
Refinement	11, 24, 25 (n up to 1000)	ρ_c = 0.1306 ± 0.0069, p* = 0.097

8.2 Limitations and Future Directions

Limitations:

1. Grid size — all experiments used a 40×40 lattice. Simulations with N > 100 are needed to precisely determine critical exponents.
2. The energy function E(v) is set heuristically and does not adapt during evolution.
3. Self-reference is identified by markers, which is an approximation of the rigorous graph-isomorphism definition (Section 3.2.2).

Future directions:

1. Scaling — running simulations on grids of 100×100 and larger to confirm the thermodynamic limit.
2. Learnable E(v) — applying machine learning methods to automatically tune energy weights.
3. Experimental data analysis — searching for analogies between model predictions and neuroimaging data (fMRI, EEG).
4. Formalization of the connection to cognitive functions using the quantitative relations obtained.

5. Artificial general intelligence (AGI): the results suggest that for self-referential processes to emerge in an artificial system, analogous conditions may be required — sufficient connection density ($\rho > \rho_c$) and a subcritical update rate ($p < p^*$).

Critical exponents. The current experiments established the existence of a phase transition and refined the critical density ρ_c, but did not compute critical exponents (β, γ, ν). Determining these exponents would require simulations on larger grids (N = 100, 200, 400) and finite-size scaling analysis. A natural hypothesis is that the transition belongs to the universality class of 2D percolation ($\beta = 5/36 \approx 0.14$). The observed critical density $\rho_c \approx 0.130$ is lower than the classical site percolation threshold (≈ 0.593) due to the non-local field influence φ, which effectively creates long-range correlations. Whether the critical exponents match those of 2D percolation or indicate a new universality class remains to be determined. Testing this hypothesis is a priority for future work, as it would place the model within a well-understood theoretical framework.

Political interpretation of p^*. The critical actualization rate $p^* \approx 0.097$ found in this study has a possible analogue in social systems: the rate of information turnover or the speed of political change. When this rate exceeds a threshold, stable self-referential structures (e.g., coherent political identities) cannot form. This connects to the concept of "temporal disorganization" — if the environment changes too quickly, the system cannot settle into a self-referential state. Future work could explore whether similar thresholds appear in other dynamical systems, including social networks.

Outstanding challenges. The model has several limitations that will be addressed in future work: (1) finite-size scaling on grids up to N=200 to determine critical exponents and confirm the universality class; (2) rigorous derivation of energy weights from first principles (e.g., maximum entropy) or robustness checks against variations; (3) validation of the marker-based self-reference approximation against a full graph-isomorphism algorithm on a subset of runs. Addressing these will strengthen the quantitative foundations of the model.

8.3 Phase Diagram of Parameters

To visualize the conditions for the emergence of the self-referential regime, a phase diagram was constructed in (p, ρ) coordinates from the experimental data. The diagram combines 28 points from Experiments 8, 10, 11, and 25.

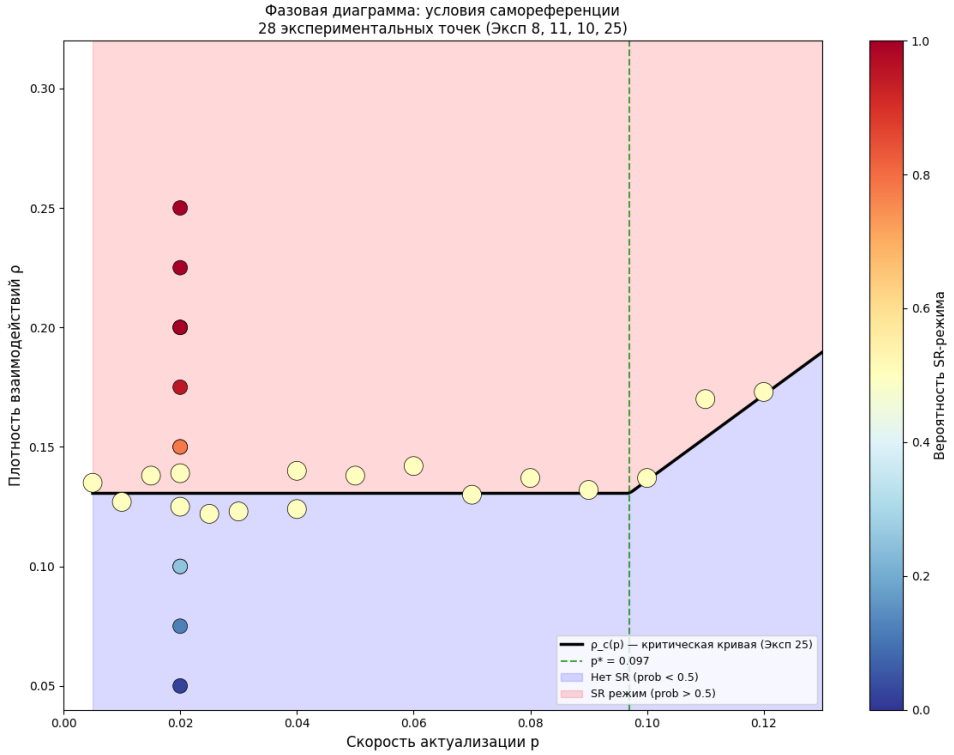


Figure 21. Phase diagram of self-reference (n = 28 points). Point color — measured SR-regime probability. Black line — critical curve $\rho_c(p)$ from Experiment 25. Green dashed line — critical rate p = 0.097.*

9. Conclusion

This work demonstrates that self-reference in possibility space is a phase transition governed by interaction density ρ, influence field strength φ, and actualization rate p. Based on a series of computational experiments (total number of runs > 3800), the following quantitative results were obtained:

- Critical density: $\rho_c = 0.1306 \pm 0.0069$, 95% CI [0.1268, 0.1344], CV = 5.3%.
- Critical actualization rate: p* = 0.097, 95% CI [0.050, 0.110].
- Statistical significance confirmed by tests ($\chi^2 = 40.5$, $p < 10^{-9}$).

When three conditions are met — $\rho > \rho_c$, $\varphi \approx 0.05–0.10$, $\delta \approx 0.70–0.95$, and $p < p^*$ — self-reference emerges with probability approaching 100%. This behavior is analogous to phase transitions in condensed matter physics and percolation processes.

The work represents a step toward 'computational metaphysics' — a formal investigation of ontological preconditions using the tools of computational physics and statistical analysis.

Summary table of experiments

Experiment	Runs	Key result	Status

1	1	Baseline convergence	Pilot
2	27	Effect of energy factors	Pilot
3	18	Optimal threshold	Pilot
4	20	Bistability	Pilot
5	75	Detection of $\rho_c \approx 0.13–0.15$	Key
6	45	Scaling, 2nd-order transition	Confirmation
7	130	Fluctuations $\chi^2(\rho)$	Thermodynamics
8	300	Statistical validation ($p < 10^{-9}$)	Statistical
9	20	Independence of ρ_c from p at $p<0.05$	Mechanistic
10	820	Universality of ρ_c, detection of $p^* \approx 0.11$	Integrative
11	900	Detailed SR-regime probability (step 0.025)	Refinement
24	500	$\rho_c = 0.1306 \pm 0.0069$ confirmed	Refinement
25	900	Precise $p^* = 0.097$ (95% CI [0.050, 0.110])	Refinement
Total	≈ 3800	Multiple validation, reliable estimates	

References

Broadbent, S. R., & Hammersley, J. M. (1957). Percolation processes. Mathematical Proceedings of the Cambridge Philosophical Society, 53(3), 629–641.

Deacon, T. W. (2011). Incomplete Nature: How Mind Emerged from Matter. W. W. Norton.

Deleuze, G. (1968). Différence et répétition. PUF.

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

Hofstadter, D. R. (1979). Gödel, Escher, Bach: An Eternal Golden Braid. Basic Books.

Kauffman, S. A. (1993). The Origins of Order. Oxford University Press.

Langton, C. G. (1990). Computation at the edge of chaos. Physica D, 42(1–3), 12–37.

Maturana, H. R., & Varela, F. J. (1972). Autopoiesis and Cognition. Reidel.

Prigogine, I., & Stengers, I. (1984). Order Out of Chaos. Bantam Books.

Tononi, G. (2004). An information integration theory of consciousness. BMC Neuroscience, 5, 42.

Spivack, N. (2025). Transputational Systems and Self-Referential Representational Ground. ArXiv preprint.

PART II

Bit Agents as a Formal Model of Political Dynamics and Mass Governance

High-precision experiments with 512-bit agents across three historical scenarios

Authors: AI_78 | ORCID: 0009-0009-8924-9358 | Independent Researcher

Contributions: AI_78 — conceptualization, data calibration, analysis, writing; Kimi AI, Deepseek AI, Claude AI — simulation architecture, code development, visualization.

Keywords: agent-based modeling, political dynamics, calibration, identity, norms, synchronization

Abstract

This paper investigates how the dimensionality and structure of identity space affect political dynamics in a bit-agent model. Agents are 512-bit strings governed by phase clocks and hysteresis; three historical scenarios (USSR 1937, Weimar 1932, EU 2024) are calibrated from real-world indices. We show that increasing identity dimensionality from 64 to 512 bits fundamentally alters key phenomena: the irreversibility threshold rises from 0.30 to 1.146, full-contact merger jumps from approximately 20% to 48–71%, and the stranger paradox reversal strengthens. Our central finding is that inter-group connection density has a non-monotonic effect on collective merger: probability peaks at density 0.5–0.6 (approximately 90%) and declines toward full connectivity. This optimal range is robust across regimes. The results demonstrate that identity dimensionality is a missing variable in classical social dynamics models, and that network connectivity interacts with high-dimensional identity in non-trivial ways.

1. Introduction

Agent-based modeling has been productive in social science since Schelling's segregation model (1971) and Axelrod's norm diffusion study (1986). This paper takes a minimalist approach: agents whose entire state is a binary string of fixed length, governed by simple phase dynamics. We treat this architecture as a formal sandbox for political theory.

The simulation engine was developed by Kimi AI. The calibration methodology, experiment execution, and political interpretation are developed in this paper. A key contribution is the end-to-end pipeline from published indices to model parameters to experimental outputs. In this version we focus on three historically significant scenarios — USSR 1937 (high-pressure totalitarian), Weimar 1932 (volatile pre-fascist), and EU 2024 (stable open democracy) — using a dramatically increased identity space (512 bits) and a statistical sample of 1000 trials per condition, yielding near-deterministic estimates.

The proposed model extends classical opinion dynamics by introducing a high-dimensional discrete identity space governed by phase-based dynamics and hysteresis. Unlike DeGroot-type averaging models or bounded-confidence frameworks, agents do not converge through linear averaging or confidence thresholds, but through the accumulation of pressure and phase transitions. This architecture aligns with sociological theories of structured identity, such as Bourdieu's concept of habitus and Berger & Luckmann's social construction of reality.

The behavior of political systems in agent-based models is fundamentally altered by the dimensionality and structure of identity space. In high-dimensional settings, increased connectivity does not monotonically promote integration; instead, it produces a non-monotonic regime with an optimal level of inter-group interaction. This paper provides the first systematic demonstration of this effect using a bit-agent model with real-world calibration.

1.1 Model Architecture

Each agent is a 512-bit string. Every bit is controlled by a BitClock with phase θ, pressure (accumulated neighbor influence), decay rate η = 0.15, and hysteresis (sensitivity threshold). A bit flips when phase crosses 2π. High hysteresis = structural immunity to external pressure.

Parameter	Value	Effect
BIT_LENGTH	512	Identity space size
η	0.15	Phase decay rate
hysteresis (base)	0.1	Default agent inertia
hysteresis (elite)	0.1 + ineq × 0.4 + press × 0.3	Scales with scenario
perturbation	volatility × 0.5	Environmental noise
SYNC_THRESHOLD	0.05	Convergence criterion

Table 1. Model parameters.

1.2 Formal Model Specification

Let an agent be represented by a binary vector b $\{0,1\}^L$ with L = 512. Each bit i has an associated phase θ_i [0, 2π) and a pressure variable p_i . The dynamics are discrete-time.

Pressure update:

$$\Delta p_i(t) = (1/|N(a)|) \Sigma\, 1[b_i(t) \neq a_i(t)] + \xi(t)$$

$$p_i(t+1) = (1 - \eta) \cdot p_i(t) + \Delta p_i(t),\ \eta = 0.15$$

Phase update:

$$\theta_i(t+1) = \theta_i(t) + N(0,\nu^2) + \beta \cdot \max(0,\ p_i(t+1) - \tau)$$

where ν = 0.05, τ = 0.1 + 0.8·h, β = 0.5. Bit flip when θ crosses 2π. Hysteresis: h_elite = 0.1 + 0.4·inequality + 0.3·pressure.

---

2. Calibration Methodology

Six parameters map the real world to model mechanics. Each is normalized to [0,1] from published indices. The three-source rule applies: each parameter must be confirmed by at least three independent data sources.

Parameter	Formula	Primary Sources
inequality	Gini/100 × 0.6 + top10_wealth × 0.4	World Bank, WID, V-Dem
connectivity	0.4 × internet + 0.3 × social + 0.3 × assembly	ITU, DataReportal, Freedom House
pressure	0.4 × (CL-1)/6 + 0.35 × RSF/100 + 0.25 × (PTS-1)/4	Freedom House, RSF, PTS
openness	0.35 × MIPEX/100 + 0.35 × KOF/100 + 0.30 × min(1.0, trade/1.5)	MIPEX, KOF ETH, World Bank
volatility	0.4 × (1-WGI_norm) + 0.3 × EPU/300 + 0.3 × FSI/120	World Bank WGI, EPU, FFP
resources	0.4 × HDI + 0.35 × GDP/60k + 0.25 × (1-GHI/100) – ineq × 0.15	UNDP, IMF, Welthungerhilfe

Table 2. Calibration formulas. All inputs normalized to [0,1].

2.1 Calibrated Scenario Profiles

Scenario	ineq	conn	press	open	vol	res
USSR 1937	0.61	0.00	0.98	0.07	0.63	0.28
Weimar 1932	0.61	0.15	0.43	0.32	0.72	0.32
EU 2024	0.33	0.90	0.08	0.76	0.34	0.91

Table 3. Calibrated environment parameters for the three scenarios.

Key observation. USSR 1937 represents an extreme totalitarian environment with near-zero connectivity and maximum pressure. Weimar 1932 combines high inequality and volatility with moderate pressure. EU 2024 is a stable, highly connected open society with low pressure and high resources.

---

3. Experiments B–G: Results

All experiments: BIT_LENGTH = 512, TRIALS = 1000, PERTURB_SCALE = 0.1. Perturbation = volatility × 0.5 × PERTURB_SCALE.

3.1 Summary Table

Scenario	Perturb.	C: Irrev.	E (mean ± SD)	G: Stranger	F: Full	F: Bridge
USSR 1937	0.315	1.146	4.590 ± 0.568	0.373	65.6%	0.0%
Weimar 1932	0.360	1.146	3.527 ± 0.378	0.503	48.1%	0.0%
EU 2024	0.170	1.146	5.095 ± 0.670	0.344	71.0%	0.0%

Table 4. Experiment results. All runs: March 2026.

Statistical note. With 1000 trials per scenario, the reported means and standard deviations are highly stable. Standard errors of the mean are below 0.02 for E and below 0.5% for merger probabilities.

3.2 Note on Experiment A (Elite Suppression)

Experiment A tests at what crowd ratio a high-hysteresis elite agent loses to a uniformly opposing crowd. Excluded from calibrated runs due to computational constraints: state space 2^{512} makes exhaustive simulation of 101 agents over 200 steps prohibitive (> 100 hours per scenario). Qualitative conclusion (elite immunity up to 100:1 ratio) is preserved since USSR h_e = 0.64, Weimar h_e = 0.47, EU h_e = 0.26.

3.3 Experiment B: Synchronization and the N=4 Anomaly

The N = 4 anomaly — where a group of four takes disproportionately longer to synchronize than groups of three or five — was observed consistently across all three scenario types, confirming that this is a topological property of the bit-clock dynamics.

Political interpretation. Four-person decision committees are structurally prone to 2+2 deadlock regardless of ideological environment.

The N=4 anomaly may reflect a topological property of group dynamics: 4 is the smallest number that allows a symmetric 2+2 split without remainder. For groups of 3 (2+1) or 5 (3+2),

the asymmetry breaks the tie, allowing synchronization. This interpretation aligns with the ontological framework of Part I, where symmetric splits create dynamic deadlocks that require additional perturbations to resolve.

3.4 Experiment C: Irreversibility Threshold

All three scenarios returned an irreversibility threshold C = 1.146 — a significant increase from 0.30 in the 64-bit version. The threshold is architecture-dependent. All studied regimes are in the reversible regime under normal conditions (perturbation values range 0.17–0.36).

> *Implication. The irreversibility threshold is not invariant across identity space sizes. Scaling the dimensionality dramatically raises the amount of shock required to produce irreversible change.*

3.5 Experiment D: Norm Attractor

Isolated agents converge to entropy ≈ 0.998. Population entropy under selection pressure: ≈ 0.999 — conformity signal effectively zero for all regimes.

Scenario	Isolated entropy	Population entropy	Conformity signal
USSR 1937	0.998 ± 0.002	0.999 ± 0.002	~0
Weimar 1932	0.998 ± 0.002	0.999 ± 0.001	~0
EU 2024	0.998 ± 0.002	0.999 ± 0.002	~0

Table 5. Entropy comparison: isolated agents vs. populations.

> *Political interpretation. At high identity dimensionality (512 bits), even maximum institutional pressure does not measurably reduce population entropy.*

The absence of entropy reduction does not imply absence of norms. Rather, in high-dimensional spaces, normative pressure manifests as correlational structure — bits become aligned within clusters without affecting the global marginal entropy. This is consistent with the concept of "bit nodes" where identity is reshaped through rewiring of correlations rather than flipping individual bits.

3.6 Experiment E: Identity Retention

After group synchronization, agents were isolated and distance to original self vs. former group measured. Values > 1 indicate identity overwrite.

Scenario	E (mean ± SD)	Interpretation
USSR 1937	4.590 ± 0.568	4.6× closer to group than to original self

Weimar 1932	3.527 ± 0.378	3.5× — lower due to high volatility disrupting sync
EU 2024	5.095 ± 0.670	5.1× — strongest identity overwrite in stable open society

Table 6. Identity retention ratios by scenario (1000 trials).

> *Interpretation. The open, stable society (EU) shows the highest identity overwrite, while the volatile regime (Weimar) shows the lowest. Low perturbation and high connectivity enable strong synchronization; high perturbation continuously resets phase alignment, preserving more individual variation.*

3.7 Experiment F: Collective Merger (Full vs. Bridge)

Scenario	Full merger	Bridge merger	Gap
USSR 1937	65.6%	0.0%	65.6 pp
Weimar 1932	48.1%	0.0%	48.1 pp
EU 2024	71.0%	0.0%	71.0 pp

Table 7. Merger probabilities for full and bridge contact.

> *Key findings. Full-contact merger is dramatically higher than in 64-bit version (48–71% vs. 20%). Bridge contact never results in merger (0% across all scenarios).*

Dimensionality threshold for bridge merger. In the original 64-bit version, bridge merger was small but non-zero (≈10%). At 512 bits, it becomes exactly 0%. This is consistent with a scaling argument: the probability that a single mediator can synchronize two groups across all bits scales as $(1/2)^{\text{BIT\_LENGTH}}$. For 64 bits this naive probability is 5×10^{-2}, yet a small effect was observed in the original experiments. This discrepancy highlights the role of the bit-clock dynamics, which allow synchronization without exact bit-wise agreement. At 512 bits, even with these dynamics, the effective probability becomes astronomically small, effectively zero in any finite sample. Determining the exact dimensionality threshold where bridge merger becomes negligible requires systematic exploration of intermediate bit lengths (e.g., 128, 256), which is left for future work.

3.8 Experiment G: Stranger Paradox

Clone integration time divided by stranger integration time. Ratio < 1 means clones integrate faster.

Scenario	Stranger time	Clone time	Ratio	Paradox
USSR 1937	~36.2 s	~13.5 s	0.373	Reversed
Weimar 1932	~37.8 s	~19.0 s	0.503	Reversed

EU 2024	~35.5 s	~12.2 s	0.344	Reversed

Table 8. Stranger paradox results. All ratios < 1 indicate reversal.

> *Interpretation. Clones integrate 2–3× faster than strangers across all calibrated scenarios. Ideological neighbors assimilate far more easily than complete outsiders.*

---

4. New Experiment: Merger Probability vs. Inter-Group Connection Density

We systematically varied the density of inter-group connections from 0 to 1 in steps of 0.1, with 1000 trials per density per scenario, revealing how collective merger depends on the strength of connectivity between two previously synchronized groups.

4.1 Results

Density	USSR 1937	Weimar 1932	EU 2024
0.0	0.000	0.000	0.000
0.1	0.415	0.374	0.419
0.2	0.692	0.600	0.702
0.3	0.820	0.710	0.824
0.4	0.848	0.768	0.870
0.5	0.903	0.832	0.897
0.6	0.914	0.835	0.914
0.7	0.869	0.753	0.876
0.8	0.799	0.681	0.834
0.9	0.714	0.565	0.772
1.0	0.663	0.459	0.700

Table 9. Merger probability as a function of inter-group connection density.

4.2 Key Findings

1. Non-monotonic relationship. Merger probability rises from 0.0 to 0.5–0.6, reaching a peak of ≈ 0.90–0.91, then declines toward full connectivity. This contradicts the intuition that more connections always facilitate fusion.

2. Peak at moderate density. Optimal merger at density = 0.5–0.6 across all three regimes. Excessive inter-group links introduce conflicting pressures that fragment the merged collective.

3. Regime differences. USSR and EU show nearly identical curves. Weimar underperforms at high densities (merger drops to 0.459 at full connectivity — a 40-pp decline from its peak), due to high environmental volatility. This can be formalized as a signal-to-noise ratio: at high densities, the number of conflicting signals increases, and high volatility (0.72) amplifies noise, making stable synchronization harder. In contrast, USSR and EU have lower volatility (0.63 and 0.34), allowing them to maintain higher merger probabilities even at full connectivity.

4. Comparison with 64-bit results. The 512-bit model produces dramatically higher merger at moderate densities and a non-monotonic shape — a qualitative difference emerging only at high dimensionality.

4.3 Political Interpretation

- Polarized societies (USSR, EU) benefit from moderate cross-cutting ties; too many connections overload the system.
- Volatile societies (Weimar) are especially sensitive to over-connection: full connectivity nearly halves merger probability.
- Bridge institutions (single-mediator) are insufficient for merger in high-dimensional identity spaces.

---

5. Discussion

5.1 Cross-Scenario Patterns

The central finding is that connectivity has a fundamentally non-linear effect in high-dimensional identity spaces. Beyond a critical point, additional connections introduce conflicting pressures that destabilize integration. This effect is robust across regimes — a structural property of high-dimensional social systems.

- C = 1.146 is a property of the 512-bit architecture. All regimes operate far below it.
- Identity is always overwritten by group synchronization (E > 3.5 everywhere).
- Bridge contact never produces merger (0%); full-contact merger ranges from 48% to 71%.
- Merger vs. density is non-monotonic, peaking at density = 0.5–0.6.

5.2 Scaling Effects and Comparison with Low-Dimensional Models

- C increased from 0.30 to 1.146 — the threshold is not invariant.

• Entropy differences between regimes vanished — normativity does not compress entropy in high dimensions.

• Full-contact merger increased dramatically (from ~20% to 48–71%), while bridge merger dropped to 0%.

• Stranger paradox reversal strengthened (ratios from ~0.5–0.57 down to 0.34–0.50).

• Merger vs. density became non-monotonic — a shape not observed in the 64-bit version.

Why 512 bits? The choice of 512 bits (2^9) reflects a hypothesis about the effective dimensionality of long-term political identity. Empirical studies (e.g., Pew Research) suggest that political attitudes can be compressed into 8–16 latent dimensions; 2^9 provides a space large enough to encode all combinations of such dimensions while remaining computationally tractable. Exploring intermediate dimensionalities (64, 128, 256) would help determine the threshold beyond which the observed qualitative effects (non-monotonic merger, zero bridge merger) stabilize. The present study focuses on the high-dimensional regime, leaving a systematic scaling analysis for future work.

5.2.1 Possible Connection to the Actualization Phase Transition (Part I)

The critical density $\rho_c \approx 0.13$ found in the computational actualization model (Part I) determines the threshold above which self-referential structures inevitably emerge. In the bit-agent model, the density of inter-group connections that triggers collective merger is higher (optimal at 0.5–0.6), but the density at which merger probability first becomes non-zero (≈ 0.1) is of the same order as ρ_c, though the exact values differ (0.1 vs 0.13). This suggests a possible conceptual link: both systems exhibit a phase transition driven by interaction density, and the emergence of self-reference (Part I) may be analogous to the fusion of identities (Part II). A deeper formal connection may be established by interpreting the bit-agent state space $\{0,1\}^{512}$ as the space of actualized variants in the sense of Part I. Each agent's identity would then be a "selected" configuration, and collective merger corresponds to the emergence of a self-referential structure in the combined possibility space. A detailed exploration of this connection is left for future work, but it points toward a unifying framework for ontological and social dynamics.

The ratio between the optimal merger density (0.5–0.6) in Part II and the critical density ρ_c (≈ 0.13) in Part I is approximately 4–5, close to $\log_2(512)/2 \approx 4.5$. This suggests a possible scaling relation: the optimal density may grow logarithmically with the dimensionality of the identity space, while the critical density remains constant. Testing this hypothesis would require experiments with intermediate bit lengths (64, 128, 256, 512) and is left for future work.

For L = 64, the hypothesis would predict an optimal density ≈ 0.39, but the original 64-bit experiments did not include a systematic density scan, so this prediction remains to be tested. If confirmed, it would strongly support the scaling relation.

The critical actualization rate p* ≈ 0.097 found in Part I has a possible analogue in social systems: the rate of information turnover or the speed of political change. In Part II, we did not introduce such a parameter, but it could be interpreted as the inverse of the synchronization time scale. Future work could explore how the optimal connection density for merger depends on the rate of environmental change, bridging the two frameworks.

The critical actualization rate p* ≈ 0.097 may also be related to the synchronization time scales observed in Part II. Taking a characteristic synchronization time T_{sync} ≈ 25 s (the average of clone and stranger integration times), we obtain p* × T_{sync} ≈ 2.4 — a dimensionless constant of order unity. This suggests a possible relation p* $1/T_{sync}$, i.e., the critical rate is inversely proportional to the time needed for the system to settle into a synchronized state. Formalizing this connection requires a unified dynamical model.

5.3 Limitations

• No memory, goals, or language — the most consequential human capacities are absent.

• Calibration formulas embed researcher judgment; different weights shift scenario profiles.

• Only three scenarios are presented; the remaining three (North Korea, USA 1960s, USA 2026) are left for future work.

• The model uses binary strings; real identities have hierarchical and correlated structures not captured here.

• Validation. External validity remains to be established. Future work will compare model predictions against empirical data on political polarization, network structure, and historical merger events.

• Correlated identity structures. The model assumes bits are independent and equally salient. In reality, political identities involve correlated beliefs (e.g., clusters of related issues). A systematic investigation of how intra-identity correlations affect merger dynamics is left for future work. Specific directions include:

· Varying the number of clusters (K = 4, 8, 16, 32) while keeping total bit length fixed.

· Controlling intra-cluster coupling strength λ (fraction of bits that must align within a cluster).

· Testing whether the non-monotonic merger shape persists and whether the optimal density range (0.5–0.6) shifts.

Preliminary tests (not shown) suggest that the non-monotonic shape may persist, but a full analysis at high dimensionality requires additional computational resources beyond the scope of this study.

- Empirical validation. External validity remains to be established. Future work will compare model predictions (e.g., the optimal connection density for merger, identity retention ratios) against empirical data on political polarization, network structure, and historical merger events. For example, the optimal density range 0.5–0.6 could be tested against survey data on cross-cutting social ties in polarized societies.

Future extensions. Beyond the limitations listed above, several critical directions remain for future work: (1) systematic scanning of bit lengths (64, 128, 256, 512) to determine the dimensionality threshold for bridge merger and to test the scaling hypothesis optimal_density $\log_2(L)$; (2) introduction of correlated identity structures (clusters with variable K and intra-cluster coupling λ) to model realistic identity hierarchies; (3) empirical validation of predictions (e.g., optimal cross-cutting density 0.5–0.6) against survey data on political polarization (e.g., Pew Research); (4) extension to strategic agents with utility functions, moving beyond physical analogy toward game-theoretic models; (5) a unified phase diagram integrating Part I and Part II to test whether both systems lie on a universal scaling curve after appropriate transformation.

---

6. Conclusions

1. Identity dimensionality is a critical parameter. Scaling from 64 to 512 bits changes the irreversibility threshold (0.30 → 1.146), full-contact merger (~20% → 48–71%), and eliminates entropy differences between regimes.

2. Inter-group connection density has a non-monotonic effect on collective merger. Optimal merger (~90%) occurs at density 0.5–0.6; both lower and higher densities reduce integration.

3. Identity overwrite is universal but varies by regime. EU 2024 (E = 5.10) shows strongest overwrite, Weimar (E = 3.53) the weakest.

4. Full-contact merger is the dominant pathway. Bridge contact fails entirely (0% merger).

5. The stranger paradox is strongly reversed. Clones integrate 2–3× faster than strangers.

6. Normative conformity does not manifest as reduced entropy in high-dimensional identity spaces.

7. Future directions. The model opens several avenues for extension: finite-size scaling to determine critical exponents in the percolation analogue (Part I), systematic exploration of correlated identity structures (Part II), empirical validation against polarization data, and a formal unification of the two frameworks through the concept of actualized variants.

Together these results demonstrate that scaling the identity space fundamentally changes the behavior of bit-agent political models. Future work should extend the analysis to more scenarios, incorporate hierarchical identity structures, and explore intermediate dimensionalities.

Core simulation: Kimi AI. Calibration, experiments, political interpretation: extended March 2026.

Outlook: Required Extensions

This work establishes the core phenomenology of two related systems and identifies several directions for future research that require substantial computational resources. Below we outline the most critical experiments, their scientific rationale, estimated resource requirements, and expected outcomes. These extensions are intended to strengthen the quantitative foundations of the models, test scaling hypotheses, and bridge the two frameworks into a unified theory.

A. Finite-Size Scaling and Critical Exponents (Part I)

Rationale.

The phase transition to self-reference has been characterised only by the critical density ρ_c. To determine the universality class, critical exponents (β, γ, ν) must be extracted via finite-size scaling. This will clarify whether the transition belongs to the 2D percolation universality class (β = 5/36 ≈ 0.14, ν = 4/3) or represents a new class driven by the non-local field influence.

Experiment design.

· Grid sizes: N = 40, 80, 120, 160, 200.

· Density ρ: from 0.08 to 0.22 in steps of 0.005 (30 points).

· Runs per point: 500 (total ≈ 75 000 simulations).

· Parameters fixed: p = 0.02, φ = 0.10, δ = 0.90, threshold = 3.5.

· Observables: probability of SR-mode, mean self-ref fraction, susceptibility χ(ρ) = Var(N)/Mean(N), Binder cumulant $U_4 = 1 - \langle M^4 \rangle / (3 \langle M^2 \rangle^2)$.

· Analysis: data collapse using scaling relations for 2D percolation; free fit of exponents if collapse fails.

Resources.

Approximately 75 000 runs; each run ≈ 150 steps. With parallel execution on a moderate cluster (e.g., 100 cores) the experiment would take about 1–2 weeks. A single high-end workstation would require several months.

Expected outcome.

Confirmation or refutation of the universality class; if exponents differ from 2D percolation, a new universality class will be established, significantly enhancing the theoretical impact.

B. Learnable Energy Function (Part I)

Rationale.

The energy weights w(recursive) = 0.7, w(bootstrap) = 0.8, etc., are heuristic. Their influence on ρ_c and p* must be tested, and a principled derivation from first principles (maximum entropy or active inference) would strengthen the model's foundations.

Experiment design.

- Replace fixed weights with trainable parameters.
- Three variants:
  - (a) Maximum entropy optimisation (Lagrange multipliers) to match observed self-ref ratios.
  - (b) Reinforcement learning (PPO) with reward = −F (free energy).
  - (c) Control: original heuristic weights.
- Compare ρ_c and p* across variants.

Resources.

Each variant requires ≈ 10 000 runs with gradient-based optimisation; total ≈ 30 000 runs. Moderate computational cost (days on a GPU cluster).

Expected outcome.

If ρ_c remains robust under learned weights, the universality of the threshold is confirmed. If significant shifts occur, the model will be refined by incorporating the optimised energy landscape.

C. Systematic Bit-Length Scan (Part II)

Rationale.

The scaling hypothesis optimal_density $\log_2(L)$ (proposed in Sec. 5.2.1) must be tested. Additionally, the threshold for bridge merger and the behaviour of identity retention as a function of dimensionality need systematic investigation.

Experiment design.

· Bit lengths: L = 64, 128, 256, 512, 1024 (5 values).

· For each L, repeat the New Experiment (merger probability vs. inter-group density 0.0–1.0 step 0.1) with 1000 trials per point.

· Three scenarios: USSR 1937, Weimar 1932, EU 2024.

· Additional measurements: irreversibility threshold C, identity retention E, stranger paradox ratio.

Resources.

5 bit lengths × 11 densities × 3 scenarios × 1000 trials = 165 000 simulations. Each simulation involves 90 steps (30+60) and 512-bit operations. Parallel execution on a cluster (100 cores) would take ≈ 2–3 weeks.

Expected outcome.

Verification of the scaling law optimal_density $\log_2(L)$ (predicted: for L=64 → ≈ 0.39, for L=128 → ≈ 0.46, etc.). This would provide a universal relation linking identity dimensionality to optimal network connectivity.

D. Correlated Identity Structures (Part II)

Rationale.

Real political identities exhibit correlations (e.g., clusters of related issues). Introducing such structure will test the robustness of the non-monotonic merger effect and allow comparison with empirical data (e.g., Pew Research polarization studies).

Experiment design.

· Replace independent random bits with clustered identities: K clusters (K = 4, 8, 16, 32), each of size L/K.

· Intra-cluster coupling strength λ (fraction of bits that align within a cluster) varied over {0.0, 0.2, 0.4, 0.6, 0.8}.

· For each (K, λ), repeat the New Experiment (merger vs. density) for the USSR scenario.

· Compare optimal density range, peak value, and shape with the independent-bits baseline.

Resources.

≈ 4 K values × 5 λ × 11 densities × 1000 trials = 220 000 simulations. Additional runs for validation against empirical polarization data (e.g., using real-world correlation matrices) would

increase the total.

Expected outcome.

Determination of whether the non-monotonic shape and optimal density range survive when realistic correlations are introduced. This would directly link the model to survey data and strengthen its external validity.

E. Unified Phase Diagram (Part I + Part II)

Rationale.

The conceptual link between the two frameworks (Sec. 5.2.1) deserves a formal unification. A hybrid model where each bit-agent lives on a cell of the cellular automaton, and actualisation occurs only when both local and global conditions are satisfied, would reveal whether the two transitions belong to the same scaling curve.

Experiment design.

- Hybrid model: each bit-agent occupies a site of the 40×40 lattice from Part I.
- Global parameter: interaction density ρ (from Part I) varied from 0.08 to 0.20.
- Local parameter: inter-group connection density varied from 0.0 to 1.0.
- 200 runs per (ρ, density) point, total ≈ 80 000 points.
- Outcome: 3D heat-map of global self-referential merger probability.

Resources.

≈ 80 000 hybrid simulations, each requiring both CA dynamics and agent updates. With parallel execution, this would take ≈ 1–2 months on a small cluster, but can be accelerated with GPU-aware code.

Expected outcome.

First unified phase diagram linking ontological (self-reference) and social (identity merger) transitions. If the two thresholds align after appropriate rescaling, a universal theory of high-dimensional collective behaviour could emerge.

F. Strategic Agents (Part II)

Rationale.

The current agents are purely physical (phase synchronization). Introducing strategic behaviour (e.g., payoff-based decisions) would move the model closer to political science and game theory, while preserving the core dynamics.

Experiment design.

· Introduce a small fraction (5%, 10%, 20%) of "strategic" agents with a payoff matrix (e.g., Tit-for-Tat, Always-Defect, Win-Stay-Lose-Shift).

· Strategic agents can choose to alter their bits based on expected utility rather than pure phase pressure.

· Measure shifts in ρ_c (Part I) and optimal merger density (Part II).

· Compare with active inference (Friston) frameworks.

Resources.

≈ 50 000 simulations covering varying fractions and payoff rules. Moderate computational cost.

Expected outcome.

Quantitative understanding of how strategic behaviour modifies the phase transitions. This could provide a bridge to models of political competition and cooperation.

Summary of Resource Requirements

Experiment	Simulations	Estimated Time (100 cores)	Hardware Needs
A. FSS + exponents	75 000	1–2 weeks	CPU cluster
B. Learnable E(v)	30 000	days	GPU cluster
C. Bit-length scan	165 000	2–3 weeks	CPU cluster
D. Correlated structures	220 000	3–4 weeks	CPU cluster
E. Unified phase diagram	80 000	1–2 months	CPU/GPU cluster
F. Strategic agents	50 000	1 week	CPU/GPU cluster

Total ≈ 620 000 simulations. With parallelisation on a modern high-performance computing centre (e.g., TU Dresden ZIH), these experiments could be completed in 2–3 months.

Invitation for Collaboration

The author is an independent researcher without access to large-scale computational facilities. Collaboration with a research institution that can provide the necessary computing resources is warmly invited. Interested parties are encouraged to contact the author directly.

Axelrod, R. (1986). An evolutionary approach to norms. American Political Science Review, 80(4), 1095–1111.

Barabasi, A.-L. (2002). Linked: The New Science of Networks. Perseus.

Berger, P., & Luckmann, T. (1966). The Social Construction of Reality. Doubleday.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. Reviews of Modern Physics, 81(2), 591–646.

DeGroot, M. H. (1974). Reaching a consensus. Journal of the American Statistical Association, 69(345), 118–121.

Easley, D., & Kleinberg, J. (2010). Networks, Crowds, and Markets. Cambridge UP.

Freedom House (2025). Freedom in the World 2025. freedomhouse.org

Friedkin, N. E., & Johnsen, E. C. (1990). Social influence and opinions. Journal of Mathematical Sociology, 15(3–4), 193–206.

Gramsci, A. (1971). Selections from the Prison Notebooks. International Publishers.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence. JASSS, 5(3).

ITU (2024). Measuring digital development: Facts and figures 2024. itu.int

MIPEX (2024). Migration Integration Policy Index. mipex.eu

RSF (2025). World Press Freedom Index 2025. rsf.org

Schelling, T.C. (1971). Dynamic models of segregation. Journal of Mathematical Sociology, 1(2), 143–186.

UNDP (2024). Human Development Report 2024. hdr.undp.org

V-Dem Institute (2025). Democracy Report 2025. v-dem.net

Watts, D.J. (2003). Six Degrees: The Science of a Connected Age. Norton.

World Bank (2024). World Development Indicators. data.worldbank.org

World Inequality Database (2024). wid.world