

**On Faultization:  
Pigeonhole Principle.  
What Perturbation Reveals About Pattern  
Access Under Irreducible Constraint**

PIATRA . INSTITUTE

March 2026

## Abstract

We apply faultization (a systematic regime of morphogenetic perturbation) to the pigeonhole principle reinterpreted as a distributed multi-agent system, asking which patterns from the latent space (Levin, 2026) the system accesses under irreducible constraint. The pigeonhole principle is itself a pattern in the Platonic Space: the theoretical minimum overload  $O_{\min} = m - n$  is a mathematical truth that the system channels without computing. In our model,  $m = 10$  autonomous pigeons self-organize into  $n = 7$  holes under four local placement policies (GREEDY, EXPLORATORY, REPULSIVE, COOPERATIVE), with no centralized controller. Because  $m > n$ , overload  $O \geq m - n = 3$  is irreducible; the system cannot solve the problem, only manage it. We subject this system to eight perturbation experiments ( $n = 30$  replications, 500 steps, paired  $t$ -tests with matched seeds): frozen-hole robustness, policy comparison, noisy perception, view-radius sweep, chimeric mixed-policy populations, dynamic damage and recovery, progressive versus sudden damage, and misleading (deceptive) holes. We classify findings into four categories. *Pattern manifestation*: the mathematical pattern  $O_{\min}$  manifests through the interface in six of eight experiments, constituting a free lunch (optimal global coordination without global optimization). *Pattern fidelity*: perceptual noise causes immediate monotonic degradation with no tolerance threshold ( $\rho = +0.638$ ,  $p < 0.0001$ ), revealing that discrete interfaces require high-fidelity perception to transmit the pattern. *Pattern corruption*: misleading holes invert the pattern-seeking mechanism, with a single misleading hole capturing 36% of pigeons and 66% of overload and increasing total overload by 25.6%. *Free lunch*: the system receives optimal overload, damage compensation, recovery, and chimeric self-organization without paying the computational cost of global optimization. Policies that are identical at the endpoint differ dramatically in failure persistence, with same-target retry after rejection ranging from 0.463 (REPULSIVE) to 0.997 (COOPERATIVE), indicating that the convergence is pattern-channeling rather than agent-level learning. Delayed gratification is absent in all conditions (DG Index = 0.0), consistent with the interpretation that convergence is driven by the pattern, not by accumulated experience.

## 1. Introduction

Levin (2026) argues that physical systems serve as interfaces for a non-physical space of patterns, the Platonic Space, whose denizens are discovered, not created, and are causal in physics, biology, and computer science. Under this framework, evolution exploits mathematical patterns as affordances it does not need to pay for: “once you find a voltage-gated ion channel, you have a transistor which can make logic gates, and truth tables are yours for free.” The Platonic Space offers what Levin calls “free lunches”: useful patterns for which the physical processes of learning, evolution, and engineering do not need to pay, or pay some but receive much more than the effort they put in.

The pigeonhole principle provides the cleanest possible test of this framework. The principle states that if  $m$  items are placed into  $n$  containers and  $m > n$ , then at least one container must hold more than one item. Formally, for any function  $f : A \rightarrow B$  with  $|A| > |B|$ ,  $f$  cannot be injective. The theoretical minimum overload  $O_{\min} = m - n$  is a mathematical truth, a pattern in the Platonic

Space. Under the Platonic Space framework, the question becomes: when a distributed system of memoryless agents faces this irreducible constraint, does the mathematical truth  $O_{\min}$  manifest through the physical interface? And if so, what does that manifestation cost?

The answer, as we show, is that it costs nothing. Local policies with no representation of the global optimum, no communication channel, and no memory reliably converge to  $O_{\min} = 3$  in six of eight experimental conditions. No agent computes the global optimum; the pattern provides it for free. This is a free lunch in Levin’s precise sense: the system receives optimal global coordination without paying the computational cost of global optimization.

Yet the pigeonhole principle describes a situation ubiquitous in distributed systems: too many agents competing for too few resources. Cells competing for niches in a tissue, packets competing for bandwidth in a network, vehicles competing for lanes on a highway, job seekers competing for positions in a labor market, all face pigeonhole-type constraints where some degree of conflict is structurally inevitable. In each of these settings, the mathematical impossibility of conflict-free allocation is a given. The operationally important question is: how faithfully does the physical interface transmit the mathematical truth  $O_{\min}$ , and what happens when we degrade that interface?

Faultization, systematic morphogenetic perturbation, is the methodology for answering this question. By perturbing the interface between algorithm and pattern space, we probe what patterns survive degradation, what fidelity the interface requires, and what happens when the interface actively misleads. The morphogenetic perturbation protocol, developed by Zhang, Goldstein, and Levin (2024) for sorting algorithms and extended by Kofman, Bhatt, and Levin (2025) to transformer training, proceeds in three stages: (1) formalize the target process as a decentralized system with explicit local policies, (2) systematically perturb the system by breaking assumptions of the nominal process, and (3) classify the resulting behaviors. We apply all three stages to the pigeonhole principle across eight experiments.

The unique advantage of the pigeonhole system for testing the Platonic Space framework is that the pattern is a known theorem. In sorting, the “pattern” is the correct permutation; in transformer training, it is the loss landscape structure. In both cases, the pattern is complex and domain-specific. In the pigeonhole system, the pattern is  $O_{\min} = m - n$ : a single number derived from a theorem every undergraduate knows. This makes free-lunch quantification precise. We can state exactly what the system receives (optimal overload) and exactly what it pays (local greedy computation with no global information). The gap between payment and receipt is the free lunch.

We classify findings into four categories that map directly onto the Platonic Space framework:

1. **Pattern manifestation:** the mathematical pattern  $O_{\min}$  successfully manifests through the interface. The system converges to the theoretical minimum overload using only local rules, channeling a global mathematical truth without computing it.
2. **Pattern fidelity:** how much interface degradation the pattern tolerates before failing to manifest. Different interface types (discrete vs continuous) have different fidelity requirements for the same kind of pattern.
3. **Pattern corruption:** the interface actively misleads, inverting the pattern-seeking mechanism.

The same machinery that efficiently finds the optimum now efficiently finds the wrong target.

4. **Free lunch:** what the system receives without paying for. Optimal overload, damage compensation, recovery, and self-organization all arrive without global optimization.

Our contributions are as follows:

- We provide the first test of Levin’s Platonic Space framework using a system where the pattern is a known mathematical theorem ( $O_{\min} = m - n$ ), enabling precise free-lunch quantification.
- We formalize the pigeonhole principle as a distributed multi-agent system with four local placement policies, a composite potential function, and a hook-based perturbation architecture that enables systematic intervention at every decision point.
- We conduct eight systematic experiments spanning substrate damage, perceptual noise, information radius, policy heterogeneity, dynamic damage/recovery, and substrate deception, with paired statistical analysis and new process metrics for post-failure persistence and deceptive-substrate bias.
- We demonstrate that local policies achieve global optimality in six of eight conditions, constituting a free lunch: the system channels the mathematical truth  $O_{\min}$  without computing it.
- We show that pattern corruption (misleading holes) is categorically more harmful than pattern unavailability (frozen holes), and that the same pattern-seeking mechanism that produces the free lunch becomes a liability under deception.
- We identify the absence of stress inoculation as evidence for pure pattern-channeling: because the memoryless system has no history to exploit, its convergence must come from the pattern, not from accumulated experience.
- We provide the first morphogenetic perturbation analysis of an impossibility theorem, extending the framework from solvable problems (sorting, training) to fundamentally unsolvable ones.

## 2. Related Work

### 2.1 Levin’s Platonic Space Framework

Levin (2026) proposes that physical systems serve as interfaces for a Platonic Space of non-physical patterns that are discovered, not created, and are causal in physics, biology, and computer science. The framework rests on several claims. (1) Some facts are not physical facts (truths of topology, the distribution of primes, the contents of discrete mathematics) and cannot be found using the tools of physics. (2) These patterns are causal in Judea Pearl’s counterfactual sense: if the prime distribution were otherwise, cicadas would emerge at different years. (3) Physical bodies (embryos, cyborgs, computers) function as interfaces through which the patterns manifest. (4) The Platonic Space offers “free lunches”: useful patterns for which physical processes do not need to pay, because “no free lunch” commitments are derived from the laws of the physical world, not from the pattern space.

The framework suggests a specific research program: build interfaces and study what unexpected patterns ingress through them that are not well-explained by their history of selection, engineering, or learning from experience. Levin notes that “biology offers the most sophisticated patterns, but it’s very hard to prove anything in biology, so we’re also making minimal computational models where we can more easily quantify the effort put in and the outcome we observed.” Our pigeonhole system is precisely such a minimal computational model: the effort put in (local greedy policies with no global information) and the outcome observed (convergence to the theoretical minimum  $O_{\min} = m - n$ ) are both exactly quantifiable, making free-lunch measurement precise.

## 2.2 Morphogenetic Perturbation of Algorithms

Zhang, Goldstein, and Levin (2024) introduced the methodology of treating classical algorithms as morphogenetic systems: decentralized collectives of elements executing local policies under imperfect conditions. Applied to sorting, this approach revealed that even minimal systems exhibit error correction, damage compensation, delayed gratification, and chimeric self-organization behaviors not specified by the original algorithm. Under the Platonic Space framework, these competencies can be reinterpreted as pattern manifestation: the correct sort order is a pattern that the system channels through the interface of local comparison-swap rules, and faultization probes how faithfully the interface transmits that pattern under degradation.

Kofman, Bhatt, and Levin (2025) extended this to transformer training, showing that gradient descent under morphogenetic perturbation (noisy gradients, frozen parameters, weight surgery) exhibits analogous competencies. They established the morphogenetic perturbation protocol: systematically break assumptions of the nominal process, record full trajectories, and classify the resulting behaviors. Under the Platonic Space interpretation, gradient descent is an interface through which the loss landscape’s structure, itself a pattern, manifests in the system’s weights.

Our work extends this line to a qualitatively different setting: an impossibility theorem where the pattern is a known mathematical truth. While sorting and training are solvable problems (there exists a correct sort, there exists a loss minimum), the pigeonhole principle with  $m > n$  is inherently unsolvable. The pattern  $O_{\min} = m - n$  is not a solution but a constraint, a mathematical fact about the minimum achievable conflict. This extension tests whether the Platonic Space framework produces meaningful results when the pattern is not a target to reach but a boundary that cannot be crossed.

## 2.3 Multi-Agent Resource Allocation

The pigeonhole system is formally a congestion game (Rosenthal, 1973): agents select resources (holes), and payoffs decrease with congestion. Nash equilibria of congestion games are well-characterized, and the price of anarchy provides worst-case bounds on decentralized performance relative to the social optimum (Roughgarden and Tardos, 2002). Our system differs in that the social optimum itself represents an irreducible conflict state, not a conflict-free solution. In standard congestion games, the social optimum is typically conflict-free or conflict-minimal; in the pigeonhole system, even the optimal state has overload  $\geq m - n$ .

Load balancing in distributed systems (Azar et al., 1999) addresses the allocation of  $m$  tasks to  $n$  processors under local information. The “power of two choices” result shows that sampling just two random processors (rather than one) and choosing the less loaded one reduces maximum load from  $\Theta(\log n / \log \log n)$  to  $\Theta(\log \log n)$ . Our view-radius experiment (Experiment 4) provides a pigeonhole analogue of this phenomenon: even a small increase in the number of visible holes dramatically accelerates convergence.

The distributed task allocation literature also addresses fault tolerance (Aguilera et al., 2004), where processors may crash or behave adversarially. Our frozen-hole and misleading-hole experiments map directly onto crash faults (silent failure) and Byzantine faults (deceptive failure), respectively. Under the Platonic Space framework, these map onto pattern unavailability (the interface loses capacity but does not mislead) and pattern corruption (the interface actively inverts the pattern-seeking mechanism).

## 2.4 Basal Cognition Framework

Levin (2019, 2022) argues that cognitive competencies exist on a continuum: from molecular networks solving constraint satisfaction problems, through cellular collectives achieving morphogenetic goals, to neural systems supporting behavioral intelligence. The key claim is that goal-directedness, error correction, and adaptive replanning are not exclusive to neural systems but appear at every scale of biological organization. Under the Platonic Space framework (Levin, 2026), these competencies can be understood as pattern manifestation: the “goal” is a pattern in the Platonic Space, and the biological system is an interface through which that pattern manifests with varying degrees of fidelity.

Our pigeonhole system instantiates this framework at the simplest possible level: agents with no memory, no communication channel, and no model of other agents, channeling a known mathematical truth ( $O_{\min} = m - n$ ) through purely local policies. The impossibility constraint adds a distinctive feature: the target state (zero overload) is unreachable, so the system must manifest a pattern (minimum overload) that it has no explicit representation of. If the system reliably reaches  $O_{\min}$  despite having no mechanism to compute it, that constitutes evidence for pattern-channeling in Levin’s sense.

## 3. Methods

### 3.1 System Specification

The system consists of  $m = 10$  pigeons and  $n = 7$  holes. Each pigeon  $k \in \{1, \dots, m\}$  occupies a hole  $x_k \in \{0, 1, \dots, n\}$ , where  $x_k = 0$  denotes the unplaced state. The load of hole  $i$  is  $\ell_i = |\{k : x_k = i\}|$ . The overload is defined as:

$$O(\mathbf{x}) = \sum_{i=1}^n \max(0, \ell_i - 1)$$

Since all pigeons must be placed and  $m > n$ , we have  $O \geq m - n = 3$  for any fully-placed state. Under the Platonic Space framework, this bound is the pattern that the system channels: a mathematical truth about the irreducible minimum conflict.

**Hole statuses.** Each hole has one of three statuses: - **ACTIVE**: accepts pigeons and reports true load. - **FROZEN**: silently rejects all placement attempts (pattern unavailability, the interface loses capacity). - **MISLEADING**: accepts pigeons but reports load as 0 regardless of true occupancy (pattern corruption, the interface actively misleads).

**Potential function.** The system minimizes:

$$\Phi(\mathbf{x}) = \alpha \sum_{i=1}^n \max(0, \ell_i - 1) + \beta \sum_{i=1}^n [\max(0, \ell_i - 1)]^2 + \gamma \cdot U(\mathbf{x})$$

where  $U(\mathbf{x}) = |\{k : x_k = 0\}|$  is the count of unplaced pigeons, and  $\alpha = 1.0$ ,  $\beta = 0.5$ ,  $\gamma = 10.0$ . The first term penalizes total overload, the second penalizes concentrated overload (encouraging even distribution of excess), and the third strongly penalizes leaving pigeons unplaced.

**Local policies.** Each pigeon executes one of four policies based on local information. These are four different access methods for the same underlying pattern, four ways the interface can channel  $O_{\min}$ :

1. **GREEDY**: Inspect visible holes (determined by view radius  $r$ ), move to the hole with lowest observed load if it is lower than the load of the pigeon's current hole. This is the simplest reasonable policy: purely self-interested, myopic load minimization.
2. **EXPLORATORY**: Sample twice as many holes as the view radius permits ( $\min(2r, n)$ ), then select the best. This policy trades computation (more sampling) for information, analogous to the "exploration vs exploitation" tradeoff in multi-armed bandits.
3. **REPULSIVE**: If the pigeon's current hole has load  $\ell_i > 1$ , escape with probability  $(\ell_i - 1)/\ell_i$  and relocate greedily. If  $\ell_i = 1$ , stay. This policy implements a stochastic crowd-avoidance heuristic: the more crowded the hole, the more likely the pigeon is to flee.
4. **COOPERATIVE**: Evaluate the change in global potential  $\Delta\Phi$  for each candidate move; move only if  $\Delta\Phi < 0$ . This is the only policy that considers the system-wide effect of its action, though it still acts locally (it cannot coordinate with other pigeons).

**Architecture.** The system uses a hook-based perturbation architecture modeled on the morphogpt framework (Kofman et al., 2025). Four named hook points provide interception at every stage of the pigeon decision cycle: - **pigeon\_view**: intercepts the list of visible (hole, load) pairs, enabling perturbations to perceived loads (noise, blindness, inversion). - **pigeon\_decision**: intercepts the chosen target hole, enabling perturbations to policy output (stubbornness, randomization, contrarianness). - **placement\_attempt**: intercepts the (pigeon, hole) pair before the frozen-hole check, enabling perturbations to the placement mechanism itself. - **post\_step**: intercepts the full system state after each step, enabling dynamic substrate changes (hole breaking, healing, progressive damage).

Each hook receives the current value and the step number, and returns either a modified value or None (pass-through). Multiple hooks can be chained on the same hook point. A probe records the complete trajectory: overload, potential, maximum load, unplaced count, load snapshots, and all move attempts with success/failure status. This architecture ensures that every experiment is implemented as a composition of hooks on the same base system, guaranteeing that all conditions share identical core dynamics.

**Dynamics.** At each step, one pigeon is activated uniformly at random, inspects a subset of holes determined by its view radius, applies its policy, and attempts placement. Frozen holes reject silently; misleading holes accept but misreport. The system runs for 500 steps after an initial random placement. This asynchronous activation scheme models real-world distributed systems where agents act independently and without synchronization.

**Theoretical minimum overload.** For a system with  $n_u$  usable (non-frozen) holes and  $m$  pigeons, the theoretical minimum overload is:

$$O_{\min}^{\text{theo}} = \max(0, m - n_u)$$

This follows from the observation that the maximum number of holes that can each hold exactly one pigeon is  $\min(m, n_u)$ , leaving  $m - \min(m, n_u)$  excess pigeons when  $m > n_u$ . Misleading holes remain usable even though they corrupt perception. All experiments use this quantity as the normalization baseline for the overload ratio. Under the Platonic Space framework,  $O_{\min}^{\text{theo}}$  is the pattern, the mathematical truth that the system either channels successfully or fails to channel.

**State space size.** The full state space has  $(n + 1)^m$  configurations (each pigeon in one of  $n$  holes or unplaced). For  $m = 10, n = 7$ , this is  $8^{10} \approx 10^9$ . The optimal states form a much smaller subset: any configuration where all  $n$  holes are occupied and the  $m - n = 3$  excess pigeons are distributed arbitrarily constitutes a minimum-overload state. The number of such states is  $\binom{m}{n} \cdot n! \cdot S(m - n, n, n)$  where  $S$  accounts for the distribution of excess, but the precise count matters less than the qualitative observation: optimal states are a tiny fraction of the state space, yet the system reliably finds them. This is the free lunch: from  $\sim 10^9$  possible states, the system converges to the optimal subset without searching it.

### 3.2 Statistical Protocol

All experiments use  $n_{\text{rep}} = 30$  replications with matched random seeds across conditions. The primary statistical test is the paired  $t$ -test (two-tailed), which exploits seed-matching to control for initialization variance. Effect sizes are reported as Cohen’s  $d$  (paired: mean difference divided by standard deviation of differences). Monotonic relationships are assessed with Spearman’s rank correlation  $\rho$ . Significance thresholds follow convention: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

For Experiment 5 (chimeric policies), where different conditions involve different random policy assignments, we use Welch’s independent  $t$ -test for between-pair comparisons.

The choice of paired  $t$ -tests (rather than independent-samples tests) is motivated by the seed-matching design: each condition is run with seeds 0, 1, ..., 29, and the same seed produces the same initial random placement and activation sequence across conditions. This eliminates between-run variance due to initialization, substantially increasing statistical power. The pairing is particularly important for experiments where the effect of interest is small (e.g., Experiment 7, where overload is identical across conditions but failed placements differ).

We do not apply multiple-comparison corrections (e.g., Bonferroni) across experiments because each experiment tests a distinct hypothesis with distinct data. Within-experiment multiple comparisons (e.g., the 6 comparisons in Experiment 1) are noted but not corrected, as the primary findings involve effects with  $p < 0.0001$  that would survive any reasonable correction.

All statistical computations use SciPy 1.x (`scipy.stats.ttest_rel` for paired tests, `scipy.stats.ttest_ind` with `equal_var=False` for Welch's test, `scipy.stats.spearmanr` for rank correlations). Effect sizes use pooled standard deviation for Cohen's  $d$ .

### 3.3 Metrics

**Overload  $O$ :** the primary outcome metric, counting excess pigeons beyond one-per-hole capacity. Formally,  $O = \sum_{i=1}^n \max(0, \ell_i - 1)$ . For a fully-placed state with  $m > n_u$  (usable holes), the minimum achievable overload is  $m - n_u$ .

**Overload ratio  $O/O_{\min}^{\text{theo}}$ :** final overload normalized by the theoretical minimum given the number of usable (non-frozen) holes. A ratio of 1.0 indicates that the mathematical pattern  $O_{\min}$  has fully manifested through the interface.

**Post-failure same-target retry:** for every failed placement that is followed by another attempted placement by the same pigeon, we ask whether the next attempted target is the same faulty hole. High values indicate retry persistence and argue against agent-level learning from rejection.

**Post-failure repeat failure rate:** for every failed placement that is followed by another attempted placement by the same pigeon, we ask whether that next attempt also fails. This captures persistence on faulty substrate even when the pigeon switches from one frozen hole to another.

**Convergence step:** the last simulation step at which overload changes, indicating when the system reaches its final configuration.

**Failed placements:** the total number of placement attempts rejected by frozen holes over the full run. A process-level metric that distinguishes policies even when outcomes are identical.

**Misleading occupancy share / bias:** the fraction of pigeons occupying misleading holes at the end of the run, and that share minus the fraction of holes that are misleading. Positive bias means the corrupted interface attracts more load than its spatial share would predict.

**Misleading overload share / bias:** the fraction of total overload concentrated on misleading holes, and that share minus the fraction of holes that are misleading. This measures whether the corruption merely perturbs the system or actively captures the excess load it creates.

**Misleading load gap:** mean load on misleading holes minus mean load on honest holes.

**Policy aggregation score:** for chimeric (mixed-policy) populations, the fraction of co-located pigeons sharing the majority policy in each multiply-occupied hole. Computed as follows: for each hole with  $\geq 2$  pigeons, count the pigeons belonging to the majority policy; sum these across all such holes and divide by the total number of pigeons in multiply-occupied holes. Values near 1.0 indicate strong same-policy clustering; values near  $1/k$  (where  $k$  is the number of distinct policies) indicate random mixing. For  $k = 2$  policies, the chance level is approximately 0.5.

**Delayed Gratification (DG) Index:** retained for compatibility with the sorting literature, measuring episodes where overload temporarily increases and later falls below the prior local minimum. In this system it is uniformly zero, consistent with the interpretation that convergence is driven by the pattern, not by accumulated experience.

### 3.4 Experiment Design

The eight experiments are designed to probe the interface between algorithm and pattern space along five perturbation axes: substrate integrity (Experiments 1, 6, 7), policy diversity (Experiments 2, 5), perceptual accuracy (Experiments 3, 8), information scope (Experiment 4), and temporal dynamics (Experiments 6, 7). Each experiment varies one axis while holding others constant, enabling attribution of effects to specific interface degradation types. Table 0 provides the complete experimental design.

**Experiment 1, Frozen Hole Robustness.** Freeze 0 through 6 holes (of 7 total), measuring overload ratio together with post-failure persistence. Tests the pattern bandwidth of the interface: how much substrate loss before the pattern  $O_{\min}$  fails to manifest.

**Experiment 2, Policy Comparison.** All four policies under identical conditions ( $m = 10, n = 7, 1$  frozen hole). Tests pattern plurality: whether different access methods channel the same underlying pattern.

**Experiment 3, Noisy Perception.** Add Gaussian noise  $\mathcal{N}(0, \sigma^2)$  to perceived hole loads, with  $\sigma \in \{0, 0.5, 1.0, 2.0, 5.0\}$ . The noise is applied via the `pigeon_view` hook, adding independent Gaussian noise to each perceived load and clamping to non-negative integers. Tests pattern fidelity: what perceptual accuracy the discrete pigeonhole interface requires to transmit the pattern.

**Experiment 4, View Radius Sweep.** Vary the number of holes each pigeon can inspect per step:  $r \in \{1, 2, 3, 5, 7\}$ . At  $r = 1$ , each pigeon sees exactly one randomly sampled hole per step; at  $r = 7$  (the total number of holes), each pigeon has full information. Tests the information geometry of pattern access: how much of the interface must be visible for the pattern to manifest, and how visibility affects convergence speed.

**Experiment 5, Chimeric Policies.** Mixed-policy populations ( $m = 12, n = 8$ ), with each pigeon independently assigned one of two policies with equal probability. Four policy pairs are tested: GREEDY+COOPERATIVE, GREEDY+EXPLORATORY, EXPLORATORY+COOPERATIVE, and

REPULSIVE+COOPERATIVE. Tests lateral pattern resonance: whether mixed-policy populations access the same pattern and whether same-policy agents cluster spatially as a free lunch.

**Experiment 6, Recovery After Damage.** Three conditions on the same system ( $m = 10, n = 7$ ): (a) control with no damage, (b) freeze hole 0 at step 167, (c) freeze hole 0 at step 167 and heal it at step 333. Tests the bidirectionality of the pattern-substrate interface: whether the pattern re-manifests after both damage and healing.

**Experiment 7, Progressive vs Sudden Damage.** Freeze 3 holes (of 7) either all simultaneously at step 100 (sudden) or one at each of steps 100, 200, and 300 (gradual). Tests whether gradual exposure smooths disruption cost, and whether the absence of stress inoculation confirms pure pattern-channeling.

**Experiment 8, Misleading Holes.** Holes that report their load as 0 regardless of true occupancy, implemented via the MISLEADING hole status in the model. Vary from 0 to 6 misleading holes (of 7 total). Tests pattern corruption: what happens when the interface does not merely degrade but actively inverts the pattern-seeking mechanism.

## 4. Results

We present results for all eight experiments. Each experiment was run with  $n = 30$  replications, 500 steps per run, and matched random seeds across conditions. We report means, significance levels ( $p$ -values from paired  $t$ -tests), and effect sizes (Cohen’s  $d$ ). For monotonic relationships, we report Spearman’s rank correlation  $\rho$ . Each subsection concludes with a classification under the Platonic Space framework. Table 1 provides a cross-experiment summary; detailed results follow in subsections 4.1–4.8.

**Table 1. Cross-experiment summary.** All tests: two-tailed paired  $t$ -test,  $n = 30$ , seeds matched across conditions. Overload ratio = 1.0 indicates optimal performance.

Experiment	Key Condition	Primary Metric	Value	$\Delta\%$	$p$	$d$
1: Frozen Robustness	frozen_3	Overload ratio	1.000	0.0	n.s.	0.0
1: Frozen Robustness	frozen_5	Repeat failure	1.000	,	,	,
2: Policy Comparison	EXPLORATORY	Same-target retry	0.940	+73.3	< 0.0001	+10.377
3: Noisy Perception	$\sigma = 1.0$	Overload	3.53	+17.8	< 0.0001	+1.051
4: View Radius	$r = 2$	Convergence	6.0	-48.7	0.0093	-0.509
5: Chimeric	REP+COOP	Aggregation	0.796	,	,	,
6: Recovery	damage_and_heal	Same-target retry	0.226	,	,	,

Experiment	Key Condition	Primary Metric	Value	$\Delta\%$	$p$	$d$
7: Progressive	gradual	Repeat failure	0.639	-14.1	< 0.0001	-2.153
8: Misleading	misleading_2	Occupancy bias	0.274	+27.4pp	< 0.0001	+1.920

#### 4.1 Experiment 1, Frozen Hole Robustness: Pattern Bandwidth

Freezing holes reduces the number of usable holes from  $n$  to  $n-k$ , raising the theoretical minimum overload from  $m-n$  to  $m-(n-k) = m-n+k$ . The system must redistribute pigeons across fewer usable holes. Under the Platonic Space framework, this experiment probes the bandwidth of the interface: how much substrate can be removed before the pattern  $O_{\min}$  fails to manifest.

Condition	Final Overload	Overload Ratio	Same-Target Retry	Repeat Failure
frozen_0	3.0	1.000	0.000	0.000
frozen_1	4.0	1.000	0.542	0.542
frozen_2	5.0	1.000	0.486	0.985
frozen_3	6.0	1.000	0.309	0.929
frozen_4	6.47	0.924	0.247	0.985
frozen_5	2.50	0.313	0.198	1.000
frozen_6	1.57	0.174	0.162	1.000

**Key finding.** The mathematical truth  $O_{\min}$  manifests perfectly through the degraded interface for 0 through 3 frozen holes (overload ratio = 1.0), demonstrating that the pattern survives up to 43% substrate damage. This is a free lunch: optimal overload despite severe interface degradation. No agent computes the global optimum; the system channels the mathematical truth  $O_{\min}$  even when nearly half the substrate is destroyed. At 4 frozen holes (57% damage), the pattern begins to fail (ratio = 0.924); at 5–6 frozen holes, the interface is too degraded for the pattern to manifest. Spearman correlation between frozen holes and overload ratio:  $\rho = -0.1498$ ,  $p = 0.030^*$ .

The process metrics show that this pattern manifestation is not driven by agent-level learning. With one frozen hole, pigeons retry the same faulty target on their next attempted move 54.2% of the time. As more holes freeze, same-target retry falls because there are fewer distinct usable targets, but repeat failure rises to nearly 1.0: the system keeps colliding with faulty substrate even when it no longer insists on the exact same hole. The convergence to  $O_{\min}$  is pattern-channeling, not learning.

Note on the overload ratio: the ratio *decreases* with more frozen holes because the theoretical minimum grows faster than the actual overload. At frozen\_5, the theoretical minimum is  $m-(n-5) = 10-2 = 8$ , but actual overload is only 2.50; the system leaves many pigeons unable to find any active hole. At frozen\_6, only one hole is active and can hold at most one pigeon, so 9 of 10 pigeons

must be either unplaced or crowded into a single hole. The system’s failure at this extreme reflects the physical impossibility of placing 10 pigeons into 1 hole efficiently, not a failure of pattern access.

The robustness curve has a clear phase transition between frozen\_3 (perfect performance) and frozen\_4 (significant degradation). At frozen\_3, the system still has 4 usable holes for 10 pigeons, giving a density of  $10/4 = 2.5$  pigeons per hole and theoretical minimum overload of 6. At frozen\_4, only 3 usable holes remain ( $10/3 = 3.33$  density), and the pattern begins to fail to manifest. This transition point at roughly 50% substrate damage defines the pattern bandwidth of the interface.

**Classification.** Perfect pattern manifestation at 0–3 frozen holes constitutes a *free lunch*: local policies channel the global mathematical truth  $O_{\min}$  despite substrate damage, with no global computation. The degradation at 4+ frozen holes marks the *pattern bandwidth limit*: the interface has lost too much capacity to transmit the pattern faithfully. The critical threshold at approximately 43% substrate damage defines the system’s pattern bandwidth.

#### 4.2 Experiment 2, Policy Comparison: Pattern Plurality

All four policies were tested with  $m = 10$ ,  $n = 7$ , and 1 frozen hole (theoretical minimum overload = 4). Under the Platonic Space framework, this experiment tests pattern plurality: whether different access methods channel the same underlying mathematical truth.

Policy	Final Overload	Failed Placements	Same-Target Retry	Convergence Step
GREEDY	4.0	212.1	0.542	7.4
EXPLORATORY	4.0	460.3	0.940	39.8
REPULSIVE	4.0	84.6	0.463	8.8
COOPERATIVE	4.0	168.3	0.997	7.6

#### Same-target retry vs GREEDY baseline:

Policy	$\Delta\%$	$p$	$d$
EXPLORATORY	+73.3%	< 0.0001***	+10.377
REPULSIVE	-14.6%	< 0.0001***	-1.437
COOPERATIVE	+83.8%	< 0.0001***	+12.963

**Key finding.** All four policies reach the identical final overload of 4.0 (the theoretical minimum) with zero variance across all 30 replications, every single run, regardless of policy, converges to exactly  $O_{\min}$ . The pattern is invariant to the access method. Four different interfaces, four different local rules with radically different process signatures, all channel the same mathematical truth with probability 1. This is pattern plurality: the pattern exists independently of the access method, and

any reasonable local policy can manifest it. The free lunch is identical regardless of how the system accesses it.

However, the process differs enormously. EXPLORATORY generates 117% more failed placements than GREEDY ( $d = +21.692$ ), because its wider sampling increases the probability of encountering the frozen hole. REPULSIVE generates 60% fewer failures ( $d = -11.116$ ) because its crowd-avoidance heuristic naturally steers away from frozen holes. The failure-persistence metric sharpens this story: after rejection, EXPLORATORY retries the same frozen hole 94.0% of the time and COOPERATIVE 99.7% of the time, whereas REPULSIVE falls to 46.3%.

The extraordinarily large effect sizes for failed placements ( $d = +21.692$  for EXPLORATORY) deserve comment. In most behavioral experiments,  $d > 0.8$  is considered “large.” The values here exceed that threshold by an order of magnitude, indicating that the policies produce almost non-overlapping distributions of failed placements. This extreme separation occurs because the failure mechanism is deterministic given the seed: once the initial placement and activation sequence are fixed, the number of times a policy directs a pigeon toward the frozen hole is a stable function of the policy’s sampling behavior.

The convergence speed difference between EXPLORATORY (39.8 steps) and the other policies (7.4–8.8 steps) is also notable. EXPLORATORY converges slowly not because it fails to find the optimum, but because its broader sampling means it continues to detect and respond to small load imbalances for longer. Once overload reaches the minimum, EXPLORATORY pigeons still see alternative holes and repeatedly attempt moves that fail at the frozen hole, prolonging the convergence tail. The COOPERATIVE policy is even more revealing: it reaches the optimum quickly, but once it has identified the frozen hole as the locally best-looking low-load option, it nearly always retries it after rejection. This is pattern manifestation without agent-level learning: the pattern provides the optimum for free, but the agents pay wasteful process costs because they have no memory of past failures.

**Classification.** The outcome equivalence across policies constitutes *pattern plurality*: the mathematical truth  $O_{\min}$  manifests through four different access methods, confirming that the pattern is invariant to the interface’s local policy. The dramatic divergence in failed placements and post-failure persistence reveals that different access methods have different process costs for the same free lunch.

### 4.3 Experiment 3, Noisy Perception: Pattern Fidelity in Discrete Domains

Gaussian noise  $\mathcal{N}(0, \sigma^2)$  was added to all perceived hole loads, rounding to non-negative integers. Under the Platonic Space framework, this experiment probes pattern fidelity: how much perceptual degradation the interface tolerates before the pattern fails to manifest.

Noise $\sigma$	Final Overload	$\Delta\%$ vs $\sigma = 0$	$p$	$d$
0.0	3.0	,	,	,
0.5	3.37	+12.2%	0.0011**	+0.659

Noise $\sigma$	Final Overload	$\Delta\%$ vs $\sigma = 0$	$p$	$d$
1.0	3.53	+17.8%	< 0.0001***	+1.051
2.0	3.87	+28.9%	< 0.0001***	+1.378
5.0	4.30	+43.3%	< 0.0001***	+1.851

Spearman correlation (noise level vs overload):  $\rho = +0.638$ ,  $p < 0.0001$ \*\*\*.

**Key finding.** The discrete pigeonhole interface requires high-fidelity perception to transmit the pattern  $O_{\min}$ . There is no noise buffer: even  $\sigma = 0.5$  produces a statistically significant 12.2% increase in overload ( $p = 0.0011$ ). This contrasts sharply with transformer training (Kofman et al., 2025), where gradient noise up to  $\sigma = 0.01$  is tolerated before degradation begins. Under the Platonic Space framework, the difference reflects the interface type: continuous systems (like gradient descent on a loss landscape) have averaging mechanisms that absorb small perturbations, providing a noise buffer for pattern transmission. Discrete placement systems with only 7 holes have no such buffer, even small noise can redirect a pigeon to the wrong hole, and there is no averaging mechanism to recover. The same kind of pattern (a mathematical optimum) requires different fidelity from different interface types.

The effect sizes grow from medium ( $d = +0.659$  at  $\sigma = 0.5$ ) to large ( $d = +1.851$  at  $\sigma = 5.0$ ), following a concave trajectory: the marginal degradation per unit of noise decreases at higher noise levels. This is consistent with a floor effect: as noise increases, the system’s placement decisions approach random assignment, and there is a limit to how bad random placement can be (the expected overload under uniform random assignment is finite and bounded).

The linear structure of the degradation is informative: a regression of final overload on  $\sigma$  yields  $O \approx 3.0 + 0.26\sigma$  ( $R^2 = 0.96$ ). Each unit of noise standard deviation adds approximately 0.26 units of overload. This provides a quantitative measure of the interface’s pattern fidelity: for any noise level, the expected pattern degradation can be estimated as roughly one-quarter of the noise magnitude.

**Classification.** The absence of a noise-tolerance threshold reveals a *pattern fidelity constraint* of discrete interfaces: the pigeonhole system has no noise buffer for pattern transmission. The monotonic relationship ( $\rho = +0.638$ ) confirms that fidelity degradation is smooth and predictable. Under the Platonic Space framework, this demonstrates that different interface types (discrete vs continuous) carry categorically different fidelity requirements for channeling the same kind of mathematical pattern.

#### 4.4 Experiment 4, View Radius Sweep: Information Geometry of Pattern Access

The view radius  $r$  determines how many holes each pigeon can inspect per step. Under the Platonic Space framework, this experiment probes the information geometry of pattern access: how much of the interface must be visible for the pattern to manifest.

Radius	Final Overload	Convergence Step	$\Delta\%$ Conv. vs $r = 1$	$p$	$d$
1	3.0	11.8	,	,	,
2	3.0	6.0	−49%	0.0093**	−0.509
3	3.0	3.9	−67%	0.0001***	0.838
5	3.0	2.3	−81%	<	−0.943
				0.0001***	
7 (full)	3.0	2.0	−83%	<	−0.965
				0.0001***	

Spearman correlation (radius vs convergence step):  $\rho = -0.494, p < 0.0001^{***}$ .

**Key finding.** The pattern  $O_{\min}$  is accessible even with radius = 1 (near-blind agents), but convergence speed scales with visibility. All view radii achieve the theoretical minimum overload of 3.0. This is a remarkable free lunch: even agents that see only one hole per step, agents with minimal information about the interface, channel the global mathematical truth  $O_{\min} = 3$ . The pattern manifests regardless of how much of the interface is visible; only the speed of manifestation varies.

Convergence speed varies by nearly 6×: full visibility ( $r = 7$ ) converges in 2.0 steps versus 11.8 steps for minimal visibility ( $r = 1$ ). This parallels the “power of two choices” result in load balancing (Azar et al., 1999): even a small increase in information access (from  $r = 1$  to  $r = 2$ ) produces a 49% speedup. The diminishing returns of additional visibility are striking: going from  $r = 1$  to  $r = 2$  saves 5.8 steps (49%), while going from  $r = 5$  to  $r = 7$  saves only 0.3 steps (13%). The relationship between radius and convergence step follows an approximate power law:  $\text{conv} \propto r^{-0.9}$ .

The fact that even  $r = 1$  (minimal information) reaches the global optimum is the strongest evidence of free-lunch pattern access in our experiments. It means that the mathematical truth  $O_{\min}$  is so “eager” to manifest that even a nearly blind interface, one that reveals only a single hole per step, suffices. The landscape has no local optima that could trap a nearly blind agent; the pattern provides a funnel that guides any local policy to the global minimum.

**Classification.** Universal convergence to  $O_{\min}$  across all radii constitutes a strong *free lunch*: the mathematical pattern manifests even through a near-blind interface. The convergence-speed gradient reveals the *information geometry* of pattern access, more visibility accelerates manifestation but is not required for it. The system never fails to channel the pattern; it only channels it more slowly with less information.

#### 4.5 Experiment 5, Chimeric Policies: Lateral Pattern Resonance

Mixed populations of  $m = 12$  pigeons and  $n = 8$  holes, with each pigeon randomly assigned one of two policies. Under the Platonic Space framework, this experiment tests lateral pattern resonance: whether mixed-policy populations access the same pattern and whether self-organization emerges as an additional free lunch.

Policy Pair	Final Overload	Aggregation Score
GREEDY + COOPERATIVE	4.0	0.738
GREEDY + EXPLORATORY	4.0	0.738
EXPLORATORY + COOPERATIVE	4.0	0.754
REPULSIVE + COOPERATIVE	4.0	0.796

No pairwise overload differences between any pair (all  $p = \text{n.s.}$ ).

**Key finding.** All chimeric combinations reach identical final overload, extending the pattern plurality result from homogeneous to heterogeneous populations: the mathematical truth  $O_{\min}$  manifests regardless of whether the interface is uniform or mixed. The pattern is invariant to the composition of access methods.

The aggregation scores (0.738–0.796) reveal an additional free lunch: self-organization by policy type. Pigeons following the same local rule tend to co-occupy holes at rates well above the chance level of 0.5. The REPULSIVE + COOPERATIVE pair shows the highest aggregation (0.796), suggesting that the spatial separation behavior of REPULSIVE pigeons creates niches that COOPERATIVE pigeons then fill. This spontaneous spatial sorting is not specified by any individual policy, it is a pattern that manifests through the interaction between different access methods. Under the Platonic Space framework, this is a lateral resonance phenomenon: different patterns (the local policies) interact and produce emergent spatial structure as an additional free lunch beyond the primary pattern  $O_{\min}$ .

The scale of this experiment ( $m = 12$ ,  $n = 8$ ) differs from the others to allow enough pigeons per hole for aggregation to be measurable. Despite the larger system, the pattern invariance persists: all pairs reach identical overload.

**Classification.** The aggregation pattern constitutes a *free lunch*: self-organization by policy type emerges without being specified by any policy, paralleling the chimeric sorting result of Zhang et al. (2024). The outcome invariance across chimeric compositions extends the *pattern plurality* result: the mathematical truth  $O_{\min}$  manifests through any composition of access methods.

#### 4.6 Experiment 6, Recovery After Damage: Bidirectional Pattern-Substrate Interface

A hole is frozen at step 167 and healed at step 333 (out of 500 total steps). Under the Platonic Space framework, this experiment tests the bidirectionality of the pattern-substrate interface: whether the pattern re-manifests after both damage and healing.

Condition	Final Overload	Failed Placements	Same-Target Retry	Repeat Failure
control	3.0	0.0	0.000	0.000
damage_only	3.0	49.4	0.257	0.257
damage_and_heal	3.0	24.7	0.226	0.211

All overload comparisons:  $p = \text{n.s.}$ ,  $d = 0.0$ .

**Key finding.** The pattern re-manifests after both damage and healing: final overload is 3.0 in all three conditions. The interface is bidirectional. When a hole is frozen, the substrate changes shift which optimum the pattern specifies (from  $O_{\min} = 3$  with 7 holes to  $O_{\min} = 4$  with 6 holes), and the system converges to the new optimum. When the hole heals, the pattern shifts back to  $O_{\min} = 3$ , and the system follows. The pattern pulls the system to the current optimum, whatever the substrate currently supports. Complete recovery is a free lunch: the system receives re-optimization without paying for any replanning computation.

The persistence metrics clarify that this recovery is pattern-channeling, not learning. Same-target retry remains non-zero under damage (0.257) and after healing (0.226), so agents do not form durable avoidance memories. Instead, recovery is a collective state-reconfiguration driven by the pattern: the distribution of pigeons adapts because the pattern  $O_{\min}$  shifts with the substrate, and the local policies channel the new pattern.

**Classification.** Complete recovery after both damage and healing constitutes a *free lunch*: the system receives re-optimization without computation. The bidirectionality of the interface, substrate changes shift the pattern, and the pattern pulls the system to the new optimum, is a direct instantiation of Levin’s (2026) claim that the physical-pattern interface is bidirectional: “patterns are modified by their projections into the physical world (it’s a 2-way interface).”

#### 4.7 Experiment 7, Progressive vs Sudden Damage: Memoryless Equals Pattern-Driven

Three holes are frozen either all at step 100 (sudden) or one every 100 steps (gradual: steps 100, 200, 300). Under the Platonic Space framework, this experiment provides the most direct evidence that the system’s convergence is pure pattern-channeling rather than learning.

Condition	Final Overload	Convergence Step	Failed Placements	Same-Target Retry	Repeat Failure
sudden	3.0	3.9	173.5	0.249	0.744
gradual	3.0	3.9	128.9	0.253	0.639

Overload comparison:  $p = 1.0$ ,  $d = 0.0$ . Convergence comparison:  $p = 1.0$ ,  $d = 0.0$ . Failed placements: gradual  $-26\%$  vs sudden,  $p < 0.0001$ ,  $d = -4.116$ . **Repeat failure: gradual**  $-14.1\%$  vs **sudden**,  $p < 0.0001$ ,  $d = -2.153$ .

**Key finding.** The absence of stress inoculation is the most theoretically significant result. There is no inoculation effect on outcome: both conditions reach identical overload and convergence speed. This contrasts with biological morphogenetic systems, where gradual exposure to stressors can build tolerance (Levin, 2022), and with transformer training, where temporal patterns in perturbation can affect learning trajectories (Kofman et al., 2025).

Under the Platonic Space framework, this absence is profound. Stress inoculation requires memory: prior exposure to a stressor triggers adaptive responses that improve tolerance to future exposure. Our pigeons have no memory. Each activation is stateless: the pigeon inspects current loads, applies its policy, and acts. Without memory, there is no substrate for inoculation. This confirms that the pigeonhole convergence is pure pattern-channeling, not learning. The system does not benefit from experience because it does not accumulate experience. The free lunch comes entirely from the pattern, the mathematical truth  $O_{\min}$ , not from any history-dependent adaptation. Unlike transformer training, where temporal pattern access may play a role, the memoryless pigeonhole system strips away all learning-like mechanisms, leaving only the pattern.

However, gradual damage does reduce wasted effort: 26% fewer failed placements ( $d = -4.116$ ) and 14.1% lower repeat failure ( $d = -2.153$ ). Same-target retry is unchanged (0.249 vs 0.253, n.s.), which is informative: gradual damage smooths the collective path through state space without making individual pigeons more avoidant of recently failed targets. This process-level difference arises not from learning but from the temporal distribution of the perturbation, the pattern  $O_{\min}$  shifts gradually rather than abruptly, producing a smoother trajectory to the same endpoint.

**Classification.** The outcome equivalence is a *pattern invariant*: the final state depends only on the total damage, not its temporal profile, because the pattern  $O_{\min}$  depends only on the number of usable holes, not on the damage history. The absence of stress inoculation confirms *pure pattern-channeling*: the free lunch comes from the pattern, not from accumulated experience.

#### 4.8 Experiment 8, Misleading Holes: Pattern Corruption

Misleading holes report their load as 0 regardless of true occupancy, actively corrupting the interface between the system and the pattern. Under the Platonic Space framework, this experiment tests pattern corruption: what happens when the interface does not merely degrade (as in frozen holes) but actively inverts the pattern-seeking mechanism.

Misleading Holes	Final Overload	Occupancy Bias	Overload Bias	Load Gap
0	3.0	0.000	0.000	0.000
1	3.77	0.217	0.518	2.533
2	4.20	0.274	0.559	1.920
3	4.33	0.265	0.511	1.544
4	4.40	0.219	0.418	1.275
5	4.37	0.156	0.286	1.090
6	4.30	0.080	0.143	0.928

Spearman correlation (misleading holes vs overload):  $\rho = +0.417, p < 0.0001^{***}$ .

**Key finding.** Misleading holes produce pattern corruption, which is categorically different from pattern unavailability (frozen holes). Frozen holes remove capacity: the pattern-seeking mechanism routes around the absent substrate. Misleading holes invert the pattern-seeking mechanism: the same machinery that efficiently finds  $O_{\min}$  now efficiently finds the wrong target. A single misleading hole occupies 14% of the substrate but captures 36% of pigeons and 66% of overload, yielding occupancy bias = 0.217 and overload bias = 0.518. With two misleading holes, deceptive substrate captures 56% of pigeons and 84.5% of overload.

Under the Platonic Space framework, this is a free lunch reversal. The same pattern-channeling mechanism that produces the free lunch (efficient convergence to  $O_{\min}$ ) now produces a “free disaster”: efficient convergence toward the corrupted attractor. The system’s efficiency at channeling patterns becomes a liability when the interface corrupts the pattern. Deception turns the free lunch into its inverse, the system pays a cost it did not incur.

A plateau-and-reversal effect is visible: degradation is steepest for the first 1–2 misleading holes (25.6% and 40.0%), peaks at 4 misleading holes (46.7%), and then decreases slightly at 5–6 misleading holes (45.6% and 43.3%). This suggests that once a critical fraction of holes is deceptive, additional deception has diminishing marginal effect and may even slightly reduce overload, likely because when nearly all holes are misleading, the deception becomes uniform and ceases to create differential attraction.

The comparison between Experiments 1 and 8 maps onto the classical distributed systems distinction between crash faults and Byzantine faults (Lamport et al., 1982). A crashed node is simply absent; a Byzantine node sends arbitrary (potentially malicious) messages. Under the Platonic Space framework, crash faults correspond to pattern unavailability (the interface loses capacity but does not mislead), while Byzantine faults correspond to pattern corruption (the interface actively inverts the pattern-seeking mechanism). The result that Byzantine faults are categorically harder to tolerate, both in classical distributed computing and in our experiments, reflects a deep asymmetry: a degraded interface merely reduces the bandwidth for pattern transmission, but a corrupted interface turns the transmission mechanism against itself.

**Classification.** Misleading holes constitute *pattern corruption*: the interface actively misleads, inverting the pattern-seeking mechanism. The system’s pattern-channeling efficiency becomes a vulnerability under corruption, producing a *free lunch reversal*, the same mechanism that provides the free lunch now provides a free disaster.

## 5. Discussion

### 5.1 The Pigeonhole Principle as a Platonic Space Pattern

The pigeonhole principle provides what may be the most direct test of Levin’s (2026) Platonic Space framework. The framework claims that physical systems serve as interfaces for non-physical patterns, and that these patterns are causal, they determine the behavior of the physical system. In

most applications, the “pattern” is implicit, complex, or debatable: in biology, the morphogenetic target; in transformer training, the loss landscape structure; in sorting, the correct permutation. In each case, one might argue that the pattern is merely an emergent property of the physical dynamics rather than a pre-existing mathematical truth.

The pigeonhole principle eliminates this ambiguity. The pattern is  $O_{\min} = m - n$ : a theorem that every undergraduate knows, derivable from first principles, invariant across implementations, and causally determinative. Under the Platonic Space framework, this theorem is a pattern in the latent space, a mathematical truth that pre-exists any particular pigeonhole system. The physical system (10 pigeons, 7 holes, local policies) is the interface through which this pattern manifests.

The evidence for pattern-channeling is strong. No agent computes  $O_{\min}$ . No agent has a representation of the global overload. No agent communicates with any other agent. Yet the system converges to  $O_{\min} = 3$  in six of eight experimental conditions, using four different local policies, at five different view radii (including near-blind  $r = 1$ ), in homogeneous and heterogeneous populations, and after dynamic damage and recovery. The mathematical truth determines the system’s behavior without any agent computing it. This is pattern manifestation in Levin’s precise sense.

## 5.2 What Faultization Reveals

Faultization, systematic perturbation of the interface, reveals four categories of behavior that map directly onto the Platonic Space framework:

**Pattern manifestation.** The system converges to  $O_{\min}$  in Experiments 1 (frozen\_0 through frozen\_3), 2 (all four policies), 4 (all view radii), 5 (all chimeric pairs), 6 (all recovery conditions), and 7 (both damage schedules). In all these cases, the mathematical pattern manifests through the interface despite perturbation. The pattern is robust to substrate damage (up to 43%), policy choice, information scope, population heterogeneity, dynamic damage, and temporal damage profile.

**Pattern fidelity.** Experiment 3 (noisy perception) reveals that the discrete pigeonhole interface requires high-fidelity perception, there is no noise buffer. Experiment 4 (view radius) reveals that the pattern manifests even with minimal information, but convergence speed depends on visibility. Together, these experiments characterize the interface’s fidelity requirements: accuracy matters, but completeness does not.

**Pattern corruption.** Experiment 8 (misleading holes) reveals that corruption is categorically different from damage. Frozen holes reduce interface capacity; misleading holes invert the pattern-seeking mechanism. The same machinery that channels  $O_{\min}$  now channels the wrong target.

**Free lunch.** Across all experiments, the system receives more than it pays for. The payment is local greedy computation with no global information. The receipt is global optimality, damage compensation, recovery, and self-organization. The gap between payment and receipt is the free lunch.

### 5.3 Free Lunch Quantification

Under the Platonic Space framework, the central question is: what does the system receive without paying for? The following table quantifies the free lunch for each experiment.

Experiment	What was specified	What was received (free lunch)
1: Frozen Robustness	Local greedy policies, damaged substrate	Optimal overload despite 43% damage
2: Policy Comparison	Four different local rules	Identical global optimum from all four
3: Noisy Perception	Noisy load perception	Near-optimal overload at low noise
4: View Radius	Minimal visibility ( $r = 1$ )	Global optimum from single-hole observations
5: Chimeric	Mixed-policy populations	Optimal overload + self-organized spatial clustering
6: Recovery	Local policies, dynamic substrate	Complete recovery after damage and healing
7: Progressive vs Sudden	Memoryless agents, temporal perturbation	Identical outcome regardless of damage schedule
8: Misleading	Pattern-seeking local policies	Free lunch reversal: efficient convergence to wrong target

Experiment 8 is the exception that proves the rule. The same pattern-channeling mechanism that produces free lunches in Experiments 1–7 produces a “free disaster” in Experiment 8. When the interface corrupts the pattern, the system’s efficiency at channeling becomes a liability. This is not a failure of the pattern but a failure of the interface: the mathematical truth  $O_{\min}$  still exists, but the corrupted interface channels a different, false signal.

### 5.4 Pattern Access in Discrete vs Continuous Domains

The comparison between the pigeonhole system and transformer training (Kofman et al., 2025) reveals that different interface types have different fidelity requirements for channeling mathematical patterns. In transformer training, gradient noise up to  $\sigma = 0.01$  is tolerated before degradation begins, the continuous loss landscape provides an averaging mechanism that absorbs small perturbations. In the pigeonhole system, there is no noise buffer: even  $\sigma = 0.5$  causes statistically significant degradation.

Under the Platonic Space framework, this reflects the interface type, not the pattern. The pattern (a mathematical optimum) is the same kind of object in both systems. But the interface differs: continuous gradient descent has smoothing mechanisms (averaging across parameters, batch gradients) that discrete pigeonhole placement lacks. A pigeon choosing among 7 holes has no averaging

mechanism, any noise-induced redirection to the wrong hole produces immediate overload. The interface's fidelity requirement depends on the interface, not the pattern.

The companion faultization study on a minimal GPT transformer sharpens this contrast. The GPT system shows approximate convergence (final loss varies across runs), a noise tolerance threshold at  $\sigma = 0.01$ , and, most strikingly, stress inoculation: gradual noise exposure builds tolerance that sudden exposure does not ( $p = 0.0001$  at  $n = 300$ ), because the optimizer's momentum state accumulates history. The pigeonhole system shows exact convergence (zero variance), no noise tolerance, and no stress inoculation, because the memoryless agents have no state to accumulate. The discrete mathematical pattern manifests with higher precision but requires higher fidelity; the continuous optimization pattern manifests with lower precision but tolerates more interface degradation and supports history-dependent pattern access.

This suggests a practical principle for the Platonic Space research program: when probing pattern access, the interface type constrains what perturbations are informative. Noise experiments are most revealing in discrete systems (where there is no buffer), while stress inoculation experiments are most revealing in continuous systems (where history-dependent access is possible). Substrate damage and information restriction are informative across both.

### 5.5 Pattern Corruption: Crash vs Byzantine Faults

The sharp distinction between frozen holes (Experiment 1) and misleading holes (Experiment 8) maps onto a fundamental asymmetry in the Platonic Space framework. Frozen holes correspond to pattern unavailability: the interface loses capacity, and the pattern adapts,  $O_{\min}$  increases with the number of frozen holes, and the system channels the new, higher  $O_{\min}$ . The pattern is still accessible; it is simply a different (less favorable) pattern for the degraded substrate.

Misleading holes correspond to pattern corruption: the interface does not merely lose capacity but actively sends false signals. The system's pattern-seeking mechanism, the same local policies that efficiently channel  $O_{\min}$  under honest conditions, is activated toward false attractors. The load of 0 reported by a misleading hole is a corrupted signal from the interface, and the system's efficient response to signals becomes its vulnerability.

Under the Platonic Space framework, this distinction is predicted. Levin (2026) argues that physical systems are interfaces for patterns, and that the interface can degrade in different ways. A degraded interface (crash fault) reduces the bandwidth for pattern transmission but does not corrupt it. A deceptive interface (Byzantine fault) turns the transmission mechanism against itself. The result that Byzantine faults are categorically harder to tolerate is not merely an engineering observation, it reflects a fundamental asymmetry in how interfaces can fail to transmit patterns.

### 5.6 Connection to Levin's Platonic Space

Our results map directly onto several claims in Levin's (2026) framework:

**"Patterns are causal."** The mathematical truth  $O_{\min} = m - n$  determines system behavior. No

agent computes it, yet the system converges to it in 6/8 experiments. The pattern is causal in Pearl’s counterfactual sense: if  $O_{\min}$  were different (different  $m$ , different  $n$ ), the system’s final state would be different. The pattern, not the policy, determines the outcome.

**“Physical systems are interfaces.”** The pigeons and holes constitute an interface through which the pattern  $O_{\min}$  manifests. The interface can be degraded (frozen holes), corrupted (misleading holes), or restricted (limited view radius), and the pattern’s manifestation changes accordingly. But the pattern itself is invariant, it is the interface that varies.

**“Free lunches.”** The system receives optimal global coordination from local greedy policies. The computational cost of finding  $O_{\min}$  by exhaustive search of  $\sim 10^9$  states would be enormous. The system pays only the cost of local greedy steps and receives global optimality for free. Under the Platonic Space framework, this is because “no free lunch” commitments are derived from the laws of the physical world, and the Platonic Space offers patterns for which the physical processes of learning, evolution, and engineering do not need to pay.

**“We need to build interfaces and study what unexpected patterns ingress through them.”** Our hook-based faultization architecture is precisely such a methodology. By systematically perturbing the interface, we observe what patterns survive degradation, what fidelity the interface requires, and what happens when the interface corrupts. The pigeonhole system is the simplest interface we can build for a known mathematical pattern, making it the ideal starting point for the research program Levin describes.

## 5.7 Limitations

Several limitations of the current study should be noted.

**Scale.** Our system is small ( $m = 10, n = 7$ ). Larger systems might exhibit qualitatively different behaviors, such as phase transitions in convergence or emergent delayed gratification driven by longer coordination chains. In particular, the “funnel-shaped” landscape property (no local optima) might break down at larger scales, where the combinatorial state space grows exponentially and the fraction of optimal states shrinks.

**Policy simplicity.** All four policies are stateless and memoryless. Policies with memory (e.g., reinforcement learning agents, or pigeons that remember which holes rejected them) might exhibit delayed gratification or more sophisticated adaptation. The universal absence of DG in our results may be an artifact of the memoryless architecture rather than a property of the pigeonhole domain.

**Spatial structure.** Our system has no spatial layout; any pigeon can attempt any hole (subject to view radius constraints). A 2D spatial version with local connectivity might exhibit richer phenomena, including traveling waves of redistribution, spatial phase separation, and boundary effects at the edges of pigeon territories.

**Potential function specificity.** The results may depend on the particular values of  $\alpha = 1.0$ ,  $\beta = 0.5$ ,  $\gamma = 10.0$ . We did not sweep these hyperparameters, which could reveal sensitivity

boundaries. In particular, the relative weight of the concentration penalty ( $\beta$ ) versus the overload penalty ( $\alpha$ ) likely affects whether the system prefers to distribute excess evenly or consolidate it.

**Fixed  $m/n$  ratio.** We tested only  $m/n = 10/7 \approx 1.43$ . More extreme ratios (higher overload density, e.g.,  $m/n = 3$ ) or near-critical ratios ( $m = n + 1$ , minimal impossibility) might produce qualitatively different behaviors. Near the critical ratio, the system might exhibit fluctuations between feasibility and infeasibility that could enable DG-like transients.

**Sample size.** While  $n = 30$  replications provide adequate power for the large effects observed (most effect sizes  $|d| > 0.8$ ), smaller effects (e.g., the Spearman correlation in Experiment 1,  $\rho = -0.15$ ) are at the edge of detectability and should be interpreted cautiously.

**Platonic Space interpretation.** The Platonic Space framework provides a productive interpretation of our results, but the results do not require this interpretation. The free-lunch phenomena we observe (global optimality from local rules, damage compensation, recovery) can also be explained by the funnel-shaped structure of the potential landscape without invoking a non-physical pattern space. Under the Platonic Space framework, the funnel shape itself is a pattern that the system channels; under a purely physicalist interpretation, the funnel shape is an emergent property of the dynamics. Our experiments do not distinguish between these interpretations. What the experiments do establish is that the system exhibits behaviors consistent with pattern-channeling, and that the Platonic Space framework provides a productive vocabulary for classifying and predicting those behaviors.

## 5.8 Future Work

Several directions emerge from the Platonic Space interpretation of our results.

**Mapping the pattern space.** Varying  $m$ ,  $n$ , and the  $m/n$  ratio would map how the pattern  $O_{\min} = m - n$  scales and whether the free lunch persists at larger scales. At what system size does the funnel-shaped landscape break down? Does the free lunch grow or shrink with scale?

**Varying the constraint type.** The pigeonhole principle is the simplest impossibility constraint ( $m > n$ ). Applying faultization to other impossibility theorems, graph coloring with too few colors, bin packing with too few bins, the party problem (Ramsey theory), would test whether free-lunch pattern access generalizes across different mathematical patterns.

**Adding memory.** Policies with memory would test whether the system transitions from pure pattern-channeling to learning-augmented channeling. If memoryful agents exhibit stress inoculation and delayed gratification, this would confirm that our memoryless results isolate the pattern contribution from the learning contribution.

**Richer corruption models.** Our misleading holes report load as 0 uniformly. More sophisticated corruption models (noisy reporting, strategic deception, time-varying corruption) would map the boundary between pattern fidelity degradation and pattern corruption.

**Cross-interface comparison.** Systematically comparing the pigeonhole system, sorting arrays, and

transformer training under identical perturbation types would characterize how different interface types transmit the same kinds of patterns. This comparative approach is central to the research program Levin (2026) describes.

## 6. Conclusion

The pigeonhole principle provides the cleanest test of the Platonic Space framework: the pattern is a known mathematical theorem ( $O_{\min} = m - n$ ), the interface is a distributed multi-agent system of memoryless agents with local policies, and the system channels the theorem without computing it. Faultization reveals what happens when we degrade this interface.

Our eight experiments demonstrate that the system receives a free lunch: optimal global coordination without global optimization, in six of eight conditions. The mathematical truth  $O_{\min}$  manifests through four different access methods, at five different visibility levels (including near-blind), in homogeneous and heterogeneous populations, and after dynamic damage and recovery. No agent computes the global optimum; the pattern provides it for free.

The most striking finding is the asymmetry between pattern unavailability and pattern corruption. Frozen holes (crash faults) reduce the interface's capacity, but the pattern adapts:  $O_{\min}$  increases, and the system channels the new pattern. Misleading holes (Byzantine faults) corrupt the interface, and the same pattern-seeking mechanism that produces the free lunch now produces a free disaster, efficient convergence toward false attractors. The system's efficiency at channeling patterns is precisely its vulnerability under corruption.

The absence of stress inoculation (Experiment 7) and the absence of delayed gratification (DG Index = 0.0 across all conditions) confirm that the convergence is pure pattern-channeling. The memoryless system has no history to exploit. The free lunch comes entirely from the pattern, from the mathematical truth  $O_{\min} = m - n$ , not from accumulated experience. This makes the pigeonhole system the ideal minimal model for studying free-lunch phenomena in the Platonic Space framework: the pattern contribution is isolated because there is no learning contribution to confound it.

Three directions for future work emerge naturally. First, scaling to larger systems ( $m, n \gg 10$ ) would test whether the free lunch persists or diminishes at scale. Second, adding memory to the pigeon policies would test whether the system transitions from pure pattern-channeling to learning-augmented channeling, potentially enabling stress inoculation and delayed gratification. Third, applying faultization to other impossibility theorems (graph coloring, bin packing, Ramsey theory) would map whether free-lunch pattern access generalizes across different mathematical patterns in the Platonic Space.

The central message of this work is that the Platonic Space framework, applied to the simplest impossibility theorem, makes testable predictions that our experiments confirm: mathematical patterns manifest through physical interfaces, the manifestation survives interface degradation but not interface corruption, and the system receives more than it pays for. The pigeonhole principle,

because the pattern is a known theorem, provides the most quantitatively precise test of these predictions available today.

## References

- Aguilera, M. K., Chen, W., and Toueg, S. (2004). Failure detection and consensus in the crash-recovery model. *Distributed Computing*, 13(2), 99–125.
- Azar, Y., Broder, A. Z., Karlin, A. R., and Upfal, E. (1999). Balanced allocations. *SIAM Journal on Computing*, 29(1), 180–200.
- Kofman, D., Bhatt, U., and Levin, M. (2025). Morphogenetic perturbation of training: probing transformer competencies beyond the nominal loss landscape. Preprint.
- Lamport, L., Shostak, R., and Pease, M. (1982). The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, 4(3), 382–401.
- Levin, M. (2019). The computational boundary of a “self”: developmental bioelectricity drives multicellularity and scale-free cognition. *Frontiers in Psychology*, 10, 2688.
- Levin, M. (2022). Technological approach to mind everywhere: an experimentally-grounded framework for understanding diverse bodies and minds. *Frontiers in Systems Neuroscience*, 16, 768201.
- Levin, M. (2026). A short argument on Platonic Space. Blog post, March 31, 2026.
- Rosenthal, R. W. (1973). A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2(1), 65–67.
- Roughgarden, T., and Tardos, E. (2002). How bad is selfish routing? *Journal of the ACM*, 49(2), 236–259.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129–138.
- Zhang, A., Goldstein, I., and Levin, M. (2024). Classical sorting algorithms as a model of morphogenesis: self-sorting arrays reveal unexpected competencies in a minimal model of basal intelligence. *arXiv preprint arXiv:2401.05375*.