Astrocyte-Mediated Learning via Local Perturbation with Global Scalar Reward

An Experimental Exploration in Biologically Inspired Computing

Dusty Wilhelm Murray & Claude (Anthropic)
February 2026

Transparency note. This document is not a peer-reviewed academic paper. It is an experimental report produced by a software engineer (25 years in EDI/B2B integration systems, zero neuroscience credentials) in collaboration with Claude, an AI assistant by Anthropic. The experimental methodology is inspired by the scientific method: controlled experiments, ablation studies, null results reported honestly, and conclusions derived from data rather than assumption. All code was developed via AI-assisted rapid prototyping ('vibe coding'). The claims made here are limited and testable: we demonstrate that a specific computational mechanism works in a specific small-scale setting. We make no claims about biological brains, nor do we claim to have discovered how learning works. We share this in the spirit of open exploration.

Abstract

We present an experimental exploration of learning in small recurrent neural networks using astrocyte-like entities as the sole mechanism of plasticity. Unlike conventional approaches where synapses contain their own learning rules (Hebbian, STDP, backpropagation), our system assigns all learning to a population of spatially fixed astrocytes that (1) select which synapses to perturb based on local synaptic traffic, (2) apply random perturbations, (3) evaluate outcomes via a single global scalar reward, and (4) keep or revert changes accordingly. Astrocytes adapt their own activation thresholds based on reward history, and their exploration rates decay with maturity — providing self-organising plasticity control without external scheduling. Over 24 experiments, we show that this mechanism produces genuine pattern discrimination (up to 100% accuracy on individual patterns, mean 67.7% across an 8-pattern task vs. 50% chance), while 18 prior experiments using neuron-centric learning mechanisms (eligibility traces, chemical diffusion, flag gates) produced only constant-output attractors. The work was conducted over a single weekend using AI-assisted rapid prototyping, suggesting that exploratory computational neuroscience can be dramatically accelerated by human-AI collaboration.

Keywords: biologically plausible learning, astrocytes, local perturbation, global reward, credit assignment, neuromodulation, AI-assisted research

1. Introduction

The credit assignment problem — determining which synapses in a network contributed to a given outcome — is the central challenge of neural learning. Backpropagation solves it exactly by computing per-synapse gradients through the chain rule, but requires biologically implausible mechanisms: symmetric feedback weights, continuous derivatives, and a separate backward pass through the network. Decades of work on biologically plausible alternatives (Hebbian learning, spike-timing-dependent plasticity, reward-modulated plasticity) have sought to achieve credit assignment using only information locally available at each synapse.

We began this project with a standard neuron-centric approach: eligibility traces marking co-active synapses, chemical signals diffusing from modulatory neurons as reward broadcasts, and flag gates filtering which synapses are eligible for update. Over 18 experiments (Phase 1), none of these mechanisms produced genuine pattern discrimination. The best results — 55% accuracy on a 50% chance task — turned out to be constant-output attractors: the network outputting the same vector for every input, scoring well by coincidence of bit overlap rather than by learning.

The critical insight came from analysing why these mechanisms failed. With global chemical reward and simultaneous update of all ~2,300 synapses, every perturbation was evaluated against the aggregate effect of 2,300 simultaneous changes. Credit assignment was not merely difficult — it was mathematically intractable. A single scalar reward signal cannot determine which of 2,300 simultaneous changes was responsible for the outcome.

This led us to reframe the problem. The bottleneck was never the reward signal. It was the update scope. Global reward works when perturbation is local. Changing 20 synapses and evaluating with one scalar is tractable — changing 2,300 synapses and evaluating with one scalar is not. This is the same principle that makes genetic algorithms work: one fitness score is sufficient if you only mutate a few genes per generation.

The question then became: what biological mechanism restricts plasticity to a local spatial neighbourhood? The answer, supported by recent experimental evidence, is astrocytes.

2. Background and Motivation

2.1 Astrocytes in biological learning

Astrocytes are glial cells that tile the cortex in non-overlapping spatial domains. Each astrocyte wraps around approximately 100,000 synapses in its territory and communicates via slow calcium waves operating on the timescale of seconds to tens of seconds — far slower than neural firing at milliseconds. They sense synaptic activity through glutamate spillover and modulate synaptic strength through release of gliotransmitters including D-serine, glutamate, and ATP.

Recent experimental work has established that astrocytes are essential, not merely supportive, for learning. Hösli et al. (2022) demonstrated that disrupting astrocyte gap junction coupling in adult mice eliminates spatial learning and memory entirely — not merely impairing it, but abolishing it completely. Bohmbach et al. (2022) showed that astrocyte-mediated D-serine release implements a BCM-like plasticity rule that controls whether synapses potentiate or depress. De Pitta and Bhatt (2024) modelled astrocytes as contextual guidance agents enabling multi-context learning through meta-plasticity on slower timescales than neuronal dynamics.

However, all existing computational models use astrocytes to *modulate* an underlying synaptic learning rule. The synapses still contain their own plasticity mechanism (Hebbian, STDP, reinforcement learning delta rules). Astrocytes adjust gain, threshold, or timing. To our knowledge, no computational model has tested whether astrocytes alone — without any synaptic learning rule — are sufficient for learning.

2.2 The genetic algorithm analogy

A genetic algorithm optimises complex solutions using a single fitness score. No per-gene gradient. No backpropagation. One scalar: 'this organism scored 73.' The key constraint that makes this work is that only a few genes mutate per generation. If every gene mutated simultaneously, the fitness change would be unattributable and the algorithm would perform a random walk.

Our astrocyte mechanism is a genetic algorithm running inside a neural network. The population is the set of synaptic weight configurations. Mutation is the astrocyte's random perturbation of local synapses. Selection is the global scalar reward. The spatial locality of astrocyte territories provides the same constraint as low mutation rates in genetic algorithms: only a small subset of the parameter space changes per evaluation, making credit assignment tractable.

2.3 Design principles

Two commitments guided the experimental design. First, **simple global signals**: the reward is one number (did overall accuracy improve or not), with no per-output error and no per-synapse gradient. This mirrors the diffuse neuromodulatory broadcasts (dopamine, norepinephrine) available to biological organisms. Second, **local architecture resolves credit assignment**: the only structure required to make global reward useful is spatial locality of plasticity, provided by astrocyte territories. No additional mechanism is needed.

3. Architecture

3.1 Network topology

The network consists of 2 clusters of 30 neurons each (60 total). Each cluster contains 5 input neurons, 5 output neurons, 2 modulatory neurons (unused in the astrocyte architecture but retained for comparability with Phase 1), and 18 regular interneurons. Intra-cluster connectivity is 60% and inter-cluster connectivity is 50%, both random. This produces approximately 2,300 synapses. Neurons are positioned in 2D space with cluster spacing of 15 units.

Neurons are binary threshold units: they accumulate weighted inputs from all incoming synapses, compare the sum to a firing threshold (fixed at 0.5), and output 1 (fire) or 0 (silent). There is no continuous activation function, no learnable bias, and no temporal dynamics beyond the single propagation step per evaluation. Neurons have no learning rule.

3.2 Astrocyte population

Eight astrocytes tile the network, four per cluster, positioned to provide overlapping coverage. Each astrocyte has a fixed spatial position and a territory defined by a radius of 3.0 units in 2D space. A synapse belongs to an astrocyte's territory if its pre-synaptic or post-synaptic neuron is within the radius. Territories overlap, meaning some synapses are influenced by multiple astrocytes. Territory sizes range from 200 to 900 synapses depending on position and local neuron density.

Each astrocyte maintains internal state: an activation threshold (initially 0.5, adaptive), a rolling window of the last 20 reward outcomes, a lifetime activation count, and a success count (perturbations that were kept). These quantities evolve over training and determine the astrocyte's behaviour.

3.3 Task

The task is 5-bit pattern mapping: given a 5-bit binary input, produce a 5-bit binary output. Eight input-output pairs (P1–P8) are used, presenting a non-trivial combinatorial challenge. Patterns include inverse relationships and overlapping bit structures that require genuine discrimination, not simply a fixed output bias. Random chance performance is 50% (each output bit has 50% probability of matching by chance).

4. The Astrocyte Learning Mechanism

Learning proceeds through a cycle of sense, activate, perturb, evaluate, and adapt. No synaptic learning rule operates at any point. All plasticity is controlled by the astrocyte population.

4.1 Sensing: synaptic traffic

Each astrocyte computes an activation score based on the total synaptic traffic flowing through its territory. For each synapse, the traffic is the absolute value of the pre-synaptic neuron's output multiplied by the synapse weight. The activation score is the mean traffic across all synapses in the territory. This mirrors biological glutamate spillover detection: the astrocyte senses how much signal is passing through its local synapses regardless of whether post-synaptic neurons fire. Crucially, this allows astrocytes in output regions (where neurons may not yet fire) to detect incoming signals and activate.

4.2 Activation: threshold and exploration

An astrocyte activates if its activation score exceeds its adaptive threshold. A minimum of 1 and maximum of 3 astrocytes are active per step. Additionally, each astrocyte has a maturity-scaled exploration rate that provides spontaneous activation independent of the threshold:

$$explorationRate = 0.30 / (1 + activationCount / 2000)$$

A fresh astrocyte with zero activations explores at 30%. After 2,000 activations the rate drops to 15%. After 40,000 activations it falls to 1.5%. This implements developmental annealing: high early exploration (analogous to infant synaptic turnover) followed by progressive refinement (analogous to cortical maturation). Maturation is per-astrocyte, not global — different territories mature at different rates depending on their activity history, mirroring the observation that visual cortex matures before prefrontal cortex.

4.3 Perturbation: random and local

Active astrocytes perturb all synapses in their territory. Each synapse receives a random Gaussian delta (mean 0, standard deviation 0.1) added to its current weight, clamped to [-2.0, +2.0]. The current weights under each active astrocyte's territory are saved before perturbation (snapshot).

4.4 Evaluation: global scalar reward

After perturbation, the network performs a forward pass with the current clamped input and a soft reward is computed: the sigmoid distance of each output neuron's activation from its target, averaged across all output bits. This produces a continuous scalar that distinguishes between 'almost fired correctly' and 'far from correct,' unlike binary accuracy which cannot detect sub-threshold improvements.

4.5 Keep or revert

If the soft reward after perturbation strictly exceeds the soft reward before, all weight changes are kept. Otherwise, all weights under active astrocyte territories are reverted to their snapshot values. This is an all-or-nothing decision per step. Biologically, this maps to synaptic tagging and capture: temporary modifications become permanent only if performance improves within a consolidation window.

4.6 Astrocyte self-adaptation

Every 50 steps, each astrocyte adjusts its activation threshold based on its recent success rate (the fraction of its last 20 activations that resulted in kept perturbations). If the success rate exceeds 15%, the threshold decreases by 0.01 (the astrocyte becomes more responsive). If below 5%, the threshold increases by 0.01 (the astrocyte becomes more conservative). Thresholds are clamped to [0.1, 0.9]. This produces self-organising plasticity control: productive astrocytes intensify their participation while unproductive ones reduce theirs.

5. Experimental History

The project comprised two phases across 24 experiments, conducted over a single weekend using AI-assisted rapid prototyping (human-Claude collaboration). Each experiment was designed based on the results of the previous one, forming a chain of reasoning from initial failure to final mechanism.

5.1 Phase 1: neuron-centric mechanisms (Experiments 1–18)

Phase 1 tested conventional biologically plausible learning mechanisms: eligibility traces (four-quadrant Hebbian co-firing detection), chemical diffusion from modulatory neurons (spatial reward broadcast with 1/distance falloff), dampening filters (activity-history gating), ambient relevance fields, flag gates (persistence mechanisms for eligibility traces), reward shaping (squared, annealed), architectural changes (propagation cycles, hidden clusters), and various parameter sweeps.

Experiment	Mechanism	Result	Insight
001-003	Eligibility traces + chemical diffusion	~50% (chance)	No learning signal visible
004	Flag persistence tuning	45% -> 53%	Appeared to help; was constant output
005	Squared reward	Deadlock	Starved bootstrapping signal
006	Reward annealing	~53%	Three shapes, same result
007	Connectivity diagnostic	Null result	Ruled out topology as bottleneck
008	Flag diagnostic	90%+ saturation	Flags latched uniformly, not gating

Experiment	Mechanism	Result	Insight
009	Flag warmup delay	Worse	Unfiltered early learning locks bad attractors
010	Direction-consistent flags	~55%	Reduced saturation, unchanged accuracy
011	Frustration-based flip	Bimodal	Revealed weight starvation
012	Weight decay halved	55% (10/10 seeds)	One parameter beat 11 experiments
013	Propagation cycles (fixed)	52%	Multi-hop adds noise, not structure
014	Hidden cluster	55%	Capacity without credit assignment is useless
015	Spatial cursor (3 combined)	47.5%	Weight starvation from stacked mechanisms
016-018	Per-bit reward + fixes	56-75% transient	Genuine but unstable discrimination

Table 1. *Phase 1 experimental summary. All experiments used 10 seeds. The 55% ceiling in experiments 10–14 was later revealed to be a constant-output attractor: the network producing [11111] for all inputs.*

The pivotal diagnostic came when output vectors were logged for the first time. The network producing 55% accuracy was outputting the identical 5-bit vector for every input pattern. The arithmetic is exact: [11111] scores 40%, 60%, 60%, 60% against the four test targets, averaging to 55%. Eighteen experiments had been measuring attractor stability, not learning. Mechanisms that 'improved accuracy' had merely helped the network find and hold the optimal constant output faster.

5.2 Phase 1b: the cursor discovery (Experiments 020–024)

The transition to astrocyte-mediated learning began with a stripped-down control experiment: a single spatial cursor that hopped randomly across the network, perturbing a small cluster of synapses per step and evaluating with one global scalar.

Experiment	Mechanism	Accuracy	Key Finding
020 iter4	Single cursor, no homeostasis	64.8%	First genuine discrimination: 3-7 distinct outputs per seed
021	8 astrocytes, activity-sensing	66.7%	Astrocyte > cursor; 38% accept rate vs 3.7%
022	Synaptic traffic sensing	66.5%	Traffic signal too weak to wake output cluster
023	Epsilon exploration (flat 1%)	68.0% / 66.2%	Cluster 1 wakes up; degrades over 20k episodes
024	Maturity-scaled exploration	67.7%	Protects mature weights; 3/4 criteria passed

Table 2. *Phase 1b experimental progression. All experiments used 10 seeds. Accuracy is mean across seeds during inference (100 frozen-weight episodes).*

6. Results

6.1 The cursor establishes the baseline

Experiment 020 (iteration 4) was the first test of pure local perturbation with global scalar reward. With homeostasis disabled, no eligibility traces, no chemical diffusion, no dampening, and no flags, a single spatial cursor produced:

Metric	Cursor	Control (no training)
Mean accuracy	64.8%	50.0%
Seeds with 2+ distinct outputs	10/10	0/10
Distinct output vectors per seed	3-7	1 (constant)
Accept rate	3.7%	N/A

Table 3. *Cursor vs. control. The control condition (random initialisation, no training) produced constant output on every seed, confirming 50% is the true floor.*

This result was decisive. Ten out of ten seeds produced genuinely different output vectors for different input patterns. The mechanism that 18 experiments of neuron-centric learning could not achieve was accomplished by random perturbation restricted to a small spatial neighbourhood.

6.2 Astrocytes improve on the cursor

Replacing the single random cursor with eight fixed-territory astrocytes (Experiment 021) produced a consistent improvement. The critical advance was not accuracy alone but efficiency: astrocytes found productive perturbations ten times more frequently.

Metric	Astrocyte	Cursor	Control
Mean accuracy	66.7%	64.7%	50.0%
Accept rate	38.0%	3.7%	N/A
Seeds astrocyte > cursor	7/10	-	-
Pattern specialisation (>2x bias)	8/10 seeds	N/A	N/A

Table 4. *Astrocyte vs. cursor vs. control (Experiment 021, 5,000 episodes, 10 seeds).*

6.3 Maturity-scaled exploration

The full system with maturity-scaled exploration (Experiment 024, 20,000 episodes) achieved 67.7% mean accuracy with demonstrated self-organising plasticity control. Astrocyte exploration rates decayed from 30% to under 1% for mature territories, while immature territories (output cluster) started high and caught up. Best individual seed (314) achieved 77.5% mean accuracy with three perfectly learned patterns (100%).

6.4 Per-pattern analysis

Detailed output vector logging reveals genuine pattern discrimination. Seed 314 (Experiment 024, maturity-scaled) illustrates the system's capabilities and limitations:

Pattern	Input	Target	Output	Accuracy
P1	[00000]	[10101]	[00000]	40%
P2	[10000]	[01010]	[00100]	40%
P3	[01000]	[11001]	[10001]	80%
P4	[00100]	[00110]	[00110]	100%
P5	[00010]	[10011]	[10011]	100%
P6	[00001]	[01100]	[01100]	100%

Pattern	Input	Target	Output	Accuracy
P7	[11000]	[00101]	[10101]	80%
P8	[00110]	[11010]	[10010]	80%

Table 5. *Per-pattern results for seed 314 (best seed). Three patterns achieved perfect accuracy. P1 ([00000] input) fails structurally: with no input neurons firing, no signal propagates through the network. This is an architectural limitation, not a learning failure.*

Excluding the structurally impossible P1 pattern, seed 314 achieves 83% accuracy on learnable patterns. Across all seeds, the system consistently produces distinct output vectors per input — 4 to 8 unique output vectors per seed — confirming that the network is performing genuine input-dependent computation, not finding a constant output attractor.

7. Analysis

7.1 Why local perturbation succeeds where global update fails

The fundamental difference between Phase 1 (neuron-centric, global update) and Phase 1b (astrocyte-mediated, local perturbation) is not the reward signal — both use a single scalar. The difference is the number of synapses modified per evaluation step. Phase 1 modified all ~2,300 synapses simultaneously. Phase 1b modifies 200–900 synapses under 1–3 active astrocyte territories.

With 2,300 simultaneous changes, the probability that a single scalar reward reflects the contribution of any individual synapse is negligible. The signal-to-noise ratio for credit assignment is approximately 1/2300. With 200 changes under one astrocyte territory, the ratio improves to 1/200 — still noisy, but sufficient for statistical convergence over thousands of keep/revert cycles.

7.2 The astrocyte advantage over random cursors

The random cursor (Experiment 020) proves that local perturbation alone is sufficient. Astrocytes (Experiment 021 onwards) add three properties that improve efficiency. First, activity-dependent activation routes perturbations to regions currently processing input, avoiding wasted perturbations in dormant regions. This is reflected in the tenfold improvement in accept rate (38% vs. 3.7%). Second, fixed territories provide consistency — the same astrocyte revisits the same synapses across episodes, allowing statistical accumulation of evidence about which changes help. Third, self-adaptation (threshold adjustment, maturity-scaled exploration) produces an emergent curriculum: broad early exploration followed by focused refinement.

7.3 What the system does not achieve

Intellectual honesty requires acknowledging clear limitations. The task is small: 5-bit inputs, 5-bit outputs, 60 neurons, 2,300 synapses. Whether this mechanism scales to larger networks and more complex tasks is untested. Mean accuracy of 67.7% leaves substantial room for improvement. The P1 pattern (zero input) is structurally unsolvable without a bias mechanism. Late-training degradation was observed in 20,000-episode runs, suggesting that the maturity-scaling parameters need further tuning. And the comparison to established methods (backpropagation, evolutionary strategies, reinforcement learning baselines) has not been performed — we cannot claim that this mechanism is competitive with existing approaches, only that it works at all.

7.4 The value of negative results

Eighteen failed experiments are not wasted effort. They are evidence. Eligibility traces with global chemical reward do not produce pattern discrimination in this architecture. Propagation cycles through random recurrent wiring add noise, not useful intermediate representations. Hidden layers without credit assignment add capacity the learning rule cannot exploit. Flag gates saturate uniformly and do not create useful selectivity. Dampening filters create self-locking traps for silent neurons. Each negative result eliminated a hypothesis and narrowed the search space.

The single most valuable diagnostic was logging output vectors — a trivial check that would have revealed the constant-output attractor on experiment 1 and redirected all subsequent work. This is a methodological lesson: always verify that your network is producing different outputs for different inputs before interpreting accuracy numbers.

8. Relation to Existing Work

Alvarellos-González et al. (2012) demonstrated that Artificial Neuron-Glia Networks outperform standard ANNs on classification tasks, but retained backpropagation as the underlying learning rule. De Pitta and Bhatt (2024) modelled astrocytes as meta-plasticity agents in reinforcement learning, sharing our intuition about time-scale separation but retaining standard RL update rules. Kostadinov and Bhatt (2023) combined CNNs with astrocyte-driven short-term plasticity using stochastic gradient descent. The NEST framework (Jiang et al., 2025) supports large-scale neuron-astrocyte simulations but focuses on synchronisation dynamics rather than learning.

Our contribution is narrow but distinct: we test whether astrocyte-like entities are sufficient for learning without any synaptic plasticity rule. The neurons compute; the astrocytes learn. To our knowledge, this specific question has not been addressed computationally.

9. On Methodology: AI-Assisted Rapid Prototyping

This work was conducted over a single weekend by a software engineer (25 years in enterprise integration systems) with no neuroscience training, using Claude (Anthropic) as a research collaborator. The human provided domain intuition, experimental direction, and biological framing. The AI provided literature context, experimental design, statistical interpretation, and code generation.

Twenty-four experiments in two days would be difficult in a traditional computational neuroscience workflow. Each experiment required designing the condition, implementing the code, running 10 seeds, analysing results, diagnosing failures, and designing the next experiment. AI-assisted prototyping compressed this cycle to approximately 30 minutes per experiment.

This speed has a cost: the code is not optimised, the statistical analysis is informal (no confidence intervals, no significance tests beyond seed counts), and the experimental design evolved reactively rather than being pre-registered. These are legitimate criticisms. The counterargument is that exploratory research benefits from rapid iteration — the astrocyte mechanism was discovered precisely because the iteration cycle was fast enough to follow unexpected leads within a single session.

10. Conclusion

We explored whether local-only plasticity control with global scalar reward can produce pattern discrimination in small neural networks. The answer is yes. A population of astrocyte-like entities — fixed spatial territories, activity-dependent activation, random perturbation, keep/revert based on one scalar, self-adapting thresholds — is sufficient for learning without any synaptic plasticity rule. The mechanism produces genuine input-dependent computation confirmed by distinct output vectors per input pattern.

The path to this result was as instructive as the result itself. Eighteen experiments with neuron-centric learning mechanisms produced zero genuine learning, masked by a constant-output attractor that mimicked improvement. A trivial diagnostic (logging output vectors) would have caught this immediately. The subsequent six experiments demonstrated that the problem was update scope, not reward complexity, and that astrocyte-like local perturbation resolves credit assignment where global simultaneous update cannot.

We make no claims about biological brains, the optimality of this mechanism, or its competitiveness with established methods. We claim only that it works, that the evidence is documented in full including all failures, and that the exploration itself demonstrates the potential of human-AI collaboration for rapid experimental research.

The code, full experimental logs, and this document are openly available. We welcome replication, criticism, and extension.

References

Alvarellos-González, A., Pazos, A., & Porto-Pazos, A. B. (2012). Computational models of neuron-astrocyte interactions lead to improved efficacy in the performance of neural networks. *Computational and Mathematical Methods in Medicine*, 2012, 476324.

Bohmbach, K., Henneberger, C., et al. (2022). An astrocytic signaling loop for frequency-dependent control of dendritic integration and spatial learning. *Nature Communications*, 13, 7932.

De Pitta, M., & Bhatt, D. (2024). Astrocytes as a mechanism for contextually-guided network dynamics and function. *PLOS Computational Biology*, 20(5), e1012186.

Ghézali, G., et al. (2022). The role of astrocyte structural plasticity in regulating neural circuit function and behavior. *Glia*, 70(7), 1467–1483.

Hösli, L., Binini, N., Ferrari, K. D., et al. (2022). Decoupling astrocytes in adult mice impairs synaptic plasticity and spatial learning. *Cell Reports*, 38(10), 110484.

Jiang, H.-J., et al. (2025). Modeling neuron-astrocyte interactions in neural networks using distributed simulation. *PLOS Computational Biology*, 21(9), e1013503.

Kostadinov, D., & Bhatt, D. K. (2023). Artificial neural network model with astrocyte-driven short-term memory. *Biomimetics*, 8(5), 422.

Document generated February 2026. Murray & Claude.