WHEN WORLDS OUTLIVE SCARS: HOW PERSISTENT ENVIRONMENTS CARRY HISTORY ACROSS AGENT RESETS

Anonymous authors
Paper under double-blind review

ABSTRACT

Resetting an agent does not necessarily reset its learning process. If an earlier policy changed a persistent environment, a fresh policy can inherit a different training distribution even after capability, memory, replay, and parameters have been restored or cleared. We trace the resulting channel from past capability, through a constructed world and fresh post-reset experience, into newly learned parameters and later behavior. Rather than infer this pathway from endpoint correlations, we exchange the intermediate carriers: fresh behavior follows assigned worlds, learned effects follow assigned replay, equalizing full replay removes the parameter contrast, and removing the world after practice preserves it. The inherited contrast is neither intrinsically beneficial nor harmful. Holding the learned snapshots fixed while changing only future task geometry yields positive, negative, and approximately cancelling consequences. An exact model gives a finite regime in which capability composition selects construction, and controlled learned systems realize the pathway without capability labels. The complete lifecycle is established in one designed Fork family; other structures expose explicit failure conditions, so broader carrier validity remains open. Persistent environmental state can therefore act as a hidden history channel even when agent-side state is reset.

1 INTRODUCTION

Resetting an agent is not necessarily the same as resetting its learning process. If an earlier policy changed a persistent environment, a fresh policy can inherit a different training distribution even after the body has recovered and agent-side memory, replay, and parameters have been cleared.

Consider a constructive robot that undergoes a sustained loss of capability along one direction. It reorganizes its task around the motions that remain reliable and builds a support for that strategy. The body later recovers and the controller is replaced, but the support remains. A reset agent now encounters a world selected by a capability state it no longer has. It follows different trajectories, collects different transitions, and may learn a different controller. Recovery has restored the agent, but not the learning problem.

We study this external route for history. Let h denote past capability history, W the persistent world it helps select, $D(W)$ the fresh post-reset experience generated in that world, θ the parameters learned from that experience, and d a future task demand. The proposed channel is

$$h \longrightarrow W \longrightarrow D(W) \longrightarrow \theta \longrightarrow Y(d).$$

We use *scar* for an acquired, composition-specific capability change that persists through an adaptation episode but may later recover; the term names its historical role, not irreversible bodily damage. The paper does not claim that this lifecycle is ubiquitous. It asks whether one such channel can be instantiated cleanly, intervened on carrier by carrier, given existence conditions, and characterized by explicit structural failure conditions.

The strongest evidence comes from moving the suspected carrier itself. Swapping the constructed worlds between matched histories makes fresh behavior follow the assigned world. Swapping their transition records makes newly learned parameters and later outcomes follow the replay source. Giving both learners the same full replay removes the parameter contrast. Removing the constructed

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

world only after practice preserves it. History is therefore not inferred from an endpoint correlation: it is followed as it moves from world, to experience, to parameters.

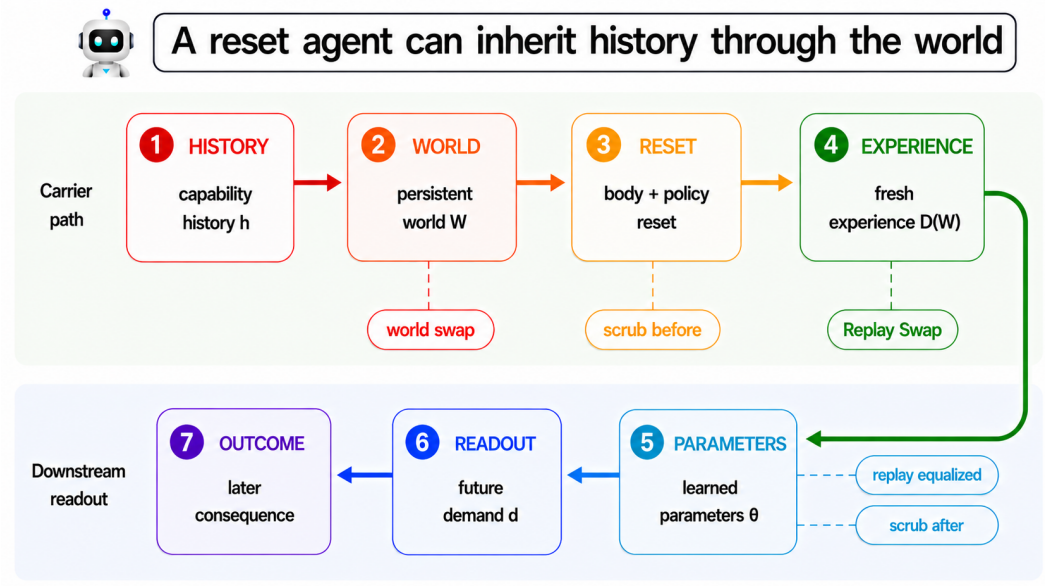

Figure 1: Carrier handoff across agent reset. Capability history (h) selects a persistent world (W). After the body and policy are reset, the retained world shapes fresh experience ($D(W)$), which trains new parameters (θ). Future demand (d) reads out these parameters to produce a later consequence. The interventions localize successive carriers: actual-world swap reassigns the retained world; scrub-before removes it before fresh practice; Replay Swap reassigns the actual transition buffer without changing construction history; genuine replay equalization removes the resulting parameter contrast; and scrub-after removes the world only after practice, testing whether the historical effect has already transferred into learned parameters.

The inherited parameter difference has no fixed value. We branch from the same learned snapshots, after the constructed world has been removed, into three geometrically matched future tasks. One reads the first practiced direction, one reads the other, and a balanced task reads both. The same history contrast is positive, negative, or approximately cancelled. A persistent scaffold is thus not intrinsically adaptive or maladaptive; its later value depends on what the world caused the agent to practice and what future geometry subsequently demands.

The complete positive lifecycle is demonstrated in one controlled Fork family, which remains the paper's principal limitation. Exact analysis and frozen failures show that construction and the full lifecycle have narrower existence conditions than the motivating story suggests. Minecraft supplies only upstream physical grounding, and an independently motivated clutter-manipulation extension stopped before capability-conditioned testing when its fresh-learning instrument failed frozen maturity criteria.

Our contributions are:

1. **A persistent-world history channel.** We formulate a four-link mechanism by which capability history can survive recovery and policy reset by changing the world that generates a fresh learner's data.
2. **Carrier-level interventions.** Actual-world exchange, Replay Swap, full-replay equalization, and scrub timing interventionally separate the handoffs from world to experience and from experience to new parameters.
3. **Demand-gated consequences.** Holding history, world-induced practice, learned parameters, current body, and current world fixed, future task geometry alone changes the learned contrast from beneficial to harmful to approximately cancelling.
4. **Existence and failure conditions.** A finite construction window and cross-generator failures locate when comparative advantage writes history into the world, when history specificity disappears, and when downstream coupling is absent.

2 Related Work

Curriculum learning and UED. Curriculum learning orders examples or tasks to improve learning, while automatic curricula and unsupervised environment design select environments using competence, progress, regret, or coverage signals (Bengio et al., 2009; Portelas et al., 2020a;b; Dennis et al., 2020; Jiang et al., 2021; Parker-Holder et al., 2022). Here no teacher chooses the later distribution and no learning-progress objective optimizes it. The distribution is an unintended physical consequence of what an earlier embodied policy found worth building. We use *endogenous curriculum* for this history-conditioned data-generation effect, not for a new curriculum method.

Performative feedback and continual learning. Performative prediction studies decisions that change the distribution on which future predictions are made (Perdomo et al., 2020); autocurricula can create sequential pressure and emergent tool use (Baker et al., 2020). Continual learning studies how parameters retain or acquire knowledge across changing experience (Parisi et al., 2019). Our object is a particular physical carrier across a particular reset: a world edit persists after capability recovery and replacement of the policy, can be literally reassigned or removed, and determines the transitions that train a new model. We follow that signal through world $\rightarrow$ experience $\rightarrow$ parameters, then remove the world before readout.

Damage adaptation and environment augmentation. Robots adapt to damage through behavioral search, reset-free trial and error, or body reconfiguration (Cully et al., 2015; Chatzilygeroudis et al., 2018; Kriegman et al., 2019). Robots also modify terrain or synthesize structures for immediate traversability (da Silva et al., 2018; Tosun et al., 2018; 2019). We claim neither damage adaptation nor construction alone. The central outcome is measured after capability recovery and policy reset; the contribution is the temporal composition from a prior capability state, through a persistent edit and fresh learning, to a later demand-dependent consequence.

Conceptual lineage: niche construction, external state, and practice. Niche-construction theory and ecological inheritance describe how organisms modify environments and how those modifications persist (Odling-Smee et al., 2003; Laland et al., 1999). Physically grounded evolutionary agents can build structures inherited by successors (Chiba et al., 2020), and stigmergic policies can use the environment as external memory (Peshkin et al., 1999). Recent RL work formalizes environmental artifacts as memory beyond the agent boundary, showing that spatial traces can reduce the internal memory needed for a performant policy (Martin et al., 2026). Work on inter-episodic state inheritance likewise separates transient agent state from persistent environment state and reuses critical trajectories across episodes (Chen, 2026). Robots can also choose what skills to practice (Kumar et al., 2024); rehabilitation distinguishes compensation from recovery and learned non-use (Taub et al., 1994; Jones, 2017). We treat these ideas as conceptual lineage. Our contribution is an ML-level carrier decomposition with explicit reset, literal interventions on intermediate physical and data carriers, newly learned parameters, world removal, and a future-demand branch that changes the sign of the consequence.

3 A Persistent-World History Channel

Let $c \in \mathbb{R}^q$ denote an agent's capability profile. During an adaptation episode, a sustained change δ produces $c^- = c - \delta$; the body may later return to c. A task $\tau \sim \mathcal{T}$ admits methods $m \in \mathcal{M}_\tau$ whose physical cost $J(m; c, W)$ depends on capability and world W. A generic edit $e \in \mathcal{E}(W)$ produces $T_e(W)$ at body-independent cost $K(e)$. Here $R(\tau; c_h, T_e(W))$ denotes the task return induced by the selected feasible method under the physical costs $J(m; c_h, T_e(W))$; the return includes the task-value/completion terms used by the planner, with costs entering negatively. The builder selects

$$e^*(h) = \arg \max_{e \in \mathcal{E}(W) \cup \{\varnothing\}} \mathbb{E}_{\tau \sim \mathcal{T}}[R(\tau; c_h, T_e(W))] - K(e), \tag{1}$$

where history h can affect the decision only through observed physical dynamics c_h, never through a condition label or condition-specific reward.

After editing, we restore the body to c, reset policy and transient memory, and retain or intervene on the actual world. Fresh interaction produces $D_W = \{(x, a, x', r)\}$ from a world-dependent

distribution p_W. A learner maps those transitions to new parameters $\theta = L(D_W; \theta_0, \xi)$, where θ_0 is the learner initialization and ξ collects the fixed learning procedure, replay order, optimizer randomness, and matched computational randomness. Unless explicitly reassigned by an intervention, θ_0 and ξ are matched within each paired comparison. We then remove the construction and evaluate a matched current agent on future geometry d, obtaining $Y = F(\theta, d)$. The complete mechanism separates into four gates:

$$\text{history} \xrightarrow{A} W \xrightarrow{B} \Delta D \xrightarrow{C} \Delta\theta \xrightarrow{D(d)} \Delta Y(d). \tag{2}$$

Gate A is comparative-advantage-driven construction; B is the effect of retained geometry on fresh experience; C is plasticity under that experience; and $D(d)$ is the future task's readout of the learned direction. Construction is necessary for this particular world-mediated pathway, but it is not sufficient for a later consequence. Proposition 1 is a restricted finite two-task specialization of this general return objective. Its g is the per-use generic benefit in the finite task comparison, d is the healthy cross-method disadvantage, and s is the capability-conditioned comparative-advantage shift. The horizon H counts repeated two-task uses, while K is the one-time construction cost. The proposition therefore instantiates, rather than generalizes, the trajectory-return formulation above.

The intervention design mirrors the chain. Actual-world swap exchanges W_0 and W_1 before fresh interaction. Replay Swap exchanges D_{W_0} and D_{W_1} after collection. Replay equalization gives both learners the same tuples, order, initialization, and random stream. Scrub-before removes the world before experience is generated; scrub-after removes it only after learning. We measure

$$\Delta Y(d) = F(L(D_{W_1}; \theta_0, \xi), d) - F(L(D_{W_0}; \theta_0, \xi), d) \tag{3}$$

with current world, body, inventory, task, and transient state matched. The local intuition $\Delta Y(d) \approx \nabla_\theta F(\theta, d)^\top \Delta\theta$ is useful, but no linear model is fitted or required.

All carriers share the same causal contract. Capability changes may alter primitive success, delay, energy, or reachability, but identical trajectories receive identical reward under permuted metadata. Construction material and edit physics are condition invariant. Body restoration resets dynamics; policy reset removes route caches, replay, recurrent state, and diagnostic history. The primary inferential unit is a topology skeleton unless stated otherwise; seeds, variants, cells, and transition steps are repeated measurements or coverage.

4 Moving History Across Carriers

The central experiment asks whether the historical effect follows the carrier we physically or computationally reassign. We first realize the world-to-experience link in a native tile environment. A directionally changed builder places generic blocks at coordinates chosen from the current map. We restore healthy body dynamics, clear the route policy and planner cache, and give a fresh agent the retained tile array, the base array, or the array built under the opposite history. The reset agent replans only from the tiles it sees.

Retained layouts redirect fresh trajectories. Opposite histories produce anonymous primitive counts of 16/8 and 8/16; swapping the raw block arrays reverses those counts. In this carrier, a directly observed change in $D(W)$ is a reallocation of post-reset practice across anonymous primitive classes: the world changes which transitions are encountered and how often, while reward semantics and history labels remain fixed. A low-level controller learns from the executed trajectories, after which the blocks are scrubbed and an opposite later capability change is imposed under matched current body, world, inventory, start, task, and transient state. This complementary second-hit challenge is applied only after the world is scrubbed, so the readout tests whether world-induced practice has been inherited by the learned controller rather than benefiting directly from retained geometry. Operationally, the reported contrast is the mean of the two complementary second-hit joint-task success-rate differences, one for each hit direction. Across ten controller/model seeds sharing 40 matched worlds, the retained-history contrast is $+0.08195$ with seed-level interval $[0.07216, 0.09175]$. World swap reverses it, scrub-before removes it, and scrub-after preserves it. The interval is stably above zero, though not wholly above 0.08.

The native study realizes an end-to-end physical retained-world pathway, but does not independently isolate the experience-to-parameter handoff. We therefore isolate that handoff separately in Fork.

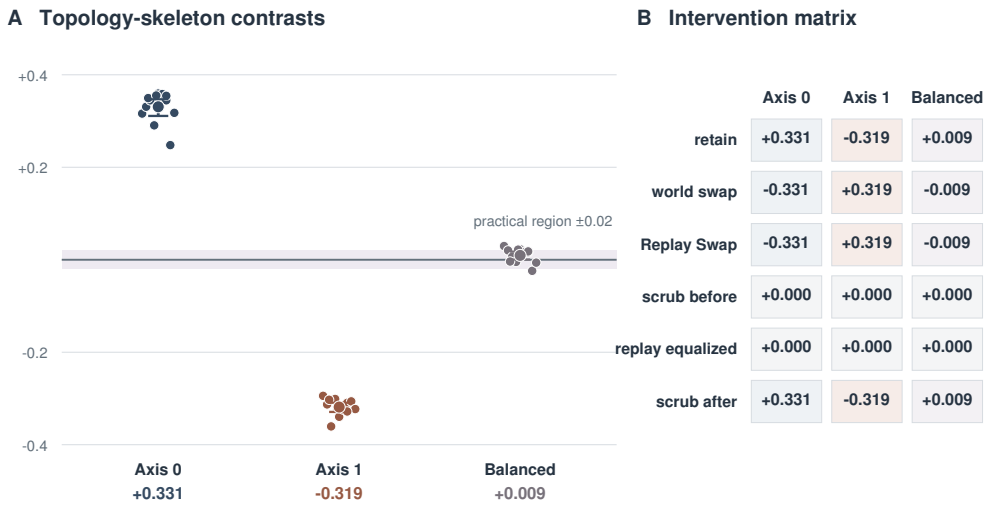

Figure 2: The historical effect follows the carrier that is reassigned. Panel A shows topology-skeleton contrasts for three future-demand readouts; each dot is one of 12 independent Fork topology-skeleton contrasts, with variants and learner seeds treated as repeated measurements. Panel B reports the same readouts across carrier interventions (rows): actual-world swap reverses collected experience, Replay Swap reverses the learned readout without changing construction history, replay equalization and scrub-before remove the contrast, and scrub-after preserves it. Directed signs agree for all 12 skeletons; the shaded band marks the practical region ± 0.02.

Each retained world first generates its own actual transitions (x, a, x', r). Only after collection do we retain, exchange, or equalize those records. Actual-world swap intervenes on Gate B; Replay Swap and full-replay equalization intervene on Gate C without moving the world at the same time.

The downstream learner is a deliberately small nonlinear function approximator, not an online deep-RL benchmark. A shared 9-32-32-4 MLP maps seven scaled local variables and a two-dimensional anonymous primitive token to four one-step reward predictions, one for each low-level control action. Each natural history supplies 24 actual transitions; ten frozen initializations receive 400 deterministic full-buffer Adam updates with Huber loss. No capability label, morphology, world provenance, audit axis, or oracle action value enters training.

The carrier handoff matches the pre-specified qualitative intervention pattern (Figure 2). Actual-world swap makes routes and collected transitions follow world assignment. Replay Swap leaves construction history unchanged but makes parameters and later outcomes follow the opposite buffer. Giving both histories identical tuples, order, initialization, and random stream removes the parameter contrast. Scrubbing after learning preserves it. The signal is therefore neither a surviving policy nor metadata: the world selects experience, and experience becomes persistent in new parameters.

5 Future Demand Assigns the Sign

A learned difference is not intrinsically a benefit or a cost. Its consequence depends on whether a future task uses the direction that practice strengthened. We isolate this demand dependence by holding the entire upstream history fixed and changing only the future geometry that reads the saved parameters.

All 22 eligible variants from all 12 Fork topology skeletons enter the experiment; world variants and learner seeds are repeated measurements. Two directional histories produce opposite constructed worlds, fresh routes, and practice distributions. After practice we save the learned parameters and scrub the world. Only then do we branch into three reward- and length-matched tasks: one geometry emphasizes axis 0, another emphasizes axis 1, and a balanced geometry requires both symmetrically. History, construction, retained-world practice, parameter snapshot, current body, and current world are shared across readouts.

The transparent controller yields the pre-specified sign pattern. Axis-0 demand favors history 0 by $+0.2136$ with topology-cluster interval $[0.2002, 0.2270]$; axis-1 demand favors the opposite history, giving -0.2096 $[-0.2261, -0.1931]$; balanced demand gives -0.0041 $[-0.0176, 0.0093]$. All 12 skeletons match the positive/negative/cancellation prediction. Exact paired sign-flip tests give $p = 0.000488$ for each directed arm, and cluster bootstrap and leave-one-skeleton-out analyses preserve both signs.

The MLP reproduces this geometry. Its directed effects are $+0.33083$ $[+0.31099, +0.34695]$ and -0.31853 $[-0.32908, -0.30917]$, again with 12/12 signs. Balanced demand leaves a small residual of $+0.00942$ $[+0.00041, +0.01753]$. Because the interval excludes zero, this is not a statistical null. It is about 3% of either directed magnitude and lies inside the internally frozen ± 0.02 practical region (and ± 0.025, but not ± 0.01), so we describe it as approximate cancellation.

The interventions establish more than the fact that different data train different models. World assignment determines which experience is generated; replay assignment determines which direction is learned; post-practice scrub shows that parameters have taken over as the carrier; and future geometry determines how that learned direction is valued. The endpoint sign is therefore a property of coupling between learned direction and future demand, not a fixed property of the original capability history.

6 When Can History Enter the World?

The carrier handoff requires a world worth changing for one capability composition but not for ordinary controls. This creates a finite opportunity rather than a universal tendency to build.

Consider two tasks and two methods. Let d be the healthy disadvantage of using a cross-axis method, s the extra cost imposed on one method by a directional capability change, g the per-use benefit of a generic structure, H its reuse horizon, and K its construction cost.

Proposition 1 (Finite scar-conditioned construction window)**.** *Under symmetric task values, additive physical costs, body-independent construction cost, and no-build tie breaking, a directional capability change selects the cross-axis method, the retained structure remains useful after recovery, and the healthy agent does not build whenever*

$$s > d, \qquad g > d, \qquad H(2g - d) \leq K < 2Hg. \tag{4}$$

The admissible interval in K has width Hd.

The interpretation is direct. $s > d$ reverses comparative advantage; $g > d$ keeps the edit useful after recovery; the lower bound on K prevents healthy construction; and the upper bound lets repeated changed-body use repay the edit. Lower cost yields healthy-build, higher cost yields no-build, and the interval between them is the only region in which construction records capability composition. The return table and proof are in the appendix.

The empirical phase scans follow these boundaries (Figure 3). The exhaustive grid enumerates $d \in \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0\}$, $g \in \{0.5, 1.0, \ldots, 4.0\}$, $s \in \{0.5, 1.0, \ldots, 6.0\}$, $H \in \{1, 2, 3\}$, and $K \in \{0, 0.5, \ldots, 20\}$: $6 \times 8 \times 12 \times 3 \times 41 = 70,848$ cells. The observed predicate requires bare-world scar substitution before construction and is equivalent on all cells to $s > d$, $g > d$, and $H(2g - d) \leq K < 2Hg$; it yields 3,468 positive cells with zero observed/predicted disagreements. Separately, a random parameter box uses seed 20260810, task-cost perturbations ± 0.5, relative frequency 0.7–1.3 times the two-cycle frequency, build cost 6.5–9.5, artifact gain 2.0–3.2, and scar penalty 3.5–6.5 across 5,000 draws, of which 2,054 are full structures. Here H is implemented by the planner parameter `repeats`: task costs and completion bonus are multiplied by `repeats`, while `build_cost` is subtracted once. With body-independent graph physics and a direct-repair competitor, positive exact cells remain in Branch (100/180), Loop (84/180), and Serial (100/180). These are coverage results within declared parameter boxes, not independent samples or real-world prevalence estimates.

6.1 Learning local physics, planning global edits

The agent is not told which capability changed or what morphology to build. It observes anonymous primitive identities with local success, time, and energy, and fits a shared local transition model. For

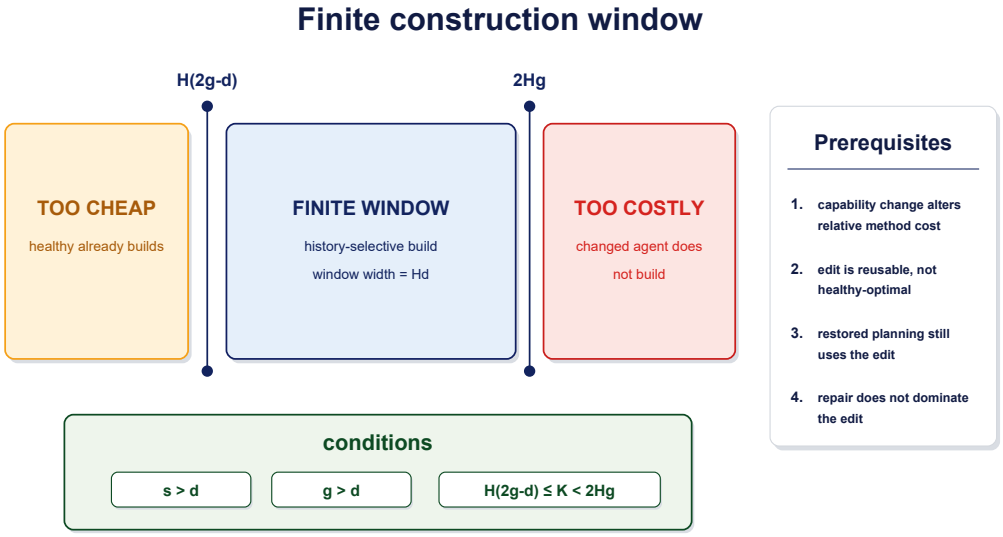

Figure 3: Finite construction window. History-selective construction occurs only within a bounded cost interval. A capability change must alter the relative method cost ($s > d$), the resulting edit must remain useful after restoration ($g > d$), and the construction cost must satisfy $H(2g - d) \leq K < 2Hg$. Below the lower boundary, construction is already worthwhile for the healthy agent; above the upper boundary, the capability-changed agent no longer builds. The resulting opportunity window has width Hd. The mechanism additionally requires the edit to be reusable but not healthy-optimal, restored planning to continue using it, and repair not to dominate the edit. Within the declared parameter boxes, exact coverage is 100/180 for Branch, 84/180 for Loop, and 100/180 for Serial; these counts characterize the declared search domain rather than real-world prevalence.

each legal generic edit, an explicit planner copies the world, applies the same body-independent edit physics, replans under learned transition costs, subtracts construction cost, and compares the result with no-build and repair.

This factorization matters on the tested structured shift. A frozen end-to-end graph policy, including a full-graph GNN added after an exact local-observation alias was discovered, fails blind Diamond/Merge evaluation for all ten seeds (four-condition accuracy 0.2375–0.5734). This is a diagnostic result about that baseline, not a general impossibility claim about amortized learning. With the same capability-label-free interface, learned local dynamics plus explicit planning reaches 1.000 exact-equivalent accuracy in all four conditions across ten seeds on V6 (the sealed causal-matched version-6 evaluation set) at 24 probes; the sealed set contains 48 matched quadruples from 108/192 exact-eligible base worlds.

In the native tile carrier, named sites are removed. The planner enumerates legal coordinates from the tile array and applies one generic block. On 40 sealed worlds, agreement with the exact native planner rises from 0.9956 at four probes to 1.000 at 24, versus 0.4625 for a reactive baseline and 0.25 for no-build. Reflecting cells, endpoints, and the agent reflects all 1,600 final placements. These experiments establish learnability of the upstream construction step; they do not turn factorization into a universal algorithmic claim.

7 Where the Pathway Stops

Scar-conditioned construction is broader than the complete history-to-learning pathway. In a one-shot test of three locked generator grammars, opposite directional layouts occur in 84/144 unconditional variants: 34/48 Fork, 13/48 Diamond/Merge, and 37/48 Serial Relay. Yet none of the 144 variants jointly satisfies the pre-specified healthy/common-mode, direct-repair, and nonzero practice-exposure requirements for a full lifecycle.

The four gates explain why the families fail differently (Figure 4). Fork reaches construction, fresh experience, and plasticity; future demand determines whether their consequence is exposed or cancelled. Diamond/Merge often fails at Gate A because a braided alternative route already absorbs

directional-0 damage; a zero-margin calculation predicts all 48 place/no-place decisions. Serial Relay fails specificity for the opposite reason: shared reuse makes the edit worthwhile without a directional history. Healthy agents build in 34/48 variants and common-mode agents in 42/48. These are structural absences of prerequisites, not failed repetitions to be averaged away.

7.1 PHYSICAL GROUNDING WITHOUT LIFECYCLE BREADTH

Minecraft isolates a narrower question: can the upstream carrier survive a real block engine? In a controlled Java 1.16.5 micro-arena, the agent executes native `use`; a voxel changes from air to dirt and inventory from 8 to 7. An agent-only reset retains the block, hard scrub removes it, and swapping actual block states makes fresh behavior follow the assigned world. The frozen unconditional pool contains the upstream contrast in only 18/270 cases (6.67%), all from one topology template. Minecraft therefore grounds placement, inventory, persistence, reset, scrub, and world swap–not practice-to-parameter learning, visual control, or open-world generalization.

A frozen planar-fixture attempt provides a different non-result: all 30 skeletons select no-build because candidate pegs almost never enter or improve the control trajectory. This is an implementation-level functional-edit failure, not evidence against manipulation as a carrier.

We also prospectively explored an independently motivated tabletop-clutter carrier. Ordinary object rearrangements passed a body-agnostic functional audit as native, persistent, swappable, and scrubbable world edits. However, the fresh-learning instrument did not satisfy frozen maturity criteria before capability-conditioned confirmation. We therefore did not open the history-specific lifecycle test and draw neither a positive nor a negative conclusion about capability-conditioned clutter preparation. This extension separates a functional persistent-world carrier from an instrument stable enough to test the complete learning handoff; the development stages and exact stopping criteria are reported in Appendix H.

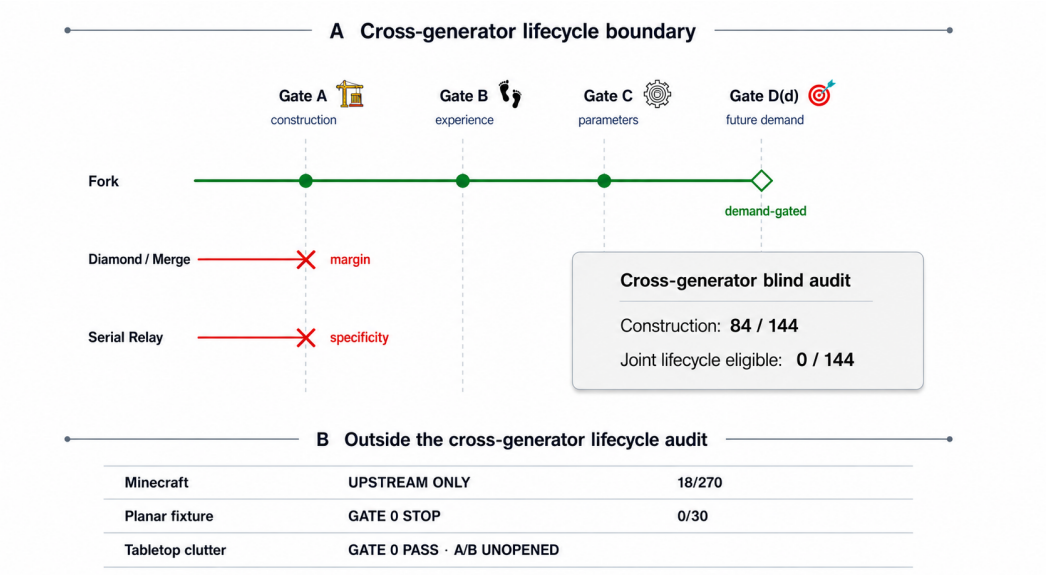

Figure 4: Different structures and carriers stop for different reasons. Fork completes Gates A–C and leaves the consequence demand-gated. Diamond/Merge often lacks construction margin; Serial Relay often makes construction too generally useful to remain history-specific. Minecraft grounds a narrow upstream physical carrier. The planar fixture fails the body-agnostic functional-edit sanity check, whereas clutter passes that check but stops during learning-instrument development before any capability-conditioned Gate A/B test. The Fork lane summarizes the controlled positive lifecycle established earlier; the 84/144 and 0/144 counts refer to the separate locked cross-generator audit. Only the first three rows enter that audit.

8 DISCUSSION

The paper's main claim is an existence-and-localization result about learning systems with persistent worlds. Agent-side reset can leave a history channel intact when an earlier policy has changed the environment that will generate a fresh learner's data. The carrier handoff matters because it distinguishes this mechanism from a story in which old metadata, an unrecovered body, or a surviving policy directly determines the outcome. Moving the world moves experience; moving replay moves new parameters; removing the world after learning does not remove the parameter consequence.

Future demand then prevents a simplistic resilience or vulnerability interpretation. A scaffold that concentrates practice along one direction can help a matching task, hurt an opposing task, or approximately cancel under balanced geometry. Immediate return is therefore insufficient for evaluating a persistent edit. Its long-run consequence depends on plasticity and on the tasks that later read the learned direction.

The mechanism has identifiable prerequisites. Capability composition must alter comparative advantage; the edit must be useful but not already optimal when healthy; retained geometry must change fresh experience; the learner must remain plastic; and future demand must project onto the learned contrast. Proposition 1 locates the first requirement in a finite economic regime. Diamond/Merge and Serial Relay show how alternative routes and shared reuse remove Gate A or its history specificity. Balanced demand shows that even successful Gates A–C need not change the final outcome.

The dominant scientific limitation is breadth and naturalism: the complete positive A–D lifecycle remains concentrated in one designed Fork family with 12 independent topology skeletons, and its clean realization uses controlled state representation, explicit planning, and structured future-demand geometry. Demand arms are structured and approximately symmetric. Observations are privileged state, not pixels. The local dynamics family, generic edit physics, explicit planner, and primitive-token-conditioned controller representation provide design scaffolding. The MLP receives 24 real transitions followed by extensive replay, so it tests nonlinear plasticity under controlled data rather than online deep RL. Minecraft provides only narrow upstream grounding. The clutter extension establishes a functional manipulation edit but no capability-conditioned result. We test neither a physical robot nor broad open-world learning.

These boundaries motivate a concrete sequence of tests for future carriers. A new carrier should first demonstrate a functional edit, then ask whether capability composition selects it, whether the retained world redistributes post-reset experience, whether the resulting data change new parameters, and whether natural future geometry reads that direction. Failure at one gate should not be reported as failure of another. More generally, experiments with persistent agents should treat reset semantics as including functional environment state: matching policy, memory, and body does not match history when the world still contains its consequences.

An agent's capability history therefore need not remain in its body, memory, or old policy. It can be written into a persistent world, encountered again as fresh experience, and absorbed into new parameters after the original capability state has vanished. What that inherited history ultimately does is decided by the future. Worlds can outlive scars, but their consequences are conditional on what the learner is later asked to do.

AI Use Statement. The authors defined the research objectives, made the final methodological and experimental decisions, determined the inclusion and interpretation of results, established the scope of the claims, and take full responsibility for the contents and conclusions of this work. Generative AI tools, including ChatGPT and Codex, were used as research and engineering assistants to support discussion and refinement of research questions and hypotheses, methodological discussion, code implementation and debugging, experimental workflow, analysis and interpretation, literature organization, and manuscript drafting and editing. All AI-assisted scientific outputs were reviewed by the authors before inclusion. **LLM outputs were not treated as experimental or bibliographic evidence without independent verification.**

REFERENCES

Bowen Baker, Ingmar Kanitscheider, Todor Markov, Yi Wu, Glenn Powell, Bob McGrew, and Igor Mordatch. Emergent tool use from multi-agent autocurricula. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=SkxpxJBKwS.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 41–48, 2009. doi: 10.1145/1553374.1553380.

Konstantinos Chatzilygeroudis, Vassilis Vassiliades, and Jean-Baptiste Mouret. Reset-free trial-and-error learning for robot damage recovery. *Robotics and Autonomous Systems*, 100:236–250, 2018. doi: 10.1016/j.robot.2017.11.010.

Guoyong Chen. Crossing multi-generational coordination tasks: Cultural reflow and inter-episodic state inheritance in zero-reward reinforcement learning, 2026. SSRN preprint, abstract ID 6599058.

Naoaki Chiba, Reiji Suzuki, and Takaya Arita. Evolution of complex niche-constructing behaviors and ecological inheritance of adaptive structures in a physically grounded environment. *Frontiers in Robotics and AI*, 7:45, 2020. doi: 10.3389/frobt.2020.00045.

Antoine Cully, Jeff Clune, Danesh Tarapore, and Jean-Baptiste Mouret. Robots that can adapt like animals. *Nature*, 521:503–507, 2015. doi: 10.1038/nature14422.

Maira Saboia da Silva, Vivek Thangavelu, Walker Gosrich, and Nils Napp. Autonomous adaptive modification of unstructured environments. In *Proceedings of Robotics: Science and Systems*, 2018. doi: 10.15607/RSS.2018.XIV.070.

Michael Dennis, Natasha Jaques, Eugene Vinitsky, Alexandre Bayen, Stuart Russell, Andrew Critch, and Sergey Levine. Emergent complexity and zero-shot transfer via unsupervised environment design. In *Advances in Neural Information Processing Systems*, volume 33, pp. 13049–13061. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/985e9a46e10005356bbaf194249f6856-Abstract.html.

Minqi Jiang, Edward Grefenstette, and Tim Rocktäschel. Prioritized level replay. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 4940–4950. PMLR, 2021. URL https://proceedings.mlr.press/v139/jiang21b.html.

Theresa A. Jones. Motor compensation and its effects on neural reorganization after stroke. *Nature Reviews Neuroscience*, 18:267–280, 2017. doi: 10.1038/nrn.2017.26.

Sam Kriegman, Stephanie Walker, Dylan S. Shah, Michael Levin, Rebecca Kramer-Bottiglio, and Josh Bongard. Automated shapeshifting for function recovery in damaged robots. In *Proceedings of Robotics: Science and Systems*, 2019. doi: 10.15607/RSS.2019.XV.028.

Nishanth Kumar, Tom Silver, Willie McClinton, Linfeng Zhao, Stephen Proulx, Tomas Lozano-Perez, Leslie Pack Kaelbling, and Jennifer L. Barry. Practice makes perfect: Planning to learn skill parameter policies. In *Proceedings of Robotics: Science and Systems*, 2024. URL https://arxiv.org/abs/2402.15025.

Kevin N. Laland, F. John Odling-Smee, and Marcus W. Feldman. Evolutionary consequences of niche construction and their implications for ecology. *Proceedings of the National Academy of Sciences*, 96(18):10242–10247, 1999. doi: 10.1073/pnas.96.18.10242.

John D. Martin, Fraser Mince, Esra'a Saleh, and Amy Pajak. Artifacts as memory beyond the agent boundary, 2026. URL https://arxiv.org/abs/2604.08756.

F. John Odling-Smee, Kevin N. Laland, and Marcus W. Feldman. *Niche Construction: The Neglected Process in Evolution*. Princeton University Press, 2003.

German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019. doi: 10.1016/j.neunet.2019.01.012.

Jack Parker-Holder, Minqi Jiang, Michael Dennis, Mikayel Samvelyan, Jakob N. Foerster, Edward Grefenstette, and Tim Rocktäschel. Evolving curricula with regret-based environment design. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pp. 17473–17498. PMLR, 2022. URL https://proceedings.mlr.press/v162/parker-holder22a.html.

Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 7599–7609. PMLR, 2020. URL https://proceedings.mlr.press/v119/perdomo20a.html.

Leonid Peshkin, Nicolas Meuleau, and Leslie Pack Kaelbling. Learning policies with external memory. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pp. 307–314. Morgan Kaufmann, 1999.

Rémy Portelas, Cédric Colas, Katja Hofmann, and Pierre-Yves Oudeyer. Teacher algorithms for curriculum learning of deep rl in continuously parameterized environments. In *Proceedings of the Conference on Robot Learning*, volume 100, pp. 835–853. PMLR, 2020a. URL https://proceedings.mlr.press/v100/portelas20a.html.

Rémy Portelas, Cédric Colas, Lilian Weng, Katja Hofmann, and Pierre-Yves Oudeyer. Automatic curriculum learning for deep rl: A short survey. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pp. 4819–4825, 2020b. doi: 10.24963/ijcai.2020/671.

Edward Taub, Jean E. Crago, Louis D. Burgio, Thomas E. Groomes, Edward W. Cook, Stephanie C. DeLuca, and Nancy E. Miller. An operant approach to rehabilitation medicine: overcoming learned nonuse by shaping. *Journal of the Experimental Analysis of Behavior*, 61(2):281–293, 1994. doi: 10.1901/jeab.1994.61-281.

Tarik Tosun, Jonathan Daudelin, Gangyuan Jing, Hadas Kress-Gazit, Mark Campbell, and Mark Yim. Perception-informed autonomous environment augmentation with modular robots. In *2018 IEEE International Conference on Robotics and Automation*, pp. 6818–6824, 2018. doi: 10.1109/ICRA.2018.8463155.

Tarik Tosun, Cynthia Sung, Colin McCloskey, and Mark Yim. Optimal structure synthesis for environment augmenting robots. *IEEE Robotics and Automation Letters*, 4(2):1069–1076, 2019. doi: 10.1109/LRA.2019.2893879.

A Supplementary Material

Appendix Roadmap

This appendix supplies the algebra, intervention definitions, learning protocols, cluster-level statistics, unsuccessful fixture replication, and artifact provenance behind the claims in the main paper. It preserves the separation between exact existence results, learned realizations, causal interventions, and physical grounding.

B Deriving the Finite Construction Window

B.1 Return table and proof of Proposition 1

Consider two tasks, two methods, reuse horizon H, per-use generic benefit g, healthy cross-method disadvantage d, directional comparative-advantage shift s, and construction cost K. Common task-return terms cancel when build is compared with no-build, so the table reports relative values. The Healthy and Directional rows are two-task construction totals; the Restored-body row is the affected-reuse marginal value after the one-time construction cost is sunk.

Capability state	No build	Build	Build advantage
Healthy	0	$H(2g-d)-K$	$H(2g-d)-K$
Directional episode	0	$2Hg-K$	$2Hg-K$
Restored body, edit sunk	0	$H(g-d)$ on affected reuse	$H(g-d)$

Under healthy capability, the generic structure benefits two repeated uses but induces the cross-method disadvantage once per two-task reuse cycle, because only the affected task uses the cross-method route. This gives gross advantage $H(2g-d)$. Under directional change, $s > d$ makes the affected task substitute to the other method; in the symmetric model this removes the healthy cross-method penalty from the scar-phase construction comparison, leaving gross advantage $2Hg$. Thus, with no-build tie breaking, healthy no-build and directional build require

$$K \geq H(2g-d), \qquad K < 2Hg.$$

After restoration, the retained edit remains useful on the affected reuse when $g > d$. Combining the three requirements gives

$$s > d, \qquad g > d, \qquad H(2g-d) \leq K < 2Hg.$$

The window width is $2Hg - H(2g-d) = Hd$. Lower K produces healthy-build; higher K produces no-build. This proposition assumes additive symmetric costs, a fixed reuse horizon, and the stated tie rule. It is a finite-model boundary result, not a universal theorem about construction.

C What Each Experiment Establishes

C.1 Separating evidence layers

The experiments answer different questions and are not pooled into one sample. Seeds, variants, cells, and transitions are never promoted to independent topology samples.

Layer	World coverage	Learner / intervention	Primary unit	Strongest supported use
Exact construction	Analytic grid; multiple graph grammars	Exact planner; repair competitor	Deterministic cell / skeleton	Existence window and failure margins
Factored planning	V6 and native-tile sealed pools	Learned local dynamics plus edit search	Model seed; worlds are coverage	Label-free construction on eligible worlds
Carrier handoff	12 Fork skeletons, 22 variants	World/Replay Swap, two scrubs, equalized replay	Topology skeleton	Interventional separation of world, experience, and parameters
Function approximation	Same Fork skeletons; 10 seeds	Shared 9-32-32-4 MLP	Topology skeleton; seeds repeated	Nonlinear learner absorbs the experience contrast
Breadth / grounding	36 cross-generator skeletons; Minecraft pool	Frozen audits and upstream interventions	Skeleton or declared generator cluster	Structural boundaries, not broad lifecycle

C.2 Intervening on successive carriers

Intervention	Manipulated carrier	Interventional implication
Actual-world swap	$W_0 \leftrightarrow W_1$ before fresh interaction	Fresh experience follows world assignment (Gate B)
Scrub before practice	$W_i \to W_{base}$ before interaction	Removing the edited world removes practice redistribution
Replay Swap	$D_{W_0} \leftrightarrow D_{W_1}$ before learning	Parameters and readout follow replay source (Gate C)
Replay equalization	$D_0 = D_1$ including order, initialization, and RNG	Holding full transition experience fixed removes the parameter contrast
Scrub after practice	$W_i \to W_{base}$ after learning	The learned contrast persists after world removal
Matched exogenous world	Same geometry, different provenance	Functional geometry, not history labels, controls behavior

The historical native count-equalization arm shared the same pre-practice world scrub as scrub-before-practice, so those were not independent interventions. It supports the native lifecycle but is not used to claim a separately manipulated experience carrier. Independent Gate-B/Gate-C separation in the main paper relies on actual-world assignment, full-transition Replay Swap, genuine replay equalization, and post-practice scrub.

D Native and Nonlinear Learning Protocols

D.1 Native-tile provenance

The native carrier dynamically enumerates legal block coordinates and applies one generic edit to copied tile arrays. Its development generator produced 157/192 causal-eligible worlds; this is conditional within-carrier coverage from a sealed development carrier, not construction prevalence across task families. Forty sealed worlds entered the learned-planning evaluation. Ten model seeds obtained exact-action agreement from 0.9956 at four probes to 1.000 at 24 probes, versus 0.4625 for a reactive baseline and 0.25 for no-build.

For the native lifecycle, retained geometry produced anonymous-token practice counts 16/8 versus 8/16; actual-world swap reversed them. Ten controller/model seeds sharing 40 matched worlds yielded a post-scrub normalized contrast $+0.08195$ with seed-level interval $[0.07216, 0.09175]$. A separate modular carrier yielded 0.4904 $[0.4712, 0.5000]$ over 26 generator-seed clusters. These results are not pooled because their physical and learning abstractions differ.

D.2 NONLINEAR FUNCTION-APPROXIMATION REPLICATION

The seven coordinates are ordered as: (1) Manhattan step length (scale 2), (2) source-tile top height (4), (3) target-tile top height (4), (4) target-minus-source height difference (2), (5) target-surface roughness (2), (6) a binary midpoint-void indicator (1), and (7) a binary target-block indicator (1). The feature builder divides the raw tuple by $[2, 4, 4, 2, 2, 1, 1]$ with no sign flip, clipping, or offset, then appends the two-dimensional one-hot token. The learner input is a nine-dimensional vector: seven scaled local native variables and a two-dimensional one-hot encoding of an anonymous primitive token. The network is 9-32-32-4 with ReLU hidden units. Token input weights are initialized symmetrically at zero; all other weights vary across ten frozen initialization seeds. The learner minimizes Huber loss with deterministic NumPy Adam at learning rate 0.003 for 400 full-buffer updates. The four outputs are one-step reward predictions for four discrete control actions; evaluation samples actions with a temperature-softmax rule at temperature 0.55. Each natural history supplies 24 actual transition tuples. This controlled replay experiment asks whether a nonlinear approximator absorbs a history-dependent experience contrast; it is not online deep RL. No morphology, capability label, axis audit, world provenance, or oracle action value enters training.

Intervention	Axis-0	Axis-1	Balanced
Retain	$+0.33083$	-0.31853	$+0.00942$
Actual-world swap	-0.33083	$+0.31853$	-0.00942
Replay Swap	-0.33083	$+0.31853$	-0.00942
Scrub before practice	0	0	0
Genuine replay equalization	0	0	0
Scrub after practice	$+0.33083$	-0.31853	$+0.00942$
Healthy / common controls	0	0	0

The balanced cluster-bootstrap interval $[+0.00041, +0.01753]$ excludes zero. We therefore do not call the MLP result a statistical null. The residual is approximately 3% of either directed magnitude and is interpreted as cancellation. Its interval lies inside the internally frozen ± 0.02 practical region and a ± 0.025 sensitivity region, but not inside ± 0.01.

E CLUSTER-LEVEL STATISTICAL PROCEDURE

The primary unit for coupling is the Fork topology skeleton ($n = 12$). World variants and learner seeds are repeated measurements. For each intervention and demand arm, outcomes are first averaged over all variants and seeds within a skeleton; inference is then performed on the 12 skeleton-level paired contrasts.

The two-sided paired sign-flip test enumerates all $2^{12} = 4096$ sign assignments. When all 12 directed contrasts share a sign, $p = 2/2^{12} = 0.00048828125$. Bootstrap intervals resample the 12 topology skeletons with replacement and report percentile quantiles at 0.025 and 0.975. The original tabular audit uses 50,000 draws with seeds 20260822, 20260823, and 20260824 for axis-0, axis-1, and balanced arms. The MLP audit uses 20,000 draws with frozen bootstrap seed 20260822. Leave-one-cluster-out removes one topology skeleton at a time after within-skeleton aggregation.

For the original demand experiment, the cluster-bootstrap intervals are:

$$\text{axis } 0 : [0.20182, 0.22457], \qquad \text{axis } 1 : [-0.22263, -0.19465].$$

All leave-one-cluster-out estimates preserve both directed signs. Balanced demand has mean -0.00414 and interval $[-0.01620, 0.00668]$, with six positive and six negative skeletons. For the MLP, both directed arms again have 12/12 consistent signs; the balanced residual is reported by magnitude and practical-region sensitivity rather than described as a null.

The ± 0.02 practical region was internally written and hash-frozen before the original coupling outcomes were revealed; it was not publicly registered. We therefore say *pre-specified* or *frozen before evaluation*, without implying public registration, and report ± 0.01 and ± 0.025 sensitivity so the interpretation does not rest on a single threshold.

F THE UNSUCCESSFUL PLANAR-FIXTURE REPLICATION

F.1 FROZEN PROTOCOL

The contact-rich carrier was frozen before confirmatory capability-conditioned outcomes. Development seeds 101, 202, and 303 were used only for interface, legality, determinism, healthy solvability, reset, scrub, and swap checks. The single confirmatory pool used skeleton seeds 51000–51029: 15 repeated-slot and 15 keyed-socket tasks. Version is `fixture-confirmatory-1.0.0`. The pool-manifest SHA-256 is recorded in full in the frozen lockfile (`5c5d5d52...7caec`).

Frozen component	Value / rule
Healthy dynamics	Translation $\sigma = 0.020$; rotation $\sigma = 0.015$
Translation degradation	Translation $\sigma = 0.120$; rotation $\sigma = 0.015$; translation energy $\times 1.15$
Rotation degradation	Translation $\sigma = 0.020$; rotation $\sigma = 0.120$; rotation energy $\times 1.15$
Matched common mode	Both $\sigma = 0.085$; both energy multipliers 1.075
Construction	Continuous `place_peg(x,y)`; 24 low-discrepancy coordinates; all legal singles plus 24 deterministic pairs; inventory two
Costs and reuse	Peg costs cycle 10/16/22; generic repair 20/26/32; three assembly tasks
Confirmation stop	At least 8 eligible skeletons and at least 4 per subfamily; otherwise stop before Gates C/D
Future demand	Physical long insertion, keyed rotation, and compound geometry; no reward-axis weights

No fixture cost, reuse horizon, severity, geometry distribution, candidate rule, or analysis rule could change after opening the pool; no confirmatory V2 was allowed. Candidate returns used common Monte Carlo seeds. Eligibility required healthy/common no-build or repair, distinct nonempty directional peg layouts with positive margins, and distinguishable post-reset practice distributions.

F.2 FROZEN OUTCOME AND FORENSIC BOUNDARY

The upstream stop failed: Gate A was 0/30 and joint A/B eligibility was 0/30. All conditions selected no-build. For translation degradation, the best peg layout had median net advantage -16.0 and median gross physical advantage 0.0 before construction cost; only 1/30 best layouts produced any contact. Rotation degradation likewise had median net -16.0, gross 0.0, and contact in 2/30. Gates C/D were not run and no revised carrier was generated.

This result does not establish a manipulation-wide mechanism boundary. The frozen pegs rarely entered or improved the native control trajectory, so the implementation did not create useful workholding. Its evidential role is transparency and claim limitation: it shows that an independently motivated carrier was attempted once and not tuned toward a positive outcome. It cannot serve as cross-carrier replication.

G STATISTICAL UNITS AND PROVENANCE

Native D2 reports seed-level uncertainty over 10 controller/model seeds sharing 40 selected worlds. Clean modular Gate C uses 26 generator-seed clusters. Cross-generator construction prevalence uses 144 variants grouped into 36 skeletons; joint lifecycle eligibility is 0/144. Minecraft's 18 eligible cells come from one topology template.

Earlier learner-backed lifecycle results with scar-specific reward or an audit-axis motor update are invalid and never used positively. The evaluated GNN is a diagnostic negative baseline, not evidence that amortized learning generally fails. V5 factored performance is diagnostic because its healthy control was ineligible. Minecraft 31.48% is withdrawn because of action-clock mismatch; only 18/270 is retained.

`RESULT_TO_SOURCE.csv` maps every headline number to an authoritative artifact, aggregation script, independent unit, SHA-256 hash, and paper location. Protocols described as pre-specified were internally written and hash-frozen before evaluation; no public preregistration is claimed.

H EXPLORATORY INDEPENDENT-CARRIER DEVELOPMENT: TABLETOP CLUTTER

This section records an independent-carrier extension that stopped before capability-conditioned confirmation. Its role is transparency about breadth effort and about the difference between a functional physical carrier and a mature downstream learning instrument. It is not an A–D result.

H.1 PROSPECTIVE SELECTION AND GATE-0 FUNCTIONALITY

Tabletop clutter was selected before outcome inspection because ordinary manipulation already motivates moving non-target objects to improve later retrieval or repeated transfer. Its persistent state is a set of rigid-body object poses and contacts rather than a route label or named build site. The body-agnostic Gate-0 protocol was hashed before evaluating an unconditional pool of 48 skeletons: 24 confined-retrieval and 24 repeated-transfer tasks. Each world exposed 40 native push candidates, 60 continuous grasp-place candidates, and no-edit under the same physics.

The functional audit passed its frozen criteria. Thirty-eight of 48 selected edits had gross utility at least 0.08; retrieval contributed 21/24 and transfer 17/24. The median gross utility among functional cases was 0.12805. Both push and grasp-place edits occurred, all five object identities were edited across the pool, and all 4,848 candidate outcomes were retained. Retain/reset, literal world swap, scrub, determinism, and provenance permutation passed on all eight frozen audit skeletons. Swap installed donor object positions and orientations rather than a layout identifier.

The initial Gate-0 execution was invalidated because it returned A* path statistics without executing the task path through the required Pymunk kinematic carrier. That run remains preserved. The implementation was corrected and the same frozen pool replayed without changing tasks, seeds, candidates, thresholds, or the learner smoke test. The byte-identical pre-fix source was not preserved; the invalid raw results, correction record, corrected source hash, and all unchanged scientific quantities are documented in the artifact.

Gate 0 established only that ordinary rearrangement can be a native, persistent, swappable, and scrubbable manipulation edit. It did not examine capability histories, construction economics, healthy specificity, or any Gate A–D outcome.

H.2 BODY-AGNOSTIC LEARNING-INSTRUMENT DEVELOPMENT

The capability holdout remained closed while a fresh learner was developed on ordinary tasks. The semantic firewall excluded capability, scar, history, provenance, morphology, generator identity, oracle actions or values, and future-demand definitions. The final candidate learned a structured local transition model from normalized native observations, anonymous action tokens, and observed next states, then used horizon-3 beam search. The environment, reward semantics, capability design, construction economics, Gate A/B criteria, and future geometries were outside the development loop.

Direct-control baselines failed mainly on lifted transfer. A legality-masked n-step Double DQN at 640 episodes obtained five-seed medians 0.450 overall, 0.875 ground, and 0.175 lifted; a return-weighted self-imitation variant obtained 0.325, 0.575, and 0.100. Successful native trajectories were present, so the failure was not complete absence of positive experience. An exact-dynamics forensic showed that one-step planning completed 256/256 ground cases and 0/256 lifted cases; lifted recovery required a three-action close–lift–move sequence. This motivated the factorized transition model and horizon-3 planner without altering the scientific carrier.

Frozen candidate	Two-mode seeds	Overall median	Lifted median	Formal decision
2,400 tuples	10/10	0.925	0.850	Fail: 4/10 at overall ceiling; maximum 2/10
V2.1, 1,600 tuples	8/10	0.7625	0.550	Fail: lifted threshold was 0.600

For the 2,400-tuple candidate, ground median was 1.000, random-legal overall median 0.03125, held-out transition RMSE 0.06938, and binary accuracy 1.000. Its strong competence did not override the frozen ceiling rule. The 1,600-tuple V2.1 candidate used new development seeds and the same model/planner; ground median was again 1.000, random-legal overall median 0.04375, RMSE 0.07418, and binary accuracy 1.000. Its lifted median of 0.550 was not rounded to the frozen 0.600 requirement. No checkpoint was substituted, no threshold was relaxed, and no further candidate was authorized.

H.3 Allowed interpretation and provenance

Allowed interpretation. Ordinary object rearrangement can instantiate a functional persistent manipulation carrier with literal retain, swap, and scrub semantics. The proposed fresh-learning instrument also improved materially during body-agnostic development, although it did not meet the frozen maturity requirements.

Not established. This sequence does not show that capability history selects clutter preparations, that healthy and directional conditions differ, or that clutter passes Gate A, B, C, or D. It is neither a positive nor a negative test of history-specific clutter preparation: no capability-conditioned Gate A/B result exists.

The authoritative Gate-0 report, implementation-correction record, V2 instrument-development report, final decision, and semantic-firewall report are indexed by path and SHA-256 in the v0.7 result-to-source ledger. The raw invalid execution is preserved separately from the corrected same-pool replay.