LEARNING TO CORRECT IS NOT ENOUGH: REUSING FEEDBACK IN WORLD MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate short-path correction does not guarantee reliable reuse along longer action paths. We study how a single observation can revise multiple cached futures of a world model without regenerating them. A matched construction adds one known return without changing the unknown response or calibration data, turning horizon-independent minimax decision risk into risk growing with path length until saturation. Legal continuations determine which response uncertainty is repeatedly exposed to the task. For known compact classes of orthogonal response dynamics on a fixed finite directed graph, we characterize when Gaussian-noisy scalar observations on a prescribed short-path bank suffice for reliable decisions along all legal continuations. Learning is possible exactly when, in every strongly connected component and at every positive task precision, only finitely many irreducible types of the joint return action remain task-significant across the candidate class. This permits infinitely many distinct return types when their task influence vanishes. Compatible internal coordinate changes cancel in composition, bounding approximation error per visited component rather than per action; under this criterion, longer paths need not demand finer calibration. We further characterize when all path responses admit a fixed finite exact linear calibration with model-independent coefficients, while showing that such exact linear tables can require exponentially many entries. On a continuous nominal base, even a smooth, uniformly observable class sharing one known response type can preclude uniform all-path learning from noisy short-path data. Controlled studies and simulated planning examine these distinctions and the trade-offs of cached correction. Calibration sufficiency therefore depends not only on what the data reveal, but also on which unresolved response differences future use can make consequential.

1 INTRODUCTION

A new observation can arrive after a world model has already imagined several possible futures. Revising those futures can avoid regeneration, but leaves a learning question: what must a correction rule learn to remain reliable beyond the short action paths used to calibrate it? Reusing a computation does not by itself justify reusing what was learned.

Consider two systems with the same unknown response and calibration data. Adding one known return action to one system lets it encounter that unresolved response repeatedly. In our matched construction, the best attainable worst-case decision risk is horizon-independent without the return, but grows with path length until

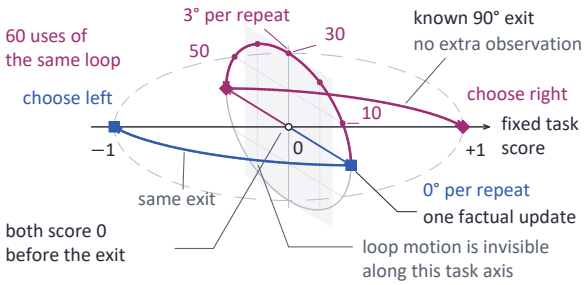


Figure 1: **An invisible response can change a later decision.** After one common fact, 0° and 3° per-loop models both score zero. After sixty loops, the same known quarter-turn exposes scores -1 and $+1$ and opposite decisions, without another observation. Curves show exact unit-response geometry, not measurements; dots mark ten-loop increments (Eq. (6); Appendix C).

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.

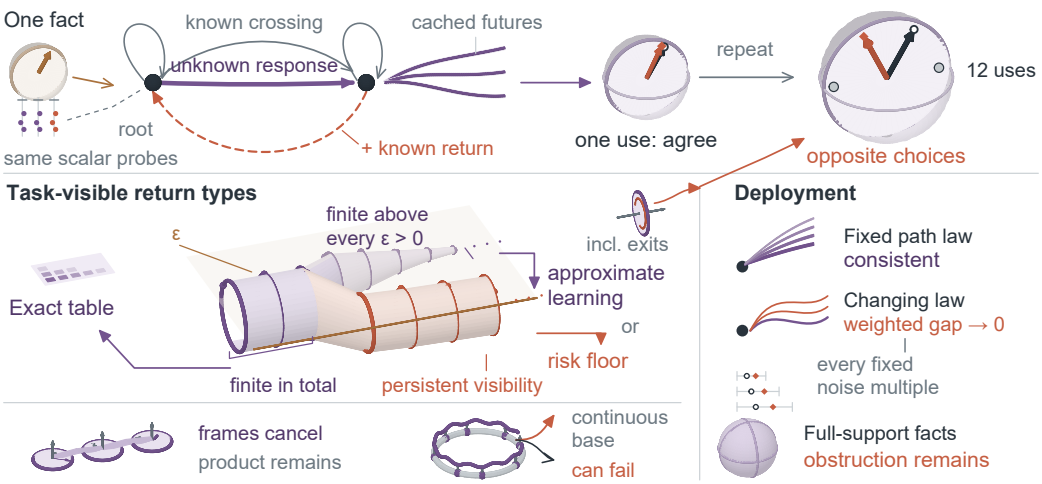


Figure 2: **Return structure and reliable reuse.** Top: calibration-near models (circle/diamond) disagree after adding a known return, without new data. Center: in a known compact orthogonal class on a fixed finite graph, ring radius measures task visibility over all legal suffixes, including exits, not norm decay. Finitely many visible types in total permit exact linear tables; the alternative infinite tails distinguish approximate all-path learning from a risk floor. Internal frames, not response products, cancel. Right and lower left: deployment-law and continuous-base boundaries (Theorems 4.4, E.1; Proposition F.1). Geometry is schematic.

saturation when the return is allowed. Every action preserves response magnitude, and both systems admit arbitrarily long paths. A later exit can also expose an invisible response difference and reverse the preferred decision (Figure 1).

Calibration is sufficient only relative to its intended future use. We characterize this dependence for a known compact class of orthogonal response dynamics on a fixed finite directed graph, calibrated through Gaussian-noisy scalar observations on a prescribed short-path bank.

The criterion concerns *return types*: elementary response patterns generated jointly by action loops, identified up to a common orthogonal change of coordinates. Repeated calibration yields vanishing worst-case expected decision regret across models, facts and legal paths exactly when each region of mutually reachable vertices (a strongly connected component) has finitely many significant types at every positive task precision, across the entire candidate class. Significance includes every legal continuation, including exits. Thus infinitely many distinct types can be compatible with learning: differences whose task influence vanishes need not all be resolved exactly.

The key is to distinguish response differences from internal coordinate choices. Compatible coordinate changes cancel under composition; response products do not. Retaining task-significant types then controls approximation error per visited graph component, rather than per action. Under this criterion, extending a path need not require finer calibration. Exact response reconstruction asks more: a fixed finite linear calibration table exists exactly when each component has finitely many visible types in total across the class. A supplied dictionary makes this constructive, although acquisition costs can be large.

Future use also changes the learning target: the prescribed calibration can support consistency under any fixed distribution of finite paths even when uniform all-path learning fails. This is not a guarantee uniform over all path distributions. A smooth continuous-base counterexample establishes a separate boundary to the finite-graph result. Calibration must therefore be matched to the response differences that future use can make consequential.

2 Related Work

Feedback and cached predictions. Feedback World Model uses execution observations to correct predictions (An et al., 2026); RIFT separates future representations from their generation (Zhang

et al., 2026). We ask when learned corrections remain reliable beyond calibration paths, with cached futures fixed before feedback and no additional branch observation or generator query.

Information sufficient for a task. Predictive states describe behavior through observables (Thon & Jaeger, 2015); proper value equivalence preserves planning information (Grimm et al., 2021), and bisimulation captures future task relevance (Zhang et al., 2021). We study whether noisy short-path calibration can learn these distinctions uniformly across models and legal continuations.

Identification versus reliable reuse. Exact identification of linear switched systems uses isomorphisms under structural minimality (Petreczky et al., 2010). Set-membership estimation and testing address observation uncertainty (Milanese & Vicino, 1991; Polyanskiy & Wu, 2026). Our matched construction changes future decision risk without changing calibration information, distinguishing exact identification from statistical stability under reuse.

Returns and coordinate freedom. Synchronization connects edge transformations with coordinates and holonomy (Gao et al., 2021); gate set tomography separates gauge freedom and uses repeated sequences to amplify observable errors (Nielsen et al., 2021). Our criterion concerns learning across a response class on legal directed paths, where inverses need not be executable. It links joint returns to task visibility, including later exits.

One calibration table across models. Finite-dimensional realization unifies automata, observable operator models and predictive states (Thon & Jaeger, 2015). Singular value automata approximate square-summable rational series (Balle et al., 2019); our responses need not decay with length. Our exact linear table requires coefficients shared across the candidate class and all horizons, not a separate realization for each model. Approximate learning need not admit such a table.

3 Problem and response interface

For root latent h and action path u, let $x_h^{u,0}$ be the nominal future and $x_h^{u,e}$ its regeneration after one common innovation e. The response for task probe g is

$$R_h(u; e, g) = g(x_h^{u,e}) - g(x_h^{u,0}). \tag{1}$$

Before feedback, compile–observe–revise fixes the futures and paths. Every editor reads the same cache, path address and innovation, without generator calls or branch-specific evidence. The rule $e \mapsto R_h(u; e, g)$ must transfer, not just fit one fact.

A connection supplies transports $\tau_{h,a}$ compatible with the nominal map Φ_a^0 and factual update U:

$$U(\Phi_a^0(h), \tau_{h,a}(e)) \approx \Phi_a^e(U(h, e)), \qquad \tau_{h,ua} = \tau_{\Phi_u^0(h),a} \circ \tau_{h,u}. \tag{2}$$

Training need not enforce this compatibility exactly. Nominal dynamics determine legal compositions; matrix inverses need not be executable. Neural transports implement the interface, without the theorem's finite-graph or orthogonality assumptions.

Prediction, response, and decision are distinct. Identical predictions can conceal opposite preferred plans (Appendix G.1). Endpoint-only response requires agreement on all legal suffixes; pooling can retain information endpoints discard (Appendix G.2). Common score shifts or large margins can preserve decisions despite response error. Our lower bounds establish a fixed-task conflict.

4 Finite calibration and the scope of reliable reuse

4.1 Graph, task and observations

Vertices determine legal continuations; edges are actions or chunks. Orthogonality isolates composition from norm growth: inter-model differences can accumulate or cancel without response decay. Let $G = (V, E)$ be a known finite directed multigraph, fixed as the deployment horizon varies, with every vertex reachable from root o. Each edge $a : x \to y$ has $T_a \in O(r)$, $r \geq 2$, and T lies in a known nonempty compact class $\mathcal{K} \subset O(r)^E$. Only legal paths compose, with $T_{pq} = T_q T_p$. A unit fact e enters once at o. For a fixed unit c, terminal actions $b \in \{-1, +1\}$ have

$$\ell_b(z) = \tfrac{1}{4}\|z - bc\|^2, \qquad r(b, f) = \tfrac{1}{2}(|f| - bf), \qquad f_T(p, e) = c^\top T_p e. \tag{3}$$

A wrong sign costs $|f|$; numerical reconstruction is not required. Put $v = |V|$, $L = (r+1)v$, and $D = r|\mathcal{P}_o^{\le L}|$. Calibration supplies

$$Y_{p,j,t} = c^\top T_p e_j + \xi_{p,j,t}, \qquad \xi_{p,j,t} \overset{\text{ind}}{\sim} N(0,\sigma^2), \quad \sigma > 0, \quad t \le n. \tag{4}$$

The nD scalar cost differences cover every root path of length at most L and every basis fact, potentially exponentially many paths. Learners may randomize with unrestricted memory and computation, but cannot query the true model at deployment. Define

$$\mathcal{R}_n = \inf_{\mathsf{L}} \sup_{T\in\mathcal{K}} \sup_{p:o\to *,\ \|e\|=1} \mathbb{E}_T r(\mathsf{L}(Y,e,p), f_T(p,e)). \tag{5}$$

The query supremum is outside the expectation: queries are fixed independently of calibration noise. This is terminal decision regret, not cumulative return.

Return types describe the *joint* action of loops, not individual spectra. In a strongly connected component C, fix a root z_C and internal paths $u_x : z_C \to x$ and $v_x : x \to z_C$, empty at z_C. The marked returns $u_x v_x$ and $u_x a v_y$, for internal edges $a : x \to y$, generate the same compact group as all legal returns at z_C. Their transports split $\mathbb{R}^r$ into minimal jointly invariant subspaces; a type π is a restricted tuple up to simultaneous orthogonal change of basis. Vertex-gauge equivalence means $S_a = Q_y T_a Q_x^\top$ internally and preserves these joint relations.

Task significance depends on *every legal suffix*, including exits that can reveal a currently invisible direction. Let $P_\pi(T)$ project onto all copies of π, or be zero if absent. Define

$$\begin{aligned} a_C(T,\pi) &= \sup_{s:z_C\to *} \|P_\pi(T) T_s^\top c\|, \\ \mathcal{F}_C(\epsilon) &= \{\pi : \exists T \in \mathcal{K},\ a_C(T,\pi) \ge \epsilon\}, \qquad \mathcal{F}_C^+ = \bigcup_{k\ge 1} \mathcal{F}_C(1/k). \end{aligned} \tag{6}$$

These are *class-wide* catalogues: each model has finitely many constituents, but their union across the candidate class need not be finite. The marking and root do not change these finiteness conditions (Appendix B).

Lemma 4.1 (Rigidity from legal returns)**.** *Within a strongly connected component, if* $\sup_{\ell:z\to z} \|T_\ell - S_\ell\|_{\mathrm{op}} < 1$ *at one vertex, then the internal connections are vertex-gauge equivalent. Conversely, a vertex gauge conjugates all marked returns by the same matrix at that vertex.*

4.2 Approximate and exact calibration

The theorem distinguishes decision sufficiency from exact reconstruction across the candidate class. Reliable approximate reuse needs finitely many significant types at each task precision; an exact linear table needs finitely many visible types in total.

Theorem 4.2 (Calibration under reuse)**.** *Under the preceding assumptions,* $\mathcal{R}_n \to 0$ *if and only if every* $\mathcal{F}_C(\epsilon)$ *is finite for every* $\epsilon > 0$*. Equivalently, for each* $\epsilon > 0$ *there are finitely many actual root paths* $p_1, \dots, p_N$ *and a selector* $i(p)$ *independent of* T *such that*

$$\sup_{T,p} \|c^\top T_p - c^\top T_{p_{i(p)}}\| \le \epsilon. \tag{7}$$

Failure gives $\mathcal{R}_n \ge a_{G,\mathcal{K}} > 0$ *at every finite* n*.*

Let $\mathcal{V}_\mathcal{K} = \mathrm{span}\{T \mapsto c^\top T_p e_j : p : o \to *,\ j \le r\}$*. Every* $\mathcal{F}_C^+$ *is finite if and only if* $B = \dim \mathcal{V}_\mathcal{K} < \infty$*, equivalently there are fixed scalar root probes* $q_i = (w_i, e_{j_i})$ *satisfying*

$$f_T(p,e) = a(p,e)^\top \mu(T), \quad \mu_i(T) = f_T(q_i), \qquad \sup_{p,\|e\|=1} \|a(p,e)\| = \Gamma < \infty. \tag{8}$$

Coefficients are model-independent and the minimum homogeneous exact linear-table dimension is B*. For semialgebraic* $\mathcal{K}$*, approximate calibration (*$\mathcal{R}_n \to 0$*) and exact calibration* (8) *are equivalent.*

Why the proof charges components, not actions. Summing per-edge error bounds incurs a path-length factor even when transports preserve norms. Retain finitely many types per component, with

all their copies, discarding only types of weight at most τ. At most r types occur per model, so $\|(I - P_{z_C})T_s^\top c\| \le \sqrt{r}\,\tau$ for every legal suffix. Transporting the retained projector gives $T_a\Pi_x = \Pi_y T_a$ internally: each entry projection moves through arbitrarily many internal actions. Telescoping over m components gives

$$\sup_{T,p} \|c^\top T_p - \widetilde{F}_T(p)\| \le m\sqrt{r}\,\tau.$$

Prefixes have norm at most one; suffixes have controlled discarded weight, including after exits.

Retained rows must share a description across $\mathcal{K}$. For each of finitely many retained type/multiplicity patterns, choose canonical internal transports M_a and isometries J_x with $T_a J_x = J_y M_a$. Intermediate coordinates cancel, not the M-products. Unknown entries remain in $J_o^\top$, $c^\top J_y$ and crossing matrices $J_y^\top T_a J_x$, all norm bounded by one. Finitely many crossing skeletons, each with at most $m-1$ crossings, expand into monomials of degree at most $m+1$; internal paths affect only known coefficients. Pointwise coordinate choices, extended by zero outside each pattern, give one bounded finite-dimensional family on $\mathcal{K}$. Auxiliary coordinates need not be continuous; query rows are. Arbitrarily accurate approximation makes these rows totally bounded, giving actual-path prototypes.

Why the original short bank suffices. Let W_x^k span task rows of suffixes from x of length at most k. The recursion $W_x^{k+1} = \operatorname{span}(c^\top, \{W_y^k T_a : a : x \to y\})$ stabilizes by $L_0 = rv - 1$: the dimension sum starts at v, is at most rv, and cannot resume growing once fixed. Root access paths have length at most $v-1$, so access, stabilized suffix and one-edge extension rows all occur in (4). In root coordinates these rows determine edge actions on the observable row spaces. Backward propagation identifies every task row without an inverse observation matrix or full observability.

Thus, for the original mean map Ψ and $d_\infty(T,S) = \sup_p \|c^\top(T_p - S_p)\|$, equality of Ψ implies $d_\infty = 0$. Under the type condition, total boundedness also gives equicontinuity. Compactness yields, for $0 < \epsilon \le 1$,

$$\delta_\epsilon = \min\{1, \inf_{d_\infty(T,S)\ge\epsilon} \|\Psi(T) - \Psi(S)\|_2\} > 0$$

(empty infimum $+\infty$). Fitting averaged observations to a finite $\delta_\epsilon/4$-net of $\Psi(\mathcal{K})$ gives all-path regret below ϵ when the noise norm is at most $\delta_\epsilon/4$. Gaussian tails give, for $0 < \beta < 1$,

$$n \ge \frac{32\sigma^2 D \log(2D/\beta)}{\delta_\epsilon^2} \quad\Longrightarrow\quad \Pr_T\{\sup_{p,\|e\|=1} \text{Regret} < \epsilon\} \ge 1 - \beta.$$

The original nD scalars suffice; separation and search cost are class-dependent.

Conversely, infinitely many significant types yield converging models with a visible type absent from the limit. Averaging paired legal returns annihilates cross-terms between inequivalent types. One common root fact and path, fixed before noise, then give opposite scores with a fixed margin; Gaussian two-point testing gives the floor. Zero truncation yields an exact table; a finite exact span cannot contain an additional visible unmatched type (Appendix B; Figure 2).

Approximate need not mean exact. On a one-vertex, one-action graph take $c = e_1$ and rotations R_x, R_z about the first and third axes. Let $a_j = 2^{-j}$, $\theta_j = 0.7 + 2^{-j}$ and $T_j = R_z(a_j)R_x(\theta_j)R_z(a_j)^\top$ for $j \ge 4$, together with $T_\infty = R_x(0.7)$. This compact class has distinct plane-rotation types with weight $\sin a_j$, and

$$\|(T_j^k)^\top c - c\| = 2\sin a_j |\sin(k\theta_j/2)| \le 2\sin a_j \quad (k \ge 0).$$

Positive but vanishing weights permit learning without a finite exact table. This class is not semialgebraic (Appendix C).

Observing each prototype row on all basis facts n times gives

$$\sup_{T,p,e} \mathbb{E}_T \,\text{Regret} \le \min\{1, \epsilon + \sigma\sqrt{2/(\pi n)}\}. \tag{9}$$

This costs nrN scalars and $nr\sum_i |p_i|$ action steps; an exact table gives $\Gamma\sigma\sqrt{2/(\pi n)}$ from nB scalars. These banks may differ from (4). Semialgebraicity gives a Hölder rate for the prescribed bank; full task observability gives a horizon-independent high-probability rate (Appendix B.5). Discovery may be costly; addressed selection is not autonomous finite-state execution.

A finite class with d independently signed crossings requires 2^d exact linear coordinates, whereas a nonlinear learner estimates d signs. Known constants need no sampling; this is not an unrestricted computational lower bound (Appendix C).

4.3 THE SAME CALIBRATION, WITH AND WITHOUT A RETURN

Proposition 4.3 (Matched information, different legal reuse)**.** *Take vertices x, y, root x, a known self-loop R_ω at x with ω/π irrational, a known self-loop q_y with response $R_{\pi/2}$ at y, and parallel crossings $x \to y$ with responses I and R_θ, $0 \le \theta \le \rho \le 0.1$; call the latter b. Compare the graph with and without a known identity return $y \to x$. Both observe the same n noisy repeats of paths b and bq_y on facts e_1, e_2. With* (5) *over paths of length at most integer $H \ge 2$ and these observations,*

$$\mathcal{R}^{\rightarrow}_{n,H} \asymp \min\{\rho, \sigma/\sqrt{n}\}, \qquad \mathcal{R}^{\mathrm{ret}}_{n,H} \asymp \min\{1, H\min(\rho, \sigma/\sqrt{n})\}. \tag{10}$$

The constants are absolute, and the one-way bound is uniform over all finite horizons.

The shared data reduce to $Y = R_\theta + \nu Z$, $\nu = \sigma/\sqrt{n}$, with independent standard Gaussian entries. Fitting the rotation or using I gives error $O(\min\{\rho, \nu\})$; the unknown edge occurs at most once without return and at most H times with it. For lower bounds take $\rho > 0$, compare 0 and $\delta = \min\{\rho, \nu/(2\sqrt{2}), 1/k\}$, and set $k = 1$ or $\lfloor H/2 \rfloor$, respectively. Calibration KL is at most $1/8$. For known return d, queries b or $(bd)^k$ separate task normals by $k\delta$. Their normalized difference gives one common fact with scores $\pm\sin(k\delta/2)$ and two-point risk at least $3\sin(k\delta/2)/8$. Models, query and fact precede noise; $\rho = 0$ gives zero risk (Appendix C.2). The query span changes from $\mathrm{span}\{1, \cos\theta, \sin\theta\}$ to all harmonics, explaining exact-table complexity but not the rates. A fixed joint return $R_\theta R_{\omega-\theta} = R_\omega$ need not amplify edge uncertainty; noiseless identification need not permit noisy all-path learning (Appendix B.1).

4.4 HOW DEPLOYMENT CHANGES THE CRITERION

The all-path guarantee is stronger than performance under any fixed future distribution. Let ν be a fixed fact law independent of the model, observations and path, and define

$$\mathcal{R}^{\nu}_{n} = \inf_{\mathsf{L}} \sup_{T,p} \mathbb{E}_Y \int r(\mathsf{L}(Y, e, p), f_T(p, e))\, d\nu(e). \tag{11}$$

Theorem 4.4 (Which averaging removes the obstruction?)**.** *If ν has full support on the sphere, $\mathcal{R}^{\nu}_{n} \to 0$ if and only if the finite-type condition of Theorem 4.2 holds; failure gives a positive floor. For uniform sphere facts σ_r, $\kappa_r = \Gamma(r/2)/(\sqrt{\pi}\Gamma((r+1)/2))$ gives*

$$(\kappa_r/4)\mathcal{R}^2_n \le \mathcal{R}^{\sigma_r}_n \le \mathcal{R}_n.$$

In contrast, for every compact class $\mathcal{K}$ on the fixed finite graph, one learner sequence L_n, independent of q, satisfies $\sup_T \sum_p q(p) \sup_{\|e\|=1} \mathbb{E}_Y r(\mathsf{L}_n(Y, e, p), f_T(p, e)) \to 0$ for every fixed probability q on finite root paths. This holds even if the finite-type condition fails. Convergence is uniform over families of laws with uniformly vanishing length tails, not over all path laws.

Fact averaging retains a positive measure of conflicts; fixed-path-law learning controls bounded lengths, then their probability tail. Positive-definite fact covariance alone is insufficient: a discrete isotropic law can miss every conflict (Appendix D).

Changing deployment. More generally, let $\Psi : \mathcal{K} \to \mathbb{R}^D$ be a continuous mean map of a fixed finite calibration experiment, with $\overline{Y} = \Psi(T) + (\sigma/\sqrt{n})Z$, $Z \sim N(0, I_D)$. The bank (4) and matched probes are distinct instances. For $u_T(p) = T_p^\top c$ and path-law family $\mathcal{Q}$, define

$$\Omega_{\mathcal{Q}}(t) = \sup_{\substack{T,S\in\mathcal{K} \\ \|\Psi(T)-\Psi(S)\|\le t}} \sup_{q\in\mathcal{Q}} \mathbb{E}_{p\sim q}\|u_T(p) - u_S(p)\|^2. \tag{12}$$

This measures calibration ambiguity, not fitted residuals. Fix the graph, $r, D, \sigma > 0$, compact nonempty $\mathcal{K}$ and nonempty deterministic law families $\mathcal{Q}_n$ independent of training noise. Define

$$\mathcal{R}_n(\mathcal{Q}_n) = \inf_{\mathsf{L}} \sup_{T\in\mathcal{K},\, q\in\mathcal{Q}_n} \mathbb{E}_{\overline{Y},p,e}\, r(\mathsf{L}(\overline{Y}, e, p), u_T(p)^\top e),$$

where $p \sim q$ and e is an independent uniform spherical fact. Theorem E.1 gives

$$\mathcal{R}_n(\mathcal{Q}_n) \to 0 \quad \Longleftrightarrow \quad \Omega_{\mathcal{Q}_n}(a\sigma/\sqrt{n}) \to 0 \text{ for every fixed } a > 0. \tag{13}$$

The learner sees $(\overline{Y}, p, e)$, not q. Finite-net fitting and Gaussian overlap of calibration-near pairs give the upper and lower bounds (Appendix E). Neither exact identification nor commutativity is required; disconnected compact classes require every fixed multiplier.

Order cannot generally be reduced to counts. On one vertex let a have unknown R_θ, $0 \le \theta \le \rho \le 0.1$, and s have known $J = \mathrm{diag}(1, -1)$. Supply a quarter-turn and the same four scalar rotation probes as in Proposition 4.3. Since $JR_\theta J = R_{-\theta}$, equal-length words with identical endpoints and action counts satisfy

$$T_{(as)^{2k}} = I, \qquad T_{a^{2k}s^{2k}} = R_{2k\theta} \quad (k \ge 1).$$

Under a uniform planar fact, these words have risks zero and $\asymp \min\{1, k^2 s_n^2\}$, respectively (Proposition E.3). Here $s_n = \min\{\rho, \sigma/\sqrt{n}\}$. In the commuting matched model, risk has order $\mathbb{E}_q \min\{1, s_n^2 N_b(p)^2\}$, where N_b counts uncertain crossings. Changing laws allow consistency exactly when $s_n N_b \to 0$ in probability; the bounds and estimator are uniform over laws (Theorem D.4).

Why a finite nominal graph matters. Over a known badly approximable circle rotation, Proposition F.1 constructs a compact smooth, uniformly task-observable class gauge equivalent to one known $SO(3)$ response. Every finite Gaussian-noisy bounded-length calibration leaves a positive worst-model, worst-path regret floor, independent of query count, length bound and positive noise level, even with adaptive queries and supplied starting points. Locally cancelling coordinates retain long-product conflicts: known type does not ensure stable calibration. Appendix F gives a sharp Sobolev threshold and transfer rates over this rotation (Avila & Kocsard, 2011; Kocsard, 2013; Karaliolios, 2017), not arbitrary visual models.

5 Learning response relations

Controlled studies isolate what short-path fit leaves unresolved and how legal use exposes it. Supplied-dictionary calibration tests the distinction constructively; risk prediction and action restriction examine its deployment consequences.

Supplied-dictionary calibration. Here the graph has one vertex, so every action word is legally queryable. For $T_a = QM_a^j Q^\top$ with unknown j, Q and J supplied orthogonal tuples, set $M_a = \mathrm{diag}(M_a^1, \dots, M_a^J)$. Its word span has an actual-word basis of dimension $b \le Jr^2$, with lengths at most $b - 1$. Known coefficients give

$$M_w = \sum_i a_i(w) M_{w_i}, \qquad c^\top T_w e = \sum_i a_i(w) c^\top T_{w_i} e. \tag{14}$$

Known coefficient updates avoid noisy generator products, without type or frame recovery, even for reducible tuples. This supplied-class construction, not class discovery, specializes the general finite-graph theorem. Appendix C.3 bounds noise independently of horizon while charging conditioning, preparation, nrb observations and probe lengths; misspecification can add horizon-dependent bias.

Calibration structure and long-query decisions. Four estimators share a two-type dictionary and noisy scalar probes on 64 systems with unknown $SO(3)$ frames (Figure 3). Joint fitting estimates type and frame; the direct table avoids generator recovery. Order-only/free fits relax joint constraints, start from the joint fit and retain no greater training loss, yet alternating-query regret at $H = 2048$ rises from 4.3×10^{-7} to $.065/.074$ with the correct dictionary. Training fit alone therefore does not ensure reliable reuse. At .01 radians of misspecification, joint/free regret is instead .268/.074: incorrect supplied relations can reverse the benefit (Appendix H.4).

Legal use and action order. Under matched calibration, return slopes exceed one-way slopes in all 95 settings with positive risk at every horizon (medians $1.051/0.083$). These finite-query means illustrate Proposition 4.3, not minimax rates (Figure 4; Appendix H.3). On four vertices, grouping the same fitted matrices raises regret in all six relation/design cells despite matched length, endpoint, edge and mixed-block counts. With exact relations and return probes at $H = 3074$, the difference is .001993 [.001357,.002705] (Appendix H.4). A separate matched-budget study leaves targeted probe superiority over D-optimal unresolved (Table 7). Composition tests with position-encoding and optimizer controls support extrapolation within the structured generator class (Appendix H.6).

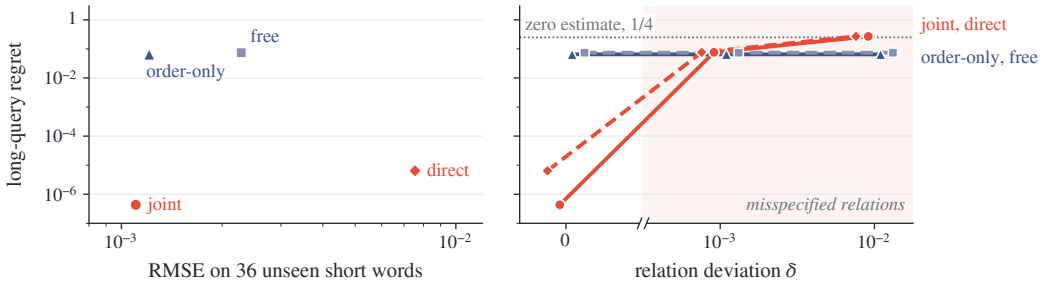


Figure 3: **Same scalar supervision, different reuse.** Alternating-query regret at $H = 2048$, $n = 64$ repeats, against short-word RMSE at $\delta = 0$ (left) and as the true relation deviates by δ from the dictionary (right; series dodged). Pointwise paired-system bootstrap intervals, where reported, are mostly hidden by the markers.

From calibration uncertainty to query risk. For three-dimensional unit rows $u_w = T_w^\top c$, $\widehat{u}_w$ and uniform spherical facts, exactly $\mathbb{E}_e r(\mathrm{sign}(\widehat{u}_w^\top e), u_w^\top e) = \|u_w - \widehat{u}_w\|^2/8$. Correct local type and negligible bias give the unscaled predictor $\mathrm{tr}(J_w \Sigma J_w^\top)/8$, with query-row Jacobian J_w and calibration parameter covariance Σ. Twelve held-out systems give median error factors 1.18 versus 3.04 for length. Both predictors fit one scale on twelve development systems; risk averages 24 fresh calibration draws. The evidence supports local risk-magnitude prediction, with reduced accuracy at higher noise, rather than a minimax rate or a demonstrated ranking advantage (Appendix H.4, Table 6).

Restriction and capability. Restricting use also removes candidates. In an analytic wrist study at length 128, removing reversals gives within-library error .0128 radians, but regret against the full-library optimum .1733 versus .1092 without restriction or calibration. At noise .30, least-squares calibration with 64 length-six words gives full-library regret .0298; one word gives .5072. The shared optimum charges lost candidates as well as selection error (Appendix H.5, Tables 9–10).

6 Cached correction in visual world models

The structured studies test what calibration must preserve under reuse. Visual planning tests the cached-feedback interface: decision quality, feedback cost and transfer beyond training length. These experiments do not identify return types or verify the theorem's all-path assumptions.

Implementation and comparisons. The observable-endomorphic world model (OEWM) transports one innovation along cached branches to revise scores:

$$z_0 = O_\psi(o_1, \widehat{o}_1, h_0), \quad z_{k,t+1} = T_\psi(h^0_{k,t}, a_{k,t}, z_{k,t}), \quad \widehat{\Delta V}_{g,k} = C_\omega(g, h^0_{k,0:H}, z_{k,0:H}). \tag{15}$$

Simulator forks provide training targets and evaluation labels, never selector inputs. Pooling averages path features; GRU reads them recurrently, sharing OEWM's cache and innovation. Reroll starts from the factual frame; conditioned reroll adds the innovation. The timed gate is innovation-dependent, not conditionally compiled (Appendices G.3, H.1).

Matched feedback, quality and cost. DINO-WM (Zhou et al., 2025) on Meta-World (Yu et al., 2020) Assembly-v3 with backlash uses six frozen seeds, 64 fresh roots from two disjoint held-out scene pools, 125 candidates and six blocks (576 cells, no headroom filtering). Analysis and checkpoints precede collection; task, effects and depth match training. Bank-normalized absorbing-success return G credits success through the remaining horizon, without strict success-first ordering (Appendices H.7, H.9).

The matched comparison forms one innovation from a shared three-frame factual prefix. Conditioned reroll uses OEWM's innovation and reader architecture, with separately trained weights; pooling is capacity-matched. A same-size no-feedback control zeros residuals throughout. For each method, Table 1 measures quality and timing with the same fixed checkpoints.

Table 1: **Matched feedback, quality and cost.** Six-block Assembly, 64 fresh roots, six seeds. Pointwise crossed seed/root 95% intervals; complete-path p99 over 4,320 calls per row on an RTX 5090, excluding acquisition and prepaid cache construction. Only regeneration uses BF16.

Reader and execution	G regret [95% CI]	Success (%)	p99 (ms)
OEWM	.0890 [.0772,.1013]	93.46	5.02
Pooling, capacity-matched	.1328 [.1151,.1541]	90.10	4.83
GRU	.1090 [.0928,.1277]	91.90	4.88
No feedback	.3386 [.3195,.3568]	72.05	.13
Conditioned reroll, FP32	.0915 [.0770,.1072]	93.49	495.18
Conditioned reroll, BF16	.0909 [.0763,.1068]	93.52	184.93

On regret, OEWM minus pooling/GRU is $-.0438[-.0644, -.0260]/-.0199[-.0369, -.0065]$ (Holm $p = .0012/.0104$). Binary success improves over pooling by $3.36[1.24, 5.93]$ points and over no feedback by $21.41[18.34, 24.36]$ points; its GRU contrast is unresolved. By contrast, an earlier six-block cohort leaves the OEWM–GRU regret contrast unresolved (Table 26). Against conditioned FP32 reroll, the primary regret difference is $-.0025[-.0166, +.0111]$; without a prespecified margin, neither equivalence nor noninferiority is established. Feedback p99 is about 99/37 times lower than conditioned FP32/BF16 reroll. BF16 also meets 187.5 ms on every measured call, so the gain is lower feedback cost rather than exclusive deadline feasibility. Preparation averages 8.05 s per root, charged separately; observed timing is not a worst-case guarantee.

Transfer beyond training length. Training on two/three/four blocks transfers to twelve on 32 fresh Assembly scenes with return G. OEWM success is 77.69%, versus 24.31% with truncated-reward training, 52.08% for capacity-matched pooling and 52.34% without feedback. Exploratory 450-epoch training gives OEWM/GRU success of 81.34/78.99%, with neither regret nor success separation; epochs, not FLOPs, match. The evidence supports finite-length transfer with aligned supervision, without establishing gate necessity (Table 29).

Supplementary scope and controls. Sixteen-chunk correction meets a two-point success-noninferiority margin against reroll but does not separate from pooling (Figure 13); pooling is also competitive with JEPA-WM (Terver et al., 2025) (Appendix H.8). These results support cached correction, not uniform architectural superiority. Additional screening, combination-transfer and task-coverage studies use distinct objectives and selection protocols (Appendices H.7 and H.10). Their task-dependent gains and failures address different comparisons from the matched G-based test.

7 Discussion and limitations

Calibration sufficiency depends on the response differences that future use can expose, not on short-path fit alone. Our characterization separates approximate all-path learning from exact linear calibration, and unrestricted reuse from performance under a fixed path distribution. The action library is therefore part of the learning specification: adding legal returns can change the required information even when the calibration observations remain unchanged.

The all-path criterion assumes a known compact orthogonal response class, a fixed finite graph and the prescribed noisy calibration protocol. Norm preservation isolates the role of legal composition from norm growth; the continuous-base counterexample shows why the finite graph matters. Finite representability need not yield inexpensive calibration, and the constructive rule uses a supplied dictionary. Enforcing incorrect relations can reverse the benefit of structural constraints.

The visual studies support cached correction and finite-length transfer in the tested simulators, without establishing neural all-path guarantees or architectural necessity. Their latency advantage concerns feedback-time execution with a prepared cache; its construction cost is reported separately. Learning suitable response classes, and when their guarantees extend to neural dynamics, is open.

AI USE STATEMENT

We used large language model assistants and an agentic coding assistant to help develop the theoretical framework and hypotheses, formulate claims, develop and write proofs, design experiments, implement methods and simulations, clean and reformat experimental outputs, and interpret results. They also assisted with experimental parameters, literature searches and summaries, reference formatting, translation, manuscript organization, drafting and editing, and figure preparation. Synthetic data were generated by simulation programs, not directly by language models; qualitative and thematic data analysis was not applicable. The authors reviewed the AI-assisted work, checked the claims and proofs, inspected and reran the experimental code, compared reported numbers with stored outputs, and verified citations against their sources. The authors take responsibility for the final manuscript and all AI-assisted claims, results and artifacts.

ETHICS STATEMENT

This work studies feedback reuse through mathematical analysis and simulation. The empirical results do not establish safety guarantees for deployment on physical robots.

REPRODUCIBILITY STATEMENT

Appendices B–G provide the theoretical assumptions, proofs and constructions. Appendix H reports simulation protocols, model and training configurations, evaluation metrics, timing procedures and uncertainty estimates. The code will be made publicly available upon acceptance.

REFERENCES

Albert Ai, Alex Lapanowski, Yaniv Plan, and Roman Vershynin. One-bit compressed sensing with non-Gaussian measurements. *Linear Algebra and its Applications*, 441:222–239, 2014. doi: 10.1016/j.laa.2013.04.002.

Tuo An, Jindou Jia, Gen Li, Jingliang Li, Chuhao Zhou, Pengfei Liu, Bofan Lyu, Jiaqi Bai, Xinying Guo, Geng Li, and Jianfei Yang. Feedback world model enables precise guidance of diffusion policy. *arXiv preprint arXiv:2605.15705*, 2026.

Artur Avila and Alejandro Kocsard. Cohomological equations and invariant distributions for minimal circle diffeomorphisms. *Duke Mathematical Journal*, 158(3):501–536, 2011. doi: 10.1215/00127094-1345662.

Borja Balle, Prakash Panangaden, and Doina Precup. Singular value automata and approximate minimization. *Mathematical Structures in Computer Science*, 29(9):1444–1478, 2019. doi: 10.1017/S0960129519000094.

Steven L. Brunton, Bingni W. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PLOS ONE*, 11(2):e0150171, 2016. doi: 10.1371/journal.pone.0150171.

Marc Burger, Narutaka Ozawa, and Andreas Thom. On Ulam stability. *Israel Journal of Mathematics*, 193(1):109–129, 2013. doi: 10.1007/s11856-012-0050-z.

Michel Coste. *An Introduction to Semialgebraic Geometry*. Institut de Recherche Mathématique de Rennes, 2002.

Víctor Elvira, Luca Martino, and Christian P. Robert. Rethinking the effective sample size. *International Statistical Review*, 90(3):525–550, 2022. doi: 10.1111/insr.12500.

Tingran Gao, Jacek Brodzki, and Sayan Mukherjee. The geometry of synchronization problems and learning group actions. *Discrete & Computational Geometry*, 65(1):150–211, 2021. doi: 10.1007/s00454-019-00100-2.

Christopher Grimm, André Barreto, Gregory Farquhar, David Silver, and Satinder Singh. Proper value equivalence. In *Advances in Neural Information Processing Systems*, volume 34, pp. 7773–7786, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/400e5e6a7ce0c754f281525fae75a873-Abstract.html.

Nikolaos Karaliolios. Differentiable rigidity for quasiperiodic cocycles in compact Lie groups. *Journal of Modern Dynamics*, 11:125–142, 2017. doi: 10.3934/jmd.2017006.

Alejandro Kocsard. On cohomological C^0-(in)stability. *Bulletin of the Brazilian Mathematical Society, New Series*, 44(3):489–495, 2013. doi: 10.1007/s00574-013-0023-9.

Michael Larsen, Alexander Lubotzky, and Claude Marion. Deformation theory and finite simple quotients of triangle groups I. *Journal of the European Mathematical Society*, 16(7):1349–1375, 2014. doi: 10.4171/JEMS/463.

Yanhong A. Liu and Tim Teitelbaum. Caching intermediate results for program improvement. In *Proceedings of the 1995 ACM SIGPLAN Symposium on Partial Evaluation and Semantics-Based Program Manipulation*, pp. 190–201, 1995. doi: 10.1145/215465.215590.

Eric Martin and Chris Cundy. Parallelizing linear recurrent neural nets over sequence length. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=HyUNwulC-.

Mario Milanese and Antonio Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27(6):997–1009, 1991. doi: 10.1016/0005-1098(91)90134-N.

Erik Nielsen, John King Gamble, Kenneth Rudinger, Travis Scholten, Kevin Young, and Robin Blume-Kohout. Gate set tomography. *Quantum*, 5:557, 2021. doi: 10.22331/q-2021-10-05-557.

Mihály Petreczky, Laurent Bako, and Jan H. van Schuppen. Identifiability of discrete-time linear switched systems. In *Proceedings of the 13th ACM International Conference on Hybrid Systems: Computation and Control*, pp. 141–150, 2010. doi: 10.1145/1755952.1755973.

Mihály Petreczky, Laurent Bako, and Jan H. van Schuppen. Realization theory of discrete-time linear switched systems. *Automatica*, 49(11):3337–3344, 2013. doi: 10.1016/j.automatica.2013.07.022.

Yaniv Plan and Roman Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Transactions on Information Theory*, 59(1):482–494, 2013. doi: 10.1109/TIT.2012.2207945.

Yury Polyanskiy and Yihong Wu. Dualizing Le Cam's method for functional estimation I: General theory. *The Annals of Statistics*, 54(1):1–24, 2026. doi: 10.1214/25-AOS2498.

Xu Shang, Jorge Cortés, and Yang Zheng. Willems' fundamental lemma for nonlinear systems with Koopman linear embedding. *IEEE Control Systems Letters*, 8:3135–3140, 2024. doi: 10.1109/LCSYS.2024.3522594.

Amit Singer and Hau-Tieng Wu. Vector diffusion maps and the connection Laplacian. *Communications on Pure and Applied Mathematics*, 65(8):1067–1144, 2012. doi: 10.1002/cpa.21395.

Basile Terver, Tsung-Yen Yang, Jean Ponce, Adrien Bardes, and Yann LeCun. What drives success in physical planning with joint-embedding predictive world models? *arXiv preprint arXiv:2512.24497*, 2025.

Michael Thon and Herbert Jaeger. Links between multiplicity automata, observable operator models and predictive state representations — a unified learning framework. *Journal of Machine Learning Research*, 16(4):103–147, 2015. URL https://www.jmlr.org/papers/v16/thon15a.html.

Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033, 2012. doi: 10.1109/IROS.2012.6386109.

Henk J. van Waarde, Claudio De Persis, M. Kanat Camlibel, and Pietro Tesi. Willems' fundamental lemma for state-space systems and its extension to multiple datasets. *IEEE Control Systems Letters*, 4(3):602–607, 2020. doi: 10.1109/LCSYS.2020.2986991.

Andrew J. Wagenmaker, Max Simchowitz, and Kevin Jamieson. Task-optimal exploration in linear dynamical systems. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 10641–10652. PMLR, 2021. URL https://proceedings.mlr.press/v139/wagenmaker21a.html.

Tianhe Yu, Deirdre Quillen, Zhanpeng He, Ryan Julian, Karol Hausman, Chelsea Finn, and Sergey Levine. Meta-World: A benchmark and evaluation for multi-task and meta reinforcement learning. In *Proceedings of the Conference on Robot Learning*, volume 100 of *Proceedings of Machine Learning Research*, pp. 1094–1100. PMLR, 2020. URL https://proceedings.mlr.press/v100/yu20a.html.

Amy Zhang, Rowan McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. Learning invariant representations for reinforcement learning without reconstruction. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=-2FCwDKRREu.

Chushan Zhang, Jinguang Tong, Xuesong Li, Yikai Wang, and Hongdong Li. Keep the future, drop the rollout: RIFT for world action models. *arXiv preprint arXiv:2608.11521*, 2026.

Gaoyue Zhou, Hengkai Pan, Yann LeCun, and Lerrel Pinto. DINO-WM: World models on pre-trained visual features enable zero-shot planning. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 79115–79135. PMLR, 2025. URL https://proceedings.mlr.press/v267/zhou25t.html.

A TECHNICAL SOURCES AND SCOPE

The finite-observability argument in Appendix B applies linear realization theory only to legal paths (Petreczky et al., 2013). The return-rigidity proof derives its operator-norm constant directly by representation averaging (Burger et al., 2013, Proposition 4.4(1)). The semialgebraic implication and calibration rate use quantifier elimination, finite cell decomposition and the standard inequality in Coste (2002, Theorems 2.6, 2.12 and Section 2.3).

Koopman and Willems-type representations have fixed-system, embedding or excitation conditions (Brunton et al., 2016; van Waarde et al., 2020; Shang et al., 2024) and are not invoked to infer class-wide calibration; the exact table here has model-independent coefficients across the known candidate class and all horizons.

Appendix D derives the spherical sign moment needed for its decision loss, rather than importing a sparse-recovery guarantee (Plan & Vershynin, 2013); discrete isotropic sign ambiguity has precedents (Ai et al., 2014, Section 1). Appendix E uses two-point testing, not the affine-functional, convex-class duality of Polyanskiy & Wu (2026).

B PROOF OF THE CALIBRATION PRINCIPLE

Appendix B proves the finite-graph characterization. Appendices C–E give the constructions, matched rates and deployment laws, Appendix F the continuous-base boundary, and Appendix G the interface facts. Three steps recur and are given in full at: the two-point decision bound, Appendix B.1; the Gaussian-overlap bound, proof of Theorem D.2; and the spherical sign-regret identity (49). Unless stated otherwise, lower bounds fix their two models, query or path law, and fact law before the calibration noise is drawn; upper bounds use unrestricted, possibly inefficient estimators; and no result certifies that a trained neural model meets its assumptions. The graph, model class, observation bank and task are those of Section 4. Facts enter only at o; paths may leave their component, and the empty path has transport I. Matrix-field and tuple norms are stacked Frobenius norms unless marked. Write m for the number of components and $L_0 = rv - 1$.

B.1 FINITE TASK ROWS AND THE ORIGINAL OBSERVATION BANK

For a vertex x, let W_x^k be the span of $c^\top T_p$ over legal paths from x of length at most k. Then

$$W_x^{k+1} = \operatorname{span}\big(c^\top, \{W_y^k T_a : a : x \to y\}\big). \tag{16}$$

The dimension sum starts at v and is at most rv; once it fails to increase, every space stays unchanged. So all row spaces stabilize by L_0, whether or not they fill $\mathbb{R}^{1\times r}$, as in linear realization theory restricted to legal paths (Petreczky et al., 2013).

Choose root paths $p_x : o \to x$ of length at most $v-1$, with p_o empty, and write

$$\begin{aligned} &P_x = T_{p_x}, \qquad H_a = P_y^\top T_a P_x, \qquad d_x^\top = c^\top P_x, \\ &\mathcal{O}_y = (c^\top T_{p_y q})_{q\in\mathcal{P}_y^{\le L_0}}, \qquad \mathcal{B}_a = (c^\top T_{p_x a q})_{q\in\mathcal{P}_y^{\le L_0}}. \end{aligned} \tag{17}$$

All their scalar entries lie in the original bank (word lengths are below $L = (r+1)v$), and

$$\mathcal{O}_y H_a = \mathcal{B}_a, \qquad c^\top T_p e = d_y^\top H_p e \quad (p : o \to y). \tag{18}$$

The unobserved internal frames P_x cancel, and $P_o = I$ preserves factual coordinates.

Let $\Psi(T)$ denote all D training means, and put $d_\infty(T,S) = \sup_{p:o\to *}\|c^\top(T_p - S_p)\|$. If $\Psi(T) = \Psi(S)$, the blocks $\mathcal{O}_y, \mathcal{B}_a$ and readers d_x coincide, so by (18) the two edge tuples agree on the common row span of $\mathcal{O}_y$. By (16) that span is invariant under backward continuation to the source row space, so propagating backward from the common endpoint reader along any query gives the same root row:

$$\Psi(T) = \Psi(S) \quad \Longrightarrow \quad d_\infty(T,S) = 0. \tag{19}$$

No inverse observation matrix or positive singular-value assumption is used here.

An identifiable class without uniform reuse. Take one vertex and one self-loop with planar response R_θ, $0 \le \theta \le \rho \le 0.1$, $\rho > 0$, and $c = e_1$. The compact class is semialgebraic in its matrix entries, and two noiseless one-step basis-fact responses $(\cos\theta, -\sin\theta)$ determine θ uniquely. Nevertheless, every finite noisy calibration has positive worst-case all-path risk for the task in (3).

Indeed, allow $N \ge 1$ adaptive scalar queries of lengths $k_i \le B_0$, $B_0 \ge 1$, with unit facts and independent $N(0, \sigma^2)$ noise. Compare $\theta = 0$ and $\delta = \min\{\rho, \sigma/(2B_0\sqrt{N}), 1/2\} > 0$. On a common observation history the learner uses the same query rule, and the mean difference is at most $k_i\delta$; the Gaussian chain rule, also conditioning on any algorithm randomness, gives $\mathrm{KL}(P_0\|P_\delta) \le NB_0^2\delta^2/(2\sigma^2) \le 1/8$, so $\mathrm{TV}(P_0, P_\delta) \le 1/4$ by Pinsker. Set $H = \lfloor 1/\delta \rfloor$; then $1/2 \le H\delta \le 1$. For $u = c$, $v = (R_\delta^H)^\top c$, the common unit fact $e_* = (u-v)/\|u-v\|$ yields scores $u^\top e_* = \sin(H\delta/2) = -v^\top e_*$. Both H and e_* are fixed before the calibration noise is drawn. On this fixed query, the two decision-error probabilities of any randomized learner sum to at least $3/4$, and a wrong sign costs the score magnitude, so at least one model has expected regret at least $3\sin(1/4)/8$, independently of N and B_0. For each fixed finite horizon, by contrast, consistent one-step angle estimation gives vanishing regret: the bound concerns uniformity over all horizons.

B.2 Returns, invariant projections, and one component expansion

The marked paths u_x, v_x of Section 4 ($z = z_C$) give marked returns $u_x v_x$ and $u_x a v_y$, for internal edges $a : x \to y$, with products R_x, R_a and

$$T_{u_y}^\top T_a T_{u_x} = R_y^{-1} R_a. \tag{20}$$

Thus the marked products generate the same compact group as all based legal returns: the closure of a semigroup in a compact orthogonal group contains inverses, since positive powers approach the identity along a subsequence and the preceding powers approach the inverse. Inverses in (20) classify representations; they are not new executable edges.

Rigidity from legal returns. We prove Lemma 4.1, used from here on.

Proof. The paired legal-return products form a semigroup in the compact orthogonal group $O(r) \times O(r)$, since concatenation is legal; by the argument after (20), their closure is a compact subgroup.

Let $\gamma < 1$ bound the displayed discrepancies. The closed convex hull $\mathcal{X}$ of the matrices $S_g T_g^\top$ lies within operator distance γ of I, so every matrix in it is invertible. Left–right multiplication $X \mapsto S_g X T_g^\top$ preserves $\mathcal{X}$ and its Frobenius norm, so the unique minimum-norm element is fixed by this action: $S_g X = X T_g$. Then $X^\top X$ commutes with every T_g, and the orthogonal polar factor Q_z of X also intertwines every return (real operator-norm averaging, cf. Burger et al., 2013, Proposition 4.4(1)); the needed constant is established directly here.

For each vertex in the component, choose legal internal paths $u_x : z \to x$ and $v_x : x \to z$, and put $Q_x = S_{u_x} Q_z T_{u_x}^\top$. The return $u_x v_x$ gives $S_{v_x} Q_x = Q_z T_{v_x}$. For an edge $a : x \to y$, the return $u_x a v_y$ gives $S_{v_y} S_a Q_x = Q_z T_{v_y} T_a = S_{v_y} Q_y T_a$, and cancelling the invertible S_{v_y} yields $S_a Q_x = Q_y T_a$. The required v_y need not exist across distinct components, hence the componentwise statement. Conversely, an edge gauge telescopes along every legal return. □

Invariant subspaces of an orthogonal tuple have invariant orthogonal complements, so the splitting into types exists; each isotypic projector P_π commutes with every return, and there are at most r distinct types per component in one model. For any sum P_z of such projectors, transport it to x as $\Pi_x = T_{u_x} P_z T_{u_x}^\top$. Equation (20) gives

$$T_a \Pi_x = \Pi_y T_a \quad (a : x \to y \text{ internal}). \tag{21}$$

Changing roots or marking paths conjugates the representation and bijectively relabels the types. Prepending a legal path in either direction preserves the projected norm, proving invariance of the weights and finiteness conditions in (6).

The shared expansion. Fix finite sets of retained types, one per component, retain all their copies, and suppose every discarded type has weight at most τ. Orthogonality gives $\|(I - P_z)T_s^\top c\| \le \sqrt{r}\tau$ for every suffix, and (21) transfers this bound to any entry vertex. Insert the retained projection on

entering each component, moving it through its internal edges, and telescope the insertions in path order: the prefix has norm at most one and the remaining untruncated suffix has discarded projection at most $\sqrt{r}\tau$. Therefore the truncated response rows satisfy

$$\sup_{T,p}\|c^\top T_p - \widetilde{F}_T(p)\| \le m\sqrt{r}\tau. \tag{22}$$

At $\tau = 0$, retaining all positively visible types is exact.

There are finitely many retained type/multiplicity patterns. For a fixed pattern choose canonical internal transports M_a on dimensions $d_C \le r$. Equivalence of the marked returns gives isometries $J_x : \mathbb{R}^{d_C} \to \mathbb{R}^r$ with $T_a J_x = J_y M_a$ internally: set $J_x = T_{u_x} J_z M_{u_x}^\top$ for a root intertwiner J_z and apply (20) to the two connections. The reduced system has input $J_o^\top$, reader $c^\top J_y$, and crossing matrices $J_y^\top T_a J_x$, all of norm at most one.

A legal path follows one of finitely many crossing skeletons in the condensation DAG, with at most $m-1$ crossings. Expanding its reduced matrix product uses finitely many monomials of degree at most $m+1$ in these input, reader and crossing entries; only their known coefficients depend on internal paths. Choose J_x pointwise within each pattern, extend its monomials by zero outside the pattern, and take the finite union. Thus all truncated query-coordinate functions lie in a single finite-dimensional subspace of $\ell_\infty(\mathcal{K})$, with norms bounded by one. The auxiliary coordinates need not be continuous; the original query functions are.

B.3 Approximate calibration and the statistical equivalence

Suppose $\mathcal{F}_C(\tau)$ is finite for every $C, \tau > 0$. Use these global sets in the expansion, retaining a listed type even when its weight is smaller in a particular model. By (22), the query-row functions $G_p : \mathcal{K} \to \mathbb{R}^{1\times r}$, $G_p(T) = c^\top T_p$, are approximated arbitrarily well by bounded subsets of finite-dimensional spaces, so they are totally bounded in $C(\mathcal{K}; \mathbb{R}^r)$. A finite net can be chosen from the family itself, yielding actual paths p_i and (7). A selector chooses the first suitable prototype according to the known class; it does not use the realized model or noise.

Finite prototype nets also imply equicontinuity of $\{G_p\}$: approximate by finitely many continuous functions and use their common continuity radius. Hence $T_j \to T$ in edge coordinates implies $d_\infty(T_j, T) \to 0$. Compactness and (19) now give, for any $0 < \epsilon \le 1$,

$$\delta_\epsilon = \min\left\{1, \inf_{d_\infty(T,S)\ge\epsilon} \|\Psi(T) - \Psi(S)\|_2\right\} > 0, \tag{23}$$

with empty infimum $+\infty$. A finite $\delta_\epsilon/4$-net of $\Psi(\mathcal{K})$ stores one model per point, and choosing the nearest point to the averaged observations has score error below ϵ whenever the joint noise norm is at most $\delta_\epsilon/4$. In particular,

$$n \ge \frac{32\sigma^2 D\log(2D/\beta)}{\delta_\epsilon^2} \quad\Longrightarrow\quad \Pr_T\{\sup_{e,p}\mathrm{Regret} < \epsilon\} \ge 1-\beta. \tag{24}$$

Fixed finite tie-breaking makes the estimator measurable. Since regret is at most one, this proves consistency using precisely the original nD observations.

An unretained significant type forces a task conflict. Suppose $\mathcal{F}_C(\epsilon)$ is infinite. Choose distinct types π_j in models T_j with weight at least ϵ, and pass to $T_j \to S \in \mathcal{K}$. Discard the finitely many types present in S. For a remaining $T = T_j$, choose a suffix $s : z \to *$ with $\|P_\pi(T)T_s^\top c\| \ge \epsilon/2$. Write $d = T_s^\top c$, $w = S_s^\top c$, $P = P_\pi(T)$, and take Haar probability over the compact closure of paired legal-return products (A, B). The matrix

$$X = \int A^\top dw^\top B\, d\mu(A,B) \tag{25}$$

intertwines the S and T representations by right invariance. Its range has no nonzero π-part, because that type is absent from S (orthogonality of inequivalent components on the same marked words); hence $PX = 0$. Expanding gives

$$\int \|(d^\top A - w^\top B)P\|^2\, d\mu = \|Pd\|^2 + \int \|w^\top BP\|^2\, d\mu \ge \epsilon^2/4. \tag{26}$$

A finite legal return therefore gives row distance at least $\epsilon/4$. Prepending a fixed root access path of length at most $v-1$ loses at most $(v-1)d_{\mathrm{loc}}(T,S)$, where $d_{\mathrm{loc}}^2 = \sum_a \|T_a - S_a\|_F^2$, so for large j the root row distance is at least $\epsilon/8$.

The two task normals u, z are unit vectors. The common fact $(u-z)/\|u-z\|$ gives scores $\pm\|u-z\|/2$, hence margin at least $\epsilon/16$; this query and fact are chosen from the two models before the noise is drawn. Training distributions satisfy

$$\mathrm{KL}(P_T^n \| P_S^n) \le \frac{nDL^2}{2\sigma^2} d_{\mathrm{loc}}(T,S)^2. \tag{27}$$

Choose j so that this is at most $1/8$. The two-point step of Appendix B.1, with margin $\epsilon/16$, gives

$$\mathcal{R}_n \ge 3\epsilon/128 \qquad \text{for every finite } n. \tag{28}$$

With N adaptive noisy unit-fact queries of length at most B_0 instead, the chain rule of Appendix B.1 bounds the conditional KL by $NB_0^2 d_{\mathrm{loc}}^2/(2\sigma^2)$, and the same obstruction holds.

This proves the statistical converse. It also excludes prototype nets: they imply equicontinuity as above, contradicting the positive row separation along $T_j \to S$.

Prototype noise, without an inverse fit. Observe the r coordinates of each prototype row with n independent repeats and predict $Y_{i(p)}e$. For a fixed query, $i(p)$ is independent of the noise, the projected noise has variance σ^2/n, and the absolute bias is at most ϵ. This proves (9). Simultaneously, a union bound over the Nr averaged coordinates gives

$$\Pr_T\left\{\sup_{e,p} \mathrm{Regret} \le \epsilon + \sigma\sqrt{2r\log(2Nr/\beta)/n}\right\} \ge 1-\beta. \tag{29}$$

The expected bound has a supremum outside the expectation; the simultaneous bound does not. Compactness and selection of a finite net are standard steps; the response expansion is what makes them applicable uniformly over this model class.

B.4 Exact calibration and finite-description priors

If every $\mathcal{F}_C^+$ is finite, use the shared expansion with zero truncation error; it gives $\dim \mathcal{V}_{\mathcal{K}} < \infty$. Conversely, suppose $B := \dim \mathcal{V}_{\mathcal{K}} < \infty$ and choose a basis μ_i from actual scalar probes. Their evaluation vectors span $\mathbb{R}^B$, so finitely many models $S_1, \dots, S_b$, $b \le B$, give coefficients independent of path and fact such that

$$c^\top T_p = \sum_i \lambda_i(T) c^\top (S_i)_p \quad \text{for every root path } p. \tag{30}$$

If a visible type π in T were absent from all S_i, choose a visible suffix s and root access u. Applying (30) to $u\ell s$ and multiplying by $T_u^\top$ gives

$$d^\top A = \sum_i \lambda_i w_i^\top B_i C_i, \quad d = T_s^\top c, \quad w_i = (S_i)_s^\top c, \quad C_i = (S_i)_u T_u^\top.$$

Average over the compact closure of joint legal returns $(A, B_1, \dots, B_b)$. Equation (25) produces X_i with $PX_i = 0$. Multiply the row identity by P and take its Haar inner product with $d^\top AP$: every right term is $\lambda_i \operatorname{tr}(PX_iC_iP) = 0$, while the left is $\|Pd\|^2 > 0$. Thus the finitely many S_i contain every visible type, proving finiteness of $\mathcal{F}_C^+$.

Norm equivalence in the finite-dimensional span gives $\lambda_* := \min_{\|z\|_2=1} \|\sum_i z_i \mu_i\|_{C(\mathcal{K})} > 0$. Since every unit-fact query has supremum norm at most one, its unique coefficients have norm at most $1/\lambda_*$. This proves (8). Any homogeneous exact linear table must span $\mathcal{V}_{\mathcal{K}}$, so it needs at least B independent functions; known constants require no sampling, and this is not a lower bound on nonlinear learners.

Why semialgebraicity closes the gap. Approximate calibration makes $\mathcal{F}_C^+ = \bigcup_{k\ge1} \mathcal{F}_C(1/k)$ countable. For constituent dimension $d \le r$, a visible irreducible marked tuple has a witness

$$T \in \mathcal{K}, \quad U^\top U = I_d, \quad R_i(T)U = UM_i, \quad M_i \in O(d), \quad \sum_{s \in \mathcal{P}_z^{\le L_0}} \|U^\top T_s^\top c\|^2 > 0.$$

Nonzero visibility is detected by these finite suffixes by (16); no lower singular-value bound is assumed. Irreducibility says that no symmetric idempotent with trace strictly between zero and d commutes with every M_i. These conditions are first-order polynomial, so quantifier elimination gives a semialgebraic projected tuple set. Each compact simultaneous orthogonal orbit has a lexicographically least vector of entries, and the condition that this vector is no larger than every orthogonal conjugate is also first-order; hence the representative set is semialgebraic. A countable semialgebraic set is finite by finite cell decomposition (Coste, 2002, Theorem 2.6 and Section 2.3). The finite union over d proves the implication; its reverse follows from $\mathcal{F}_C(\epsilon) \subseteq \mathcal{F}_C^+$.

The two statistical interfaces. For a fixed exact table, $\widehat{f} = a(p,e)^\top Y$ has score-noise variance at most $\Gamma^2\sigma^2/n$. Consequently

$$\sup_{T,p,e} \mathbb{E}_T \operatorname{Regret} \le \min\{1, \Gamma\sigma\sqrt{2/(\pi n)}\}, \tag{31}$$

$$\Pr_T\{\sup_{p,e} \operatorname{Regret} \le \Gamma\sigma\sqrt{2B\log(2B/\beta)/n}\} \ge 1-\beta. \tag{32}$$

The second line bounds the whole calibration-noise vector, not a union over paths. The cost is nB scalars and $n\sum_i |w_i|$ independent-query action steps, excluding known coordinates.

To keep the original bank instead, note that for compact semialgebraic learnable $\mathcal{K}$ both a finite table μ and Ψ are polynomial maps. On $\mathcal{K}\times\mathcal{K}$, (19) gives $\|\Psi(T)-\Psi(S)\| = 0 \Rightarrow \|\mu(T)-\mu(S)\| = 0$, and the semialgebraic Łojasiewicz inequality (Coste, 2002, Theorem 2.12) yields an integer $N_0 \ge 1$ and $C < \infty$ such that

$$d_\infty(T,S) \le \Gamma\|\mu(T)-\mu(S)\| \le C\|\Psi(T)-\Psi(S)\|^{1/N_0}. \tag{33}$$

With probability at least $1-\beta$, the norm of all averaged scalar noises is at most $t = \sigma\sqrt{2D\log(2D/\beta)/n}$. On this event a finite t-net fit to the original means has mean-vector error at most $3t$, so its all-path regret is at most $C(3t)^{1/N_0}$.

B.5 Recovery of the quantitative observable case

For a learnable class satisfying Theorem 4.2, assume in addition that the bounded-length task rows from every vertex span $\mathbb{R}^r$. Using the prescribed observations (4), a nearest-point estimator then satisfies, for $0 < \beta < 1$,

$$\Pr_T\left\{\sup_{e,p} \operatorname{Regret} \le C_{G,\mathcal{K}}\sigma\sqrt{D\log(2D/\beta)/n}\right\} \ge 1-\beta. \tag{34}$$

The bound holds beyond a class-dependent sample size. Its constant and threshold depend on observability, separation of component-type orbits and calibration gaps, not on the deployment horizon. The following argument specifies these quantities.

Assume every $W_x^{L_0}$ has dimension r. Let $O_x(T)$ stack the corresponding unframed rows and set $s_* := \min_{T,x}\sigma_{\min}(O_x(T)) > 0$ and $M := \max_x |\mathcal{P}_x^{\le L_0}|$; compactness gives the positive minimum. A unit vector in any nonzero isotypic subspace has scalar product at least $s_*/\sqrt{M}$ with one of these rows, so every occurring type has weight at least $s_*/\sqrt{M}$. Thus the main type condition is equivalent to a finite total set of constituents, hence to finitely many full component gauge types at fixed dimension, and the statistical necessity already proved applies.

For the rate, fix a representative T of one full component type. On vertex matrix fields define

$$(\mathcal{D}_T X)_a = X_y T_a - T_a X_x, \quad \kappa_T = \text{smallest positive singular value of } \mathcal{D}_T. \tag{35}$$

If $S_a = Q_y T_a Q_x^\top$ and internal tuple distance is δ_C, then $\|\mathcal{D}_T Q\| = \delta_C$. Projection onto $\ker \mathcal{D}_T$ gives a field X with $\|X-Q\| \le \delta_C/\kappa_T$; when this is below one, every X_x is invertible. Since $X_y = T_a X_x T_a^\top$, uniqueness of polar decomposition makes its orthogonal factors satisfy $V_y = T_a V_x T_a^\top$. Nearest-orthogonal projection gives $\|V_x - Q_x\|_F \le 2\|X_x - Q_x\|_F$. The gauge $U_x = Q_x V_x^\top$ obeys $\|U_x - I\|_{\mathrm{op}} \le 2\delta_C/\kappa_T$ and $S_p = U_y T_p U_x^\top$, hence $\|S_p - T_p\|_{\mathrm{op}} \le 4\delta_C/\kappa_T$, though Q_x need not be small. The gap is gauge invariant. If $\mathcal{D}_T = 0$, the orbit is a singleton and there is no calibration error.

Distinct finitely many compact component orbits have minimum positive separation μ (set $\mu = \infty$ when no distinct pair exists). Let κ_* be the minimum of one and their positive gaps. Telescoping across at most m components, with crossing edges free to vary continuously, for $d_{\mathrm{loc}} < \min(\mu, \kappa_*)$,

$$\sup_p \|S_p - T_p\|_{\mathrm{op}} \le C_0 d_{\mathrm{loc}}, \qquad C_0 = 4m/\kappa_* + m - 1. \tag{36}$$

On the probability-$(1-\beta)$ event of the preceding subsection, every block noise norm in (17) is at most t, since each path appears once in each block. For $t < s_*/2$, estimate $\widetilde{H}_a = \widetilde{\mathcal{O}}_y^{\dagger}\widetilde{\mathcal{B}}_a$ and the endpoint rows. Since $\mathcal{B}_a = \mathcal{O}_y H_a$ and H_a is orthogonal, $\|\widetilde{H}_a - H_a\|_F \le 4t/s_*$ and $\|\widetilde{d}_x - d_x\| \le t$, so the combined transition/reader error is at most $b_n = \sqrt{16|E|/s_*^2 + v}\, t$. Fit a fixed finite b_n-net of the compact root-frame image of $\mathcal{K}$; nearest-point error is at most $3b_n$. For sufficiently large n, it lies in the correct component orbits. Applying (36) to the transitions and adding reader error proves (34) with $C_{G,\mathcal{K}} = 3(1 + C_0)\sqrt{2(16|E|/s_*^2 + v)}$.

C Constructions, rates, and indispensable boundaries

C.1 Weak visibility, exact calibration, and delayed task relevance

On a one-vertex, one-action graph, let $c = e_1$ and

$$a_j = 2^{-j}, \quad \theta_j = 0.7 + 2^{-j}, \quad T_j = R_z(a_j)R_x(\theta_j)R_z(a_j)^{\top}\ (j \ge 4), \quad T_\infty = R_x(0.7).$$

The class is compact; its axis type is shared, while its plane-rotation types are distinct with weight $\sin a_j \to 0$. Resolving c along axis and plane gives

$$\|(T_j^k)^{\top}c - c\| = 2\sin a_j |\sin(k\theta_j/2)| \le 2\sin a_j \quad (k \ge 0). \tag{37}$$

Every finite j is task observable because $\sin a_j \cos a_j \sin\theta_j(1-\cos\theta_j) \ne 0$; the limit is not. At each precision only a finite prefix needs distinguishing, and the scalar $c^{\top}T_j e_1 = 1 - \sin^2 a_j(1 - \cos\theta_j)$ separates each isolated model from its compact complement, so noise repetitions suffice. Theorem 4.2 also gives finite approximate prototypes, but infinitely many positively visible types preclude a finite exact linear table; the class is not semialgebraic.

Same type is not task invariance: for $T = \mathrm{diag}(1,-1)$, $S = R_{\pi/4}TR_{\pi/4}^{\top}$, $c = e_1$ and $e = (1,-1)^{\top}/\sqrt{2}$, the scores are $\pm 1/\sqrt{2}$; a shared quarter-turn makes both tuples observable and preserves equivalence, so calibration matters without accumulation. Immediate invisibility is also insufficient: with $A = R_x(\theta)$, $B = R_z(\pi/2)$ and $c = e_1$, all $c^{\top}A^k$ agree, but $A^H B$ at fact e_2 has scores $-1, +1$ for $\theta = 0, \pi/H$. The all-suffix weight in (6) detects B.

An exponential cost boundary. Take a chain of $m+1$ vertices with two edges at stage i, of responses I_2 and $\theta_i I_2$, $\theta_i \in \{-1,+1\}$, and known identity self-loops at all vertices. The class is finite, semialgebraic and orthogonal, and paths can be arbitrarily long. With $c = e_1$, the paths selecting unknown edges exactly in S have score $\chi_S(\theta) = \prod_{i\in S}\theta_i$ at fact e_1. Uniform independent signs give $\mathbb{E}\chi_S\chi_R = \mathbf{1}[S = R]$, hence $\dim \mathcal{V}_{\mathcal{K}} = 2^m$ (the constant coordinate is known). This bounds homogeneous exact linear interfaces only, not computation or unrestricted queries. A nonlinear learner instead observes m length-m paths, each isolating one sign. Thresholding n Gaussian repeats per sign has joint error at most $m\exp[-n/(2\sigma^2)]$, so $n \ge 2\sigma^2\log(m/\epsilon)$ gives worst-query expected regret at most ϵ with mn scalar samples and $m^2 n$ action steps.

C.2 Proof of the matched-return rates

The four common scalar queries of Proposition 4.3 have mean $B_c R_\theta$, where $B_c = [c^{\top}; c^{\top}R_{\pi/2}]$ is orthogonal for every known planar unit vector c. Premultiplication by $B_c^{\top}$ preserves independent Gaussian entries and gives $Y = R_\theta + \nu Z$, $\nu = \sigma/\sqrt{n}$, with Z having independent standard Gaussian entries. Nearest-curve fitting has expected matrix error at most 4ν, and I has error at most $\sqrt{2}\rho$; choose the better rule. The uncertain edge appears at most once on one-way paths and at most H times with a return; product telescoping and regret no larger than score error give both upper bounds, capped by one.

For the lower bound compare $\theta = 0$ and $\theta = \delta = \min\{\rho, \nu/(2\sqrt{2}), 1/k\}$, with $k = 1$ without return and $k = \lfloor H/2 \rfloor$ with return. Their KL is at most $\delta^2/\nu^2 \le 1/8$. The queries b or $(bd)^k$ have task normals separated by angle $k\delta$, and their normalized difference is one common fact giving scores $\pm \sin(k\delta/2)$. The two-point step of Appendix B.1 yields risk at least $3\sin(k\delta/2)/8 \ge 3k\delta/(8\pi)$. This matches (10) uniformly for $H \ge 2$; $\rho = 0$ gives zero risk. The empty and quarter-turn suffixes establish observability, and known irrational self-loops already allow arbitrary length without return. Without return every query function of θ lies in $\mathrm{span}\{1, \cos\theta, \sin\theta\}$; with return and $\rho > 0$ the queries include all harmonics $\cos(k\theta), \sin(k\theta)$, linearly independent on $[0, \rho]$. This explains the exact-calibration boundary; the finite-horizon rates use the two-point argument, not rank.

For a concrete decision contrast in these matched graphs, take $c = (1, 0)^\top$, $e = (\cos(3\pi/8), \sin(3\pi/8))^\top$, $\theta_A = 0$ and $\theta_B = \pi/36$, with $\rho = 0.1$. The paths bq_y^{16} and $(bd)^8 b$ both have length 17 and end at y. Since $R_{\pi/2}^{16} = I$, their unit response directions are respectively $67.5°, 72.5°$ and $67.5°, 112.5°$. Under (3), with right the positive and left the negative terminal action, both models prefer right on the first path and differ on the second. At averaged scalar noise $\nu = 1/4$, the common four-query experiment has $D_{\mathrm{KL}}(P_A \| P_B) = 2[1 - \cos(\pi/36)]/\nu^2 \simeq 0.12177 < 1/8$. These four exact responses do not imply agreement on every one-way query or replace the minimax argument above.

Why the scope cannot be removed. A two-edge cycle $R_\theta, R_{\omega-\theta}$ has fixed return R_ω, one vertex-gauge type, and harmless calibration uncertainty even when that return is nonidentity. Conversely, on a continuous nominal circle let both actions advance by known irrational α, with response R_θ on b and a known quarter-turn on q. There is no nonempty exact nominal return, but b^H has response $R_{H\theta}$; the same four-query experiment and two-point argument give horizon-growing risk, and empty and q suffixes maintain observability. Unfolding the graph into a depth-H tree changes its size, probe count and constants. None of the finite-graph results covers arbitrary continuous bases, unknown graphs, general nonlinear or nonorthogonal transports, image-only observations or uncalibrated root facts.

C.3 A SCALAR SUPPLIED-DICTIONARY COMPILER

This construction uses a single-vertex graph, so every action word is a legal calibration query. It is a constructive special case, not a restriction on the finite-graph characterization in Theorem 4.2. Let $M^1, \dots, M^J$ be supplied tuples in $O(r)^{\mathcal{A}}$ and $T_a = QM_a^j Q^\top$ for unknown j, Q. Shared constituents and nonscalar commutants are allowed. The class is compact semialgebraic, a finite union of projections of $Q^\top Q = I$ and $T_a Q = QM_a^j$.

Set $\mathbb{M}_a = \mathrm{diag}(M_a^1, \dots, M_a^J)$ and let $\mathcal{W} = \mathrm{span}\{\mathbb{M}_w\}$, of dimension $b \le Jr^2$. Starting with the identity, repeatedly left-multiply current basis matrices by every generator and add independent products; once the span stops growing it contains every word. Thus an actual-word basis w_i exists with w_1 empty and $|w_i| \le b - 1$. Stack their vectorized diagonal blocks as columns of V. Then

$$a(w) = V^\dagger \mathrm{vec}_{\mathrm{diag}}(\mathbb{M}_w), \qquad \|a(w)\| \le \Gamma_0 := \frac{\sqrt{Jr}}{\sigma_{\min}(V)}. \tag{38}$$

Every block word is orthogonal, the matrix expansion holds in every block, and conjugating by any Q proves (14). Left multiplication induces known coordinate updates $a(wa) = L_a a(w)$.

Observe each scalar $c^\top T_{w_i} e_j$ with n Gaussian repeats, store averaged rows Y_i, and predict $\sum_i a_i(w) Y_i e$. For each fixed unit fact the score noise has variance at most $\Gamma_0^2 \sigma^2/n$, so (31) and (32) hold with $\Gamma = \Gamma_0$ and $B = rb$. Known rows need not be sampled.

Conservative dense exact-real preprocessing costs $O(|\mathcal{A}| J^3 r^6)$ arithmetic operations, not bit operations. Acquisition uses nrb scalars and $nr \sum_i |w_i|$ action steps before omitting known rows; storage is rb measured numbers plus $|\mathcal{A}| b^2$ coefficient-map numbers. Dense coefficient propagation costs $O(|w| b^2)$, and response application $O(br)$, or $O(b)$ after row–fact products are formed. Coefficients or complete rows may be prepared before feedback, with their computation and storage charged. The statistical statement assumes exact coefficient evaluation.

For a true orthogonal tuple S outside the dictionary with $\max_a \|S_a - T_a\|_{\mathrm{op}} \le \zeta$ for some realization T, telescoping probes and query gives

$$\left| \sum_i a_i(w) c^\top S_{w_i} e - c^\top S_w e \right| \le \left(H + \Gamma_0 \sqrt{\sum_i |w_i|^2} \right) \zeta, \quad |w| \le H. \tag{39}$$

Sampling noise adds the previous term.

C.4 Alternative spectral fitting and the experimental dictionary

The controlled calibration studies also use a vector-generator protocol. Suppose supplied tuples M^j are pairwise nonconjugate and have real commutants $\mathbb{R}I$. Averaged noisy generator matrices satisfy $\|Y - T\|_F \le t_n := \sigma\sqrt{2p\log(2p/\beta)/n}$ with probability $1-\beta$, where $p = |\mathcal{A}|r^2$. Define

$$\mathcal{L}_j(Y)X = (Y_a X - X M_a^j)_a, \quad \gamma = \min_j \sigma_{r^2-1}(\mathcal{L}_j(M^j)), \quad \tau = \min_{i \ne j} \sigma_{r^2}(\mathcal{L}_j(M^i)),$$

with singular values decreasing and $\tau = \infty$ for $J = 1$. Select the smallest final singular residual, reshape its unit right singular vector, rescale to Frobenius norm $\sqrt{r}$, polar-project (Singer & Wu, 2012) to $\widehat{Q}$, and predict through $\widehat{Q} M_w^j \widehat{Q}^\top$.

Proposition C.1 (Relation-preserving spectral fit)**.** *The two gaps above are positive, and for* $\epsilon = \|Y - T\|_F < \min\{\tau/2, \gamma/(2\sqrt{2r})\}$ *the fit selects the correct type and the operator-norm error obeys* (40)*.*

A scalar commutant implies irreducibility, so nonzero intertwiners between such tuples are invertible and their polar factors would give conjugacy. Thus cross-type maps are injective while the correct map has kernel $\mathbb{R}Q$, proving both gaps positive. Perturbing T to Y changes each map by at most $\epsilon := \|Y - T\|_F$, so under the stated condition on ϵ type selection is correct and the fitted singular vector obeys $\sin\angle(v, Q/\sqrt{r}) \le 2\epsilon/\gamma$; after choosing its immaterial sign and polar projection, $\|\widehat{Q} - Q\|_F \le 4\sqrt{2r}\epsilon/\gamma$. Therefore

$$\sup_w \|\widehat{T}_w - T_w\|_{\mathrm{op}} \le \frac{8\sqrt{2r}}{\gamma}\epsilon. \tag{40}$$

The same bound, capped by one, controls fixed-task regret. Dense fitting costs $O(J|\mathcal{A}|r^6)$ and all dictionary gap checks $O(J^2|\mathcal{A}|r^6)$ arithmetic operations; this vector access differs from the scalar-root bank. For misspecified true S at tuple distance ζ from a dictionary realization and observation error ν, the bound becomes $8\sqrt{2r}(\nu+\zeta)/\gamma + H\zeta$ under the same small-noise condition.

The two canonical three-dimensional tuples used in the experiments are (A, B_θ), where

$$A = R_{e_3}(\pi), \quad t_\theta = 2\cos(\theta/2)/\sqrt{3}, \quad B_\theta = R_{(\sqrt{1-t_\theta^2}, 0, t_\theta)}(2\pi/3), \quad \theta \in \{4\pi/7, 6\pi/7\}. \tag{41}$$

Distinct axis inner products preclude common conjugacy. The axis of B_θ is not preserved by A, excluding a common line and, by orthogonality, a plane. A commuting symmetric matrix is therefore scalar; a commuting skew 3-by-3 matrix has invariant kernel and must vanish. Quaternion multiplication verifies $A^2 = B_\theta^3 = (AB_\theta)^7 = I$. This supplied prior is consistent with classical triangle-group representation theory (Larsen et al., 2014); perturbing θ preserves the individual orders but can violate the mixed relation.

D Which deployment uniformity matters?

The graph, class $\mathcal{K}$, calibrated basis facts, bank (4) and loss (3) are unchanged, except that the matched-rate result uses the four-scalar experiment of Proposition 4.3. Only the deployment risk changes; no labels, generator queries or OEWM components are added. With $u_T(p) = T_p^\top c \in \mathbb{S}^{r-1}$ and $\Psi(T) \in \mathbb{R}^D$ the mean vector of the bank, the identification fact

$$\Psi(T) = \Psi(S) \quad \Longrightarrow \quad u_T(p) = u_S(p) \quad \text{for every legal root path } p \tag{42}$$

is (19), proved in Appendix B without full observability. Let ν be a fixed Borel probability on $\mathbb{S}^{r-1}$, independent of the model, data and path, and define

$$\mathcal{R}_n^\nu = \inf_L \sup_{T\in\mathcal{K},\ p:o\to *} \mathbb{E}_{Y\sim P_T^n} \int r\big(L(Y,e,p), u_T(p)^\top e\big)\, d\nu(e). \tag{43}$$

Expectations include learner randomness; the infimum runs over the measurable, unrestricted class of (5). For a probability q on the countable set of finite legal root paths, let

$$\mathcal{R}_n(q,\nu) = \inf_L \sup_{T\in\mathcal{K}} \mathbb{E}_Y \sum_p q(p) \int r\big(L(Y,e,p), u_T(p)^\top e\big)\, d\nu(e). \tag{44}$$

The law q is fixed as n varies. An upper bound uniform in all facts also covers path- or model-dependent fact laws.

D.1 Random facts preserve the structural obstruction

Let σ_r be uniform probability on the sphere and $\kappa_r = \int |e_1|\, d\sigma_r(e) = \Gamma(r/2)/(\sqrt{\pi}\,\Gamma((r+1)/2))$, so $\kappa_2 = 2/\pi$ and $\kappa_3 = 1/2$. The spherical first-moment calculation below is classical; compare Plan & Vershynin (2013) and Appendix A.

Lemma D.1 (A readout reduction without new observations)**.** *For every experiment with a unit true task row and loss* (3)*,*

$$\frac{\kappa_r}{4}\mathcal{R}_n^2 \;\le\; \mathcal{R}_n^{\sigma_r} \;\le\; \mathcal{R}_n. \tag{45}$$

For any $\nu \ge \lambda\sigma_r$ as measures, with $\lambda > 0$, the left side is multiplied by λ and $\mathcal{R}_n^\nu \le \mathcal{R}_n$. The statement does not require a proper or linear input learner.

Proof. Average internal randomization into a conditional mean decision $g_Y(e,p) \in [-1,1]$. Fix Y, p and put $h = \int e\, g_Y(e,p)\, d\sigma_r(e)$. For each unit vector a, $a^\top h \le \int |a^\top e| d\sigma_r(e) = \kappa_r$, so $\|h\| \le \kappa_r$. For the unit true row u, the conditional risk is $\overline{r}_Y(u) = \frac{1}{2}\int(|u^\top e| - g_Y(e,p)u^\top e)d\sigma_r(e) = (\kappa_r - u^\top h)/2$. Consequently, for $z = h/\kappa_r$, $\|u - z\|^2 = 1 + \|z\|^2 - 2u^\top z \le 2 - 2u^\top z = 4\overline{r}_Y(u)/\kappa_r$. Decode a query using $\operatorname{sign}(z^\top e)$, with a fixed tie rule. A wrong sign has regret $|u^\top e| \le |(u - z)^\top e| \le \|u - z\|$; a right sign has zero regret. Taking expectations and using Cauchy–Schwarz yields

$$\sup_e \mathbb{E}_Y r(\operatorname{sign}(z^\top e), u^\top e) \le \frac{2}{\sqrt{\kappa_r}}\sqrt{\mathbb{E}_Y \overline{r}_Y(u)}.$$

The transformation uses only Y, p and the original learner's function, not the true u, and its parameter integrals are measurable because the integrands are bounded and jointly measurable. The supremum over models and paths, then an arbitrarily near-optimal input learner, give $\mathcal{R}_n \le 2\sqrt{\mathcal{R}_n^{\sigma_r}/\kappa_r}$. The opposite inequality is integration below a supremum, and nonnegativity of the loss gives $\mathcal{R}_n^\nu \ge \lambda\mathcal{R}_n^{\sigma_r}$. □

Theorem D.2 (Full-support random facts)**.** *If every nonempty open subset of the sphere has positive ν-mass, then $\mathcal{R}_n^\nu \to 0$ if and only if every $\mathcal{F}_C(\epsilon)$ in* (6) *is finite for every $\epsilon > 0$. Failure gives a positive risk floor for every finite n. The floor can depend on ν and is not uniform over all full-support laws. No density lower bound is needed for this qualitative statement.*

Proof. Sufficiency follows from $\mathcal{R}_n^\nu \le \mathcal{R}_n$ and Theorem 4.2. For necessity, the proof of Theorem 4.2 supplies $\eta > 0$ and, for every n, two models and one common legal path with task rows u, v satisfying $\|u - v\| \ge \eta$ and training total variation at most $1/4$. We replace only its final choice of an adversarial fact. Define

$$A_\nu(u,v) = \frac{1}{2}\int \big(|u^\top e| + |v^\top e| - |(u+v)^\top e|\big)\, d\nu(e),$$

the integral of $\min_{b\in\{-1,1\}}\{r(b,u^\top e) + r(b,v^\top e)\}$. For $u \ne v$, the direction $(u-v)/\|u-v\|$ gives opposite, nonzero scores, and so does an open neighborhood of it; hence $A_\nu(u,v) > 0$. Bounded continuity of the integrand makes A_ν continuous on the product of spheres, and compactness gives $a_\nu(\eta) = \min_{\|u-v\|\ge\eta} A_\nu(u,v) > 0$. For the two training laws let M have the minimum of their

densities with respect to their sum; M has mass $1 - \mathrm{TV}$. Since losses are nonnegative, the sum of the two risks is at least

$$\int dM(Y) \int \{r(g_Y(e,p), u^\top e) + r(g_Y(e,p), v^\top e)\}\, d\nu(e) \geq \tfrac{3}{4} A_\nu(u,v).$$

Here $r(g,f) = (|f| - gf)/2$ extends the loss to mean decisions; its minimum over $g \in [-1,1]$ is attained at an endpoint. One model thus has risk at least $3a_\nu(\eta)/8$. The models and path are fixed before the observations, and the fact is drawn from the same predeclared ν in both models. □

The same argument applies to the separated unit rows of Proposition F.1, so its continuous-base floor survives every fixed full-support fact law; for uniform facts, Lemma D.1 gives $\mathcal{R}^{\sigma_3} \geq \kappa_3 a_0^2/4$ from its original floor a_0.

Why covariance isotropy is not enough. Take one loop in dimension three, $d = (1,1,1)/\sqrt{3}$, $h = (1,-1,0)/\sqrt{2}$, $c = \cos\beta\, d + \sin\beta\, h$ with $0 < \beta < \arctan(1/\sqrt{2})$, $T_\theta = R_d(\theta)$ for $0 \leq \theta \leq \rho < \pi$, $\rho > 0$, and ν uniform on $\{\pm e_1, \pm e_2, \pm e_3\}$. This law has zero mean and covariance $I/3$, yet every coordinate of $(T_\theta^k)^\top c$ is at least $(\cos\beta - \sqrt{2}\sin\beta)/\sqrt{3} > 0$, since a unit vector in $d^\perp$ has coordinates at most $\sqrt{2/3}$. So the rule returning $+1$ on e_i and -1 on $-e_i$ has zero risk without training, for every k, θ. For $0 < \theta < \rho$, however, the planar irreducible return types vary continuously, each with task weight $\sin\beta > 0$, so the type condition fails in this compact orthogonal class of the original graph setting, which is not a failure of that theorem. Discrete sign ambiguity has precedents (Ai et al., 2014, Section 1); the example identifies its implication for this reuse criterion.

D.2 A FIXED CONTINUATION DISTRIBUTION REMOVES THIS OBSTRUCTION

Theorem D.3 (One learner for every fixed continuation law)**.** *For every compact orthogonal class in Section 4, whether or not its return-type condition holds, there is a single deterministic sequence L_n^* using exactly the observations* (4) *such that, for every fixed probability q on finite legal root paths,*

$$\sup_T \sum_p q(p) \sup_{\|e\|=1} \mathbb{E}_Y r\big(L_n^*(Y,e,p), u_T(p)^\top e\big) \longrightarrow 0. \tag{46}$$

The learner is independent of q. Convergence is uniform over any family $\mathcal{Q}$ satisfying $\lim_{H\to\infty} \sup_{q\in\mathcal{Q}} q(|p| > H) = 0$. No uniformity over all continuation laws or computational-efficiency guarantee is asserted.

Proof. For each finite H set

$$\omega_H(t) = \sup\{\|u_T(p) - u_S(p)\| : T, S \in \mathcal{K},\ |p| \leq H,\ \|\Psi(T) - \Psi(S)\| \leq t\}.$$

This modulus is at most two and tends to zero as $t \downarrow 0$: a contrary sequence has, by compactness and because only finitely many paths have length at most H, a subsequence with convergent models and one fixed path, whose limiting models share their mean vector; (42) and continuity contradict the positive limiting row distance. Put $t_n = n^{-1/4}$, take a finite t_n-net of the compact set $\Psi(\mathcal{K})$ whose points are means of retained models, and let $\widehat{T}_n$ be the model of the net point closest to the averaged observation $\overline{Y}$, with fixed ties; this finite-valued rule is measurable. If $Z = \overline{Y} - \Psi(T)$ satisfies $\|Z\| \leq t_n$, then $\|\Psi(\widehat{T}_n) - \Psi(T)\| \leq t_n + 2\|Z\| \leq 3t_n$. Since $\mathbb{E}\|Z\|^2 = D\sigma^2/n$, the complementary event has probability at most $D\sigma^2/\sqrt{n}$. Use $L_n^*(Y,e,p) = \mathrm{sign}(u_{\widehat{T}_n}(p)^\top e)$. On the event $\|Z\| \leq t_n$ its regret is at most $\min\{1, \omega_H(3t_n)\}$ on all paths with $|p| \leq H$ and all unit facts; on every other query it is at most one. Thus the left side of (46) is bounded by

$$\min\{1, \omega_H(3n^{-1/4})\} + \frac{D\sigma^2}{\sqrt{n}} + q(|p| > H). \tag{47}$$

Choose H to control the fixed law's tail, then increase n; the uniform tail condition does the same for $\mathcal{Q}$. Neither the net nor the learner depends on H, q or the test facts. □

Order of quantifiers. Every probability on finite paths has a vanishing length tail, even with unbounded support. When the type condition fails, however, Theorem D.2 still gives $\liminf_n \inf_L \sup_q \sup_T \mathbb{E}_{Y,p\sim q,e\sim\nu} r > 0$, since q may be a point mass on the difficult path for that n; thus $\exists(L_n)\,\forall q$ $\lim_n$ cannot be replaced by $\lim_n \inf_L \sup_q$. Uniform tail control is sufficient, not necessary when long paths repeat known response functions, and the result relies on (19) on a fixed finite graph; it does not apply to the continuous-base experiment with a permanently fixed training prefix.

D.3 A SHARP PRICE FOR THE DISTRIBUTION OF REPEATED USE

For the matched planar construction the full law-dependent rate is explicit. Its lower bound needs one common two-point pair: summing per-horizon minimax bounds would let the hard model vary with the sampled path.

Theorem D.4 (Distribution-dependent risk in the matched return model)**.** *Use either matched two-vertex graph, the angle interval $[0,\rho]$, and the unchanged four-scalar Gaussian calibration experiment of Proposition 4.3, with $0 < \rho \le 0.1$. Deployment samples an arbitrary finite legal root path p from a predeclared law q, independently of the unknown model and training data, and an independent uniform planar fact. Let $N_b(p)$ count traversals of the unknown crossing b, and let $q_k = q\{p : N_b(p) = k\}$. In either matched graph, let*

$$\mathfrak{R}_n(q) = \inf_L \sup_{0\le\theta\le\rho} \sum_p q(p)\,\mathbb{E}_Y \int r(L(Y,e,p), c^\top T_p(\theta)e)\,d\sigma_2(e), \qquad s_n = \min\{\rho, \sigma/\sqrt{n}\}.$$

There are absolute constants $0 < c_* \le C_* < \infty$ such that

$$c_* \sum_{k\ge 0} q_k \min\{1, k^2 s_n^2\} \;\le\; \mathfrak{R}_n(q) \;\le\; C_* \sum_{k\ge 0} q_k \min\{1, k^2 s_n^2\}. \tag{48}$$

The constants are uniform over q, n, σ, ρ. The statement also holds when $q = q_n$ is fixed before the data for each n. The upper bound uses a single angle estimator independent of q; the lower bound permits arbitrary randomized decoders. At $\rho = 0$, or when $q_0 = 1$, the risk and the displayed sum are zero.

Proof. For unit rows u, v, rotational symmetry gives $\int e\,\mathrm{sign}(v^\top e) d\sigma_r(e) = \kappa_r v$. Thus the loss in (3) satisfies

$$\int r(\mathrm{sign}(v^\top e), u^\top e) d\sigma_r(e) = \frac{\kappa_r}{2}(1 - u^\top v) = \frac{\kappa_r}{4}\|u - v\|^2. \tag{49}$$

The matched edge responses are commuting planar rotations, so every legal word has $T_p(\theta) = R_{\gamma(p)+N_b(p)\theta}$ with $\gamma(p)$ known. Under uniform planar facts the known rotation $R_{\gamma(p)}$ affects neither the integrated loss nor the two-model separation, so only $N_b(p)$ matters, not the length or known self-loops. By the reduction of Appendix C.2, with $\tau = \sigma/\sqrt{n}$ in place of ν, the experiment is $Y = R_\theta + \tau Z$ with four independent standard normal entries of Z. A nearest-curve fit has $\|R_{\widehat{\theta}} - R_\theta\|_F \le 2\tau\|Z\|_F$, hence mean squared error at most $16\tau^2$. If $\rho \le \tau$, use instead the fixed estimate zero, whose squared matrix error is at most $2\rho^2$; this choice depends only on the known ρ, τ. Orthogonal telescoping gives $\|R_{k\widehat{\theta}} - R_{k\theta}\|_{\mathrm{op}} \le \min\{2, k\|R_{\widehat{\theta}} - R_\theta\|_{\mathrm{op}}\}$, so for $k = N_b(p)$, equation (49) bounds the risk of each path by an absolute constant times $\min\{1, k^2 s_n^2\}$. Summing over paths grouped by $N_b(p)$ gives the upper bound, with one estimator for every k.

For the lower bound fix any learner and set $a = \min\{\rho, \tau/(2\sqrt{2})\}$, so $s_n/(2\sqrt{2}) \le a \le s_n$. For every $\delta \in [0, a]$, the training laws at angles zero and δ have $\mathrm{KL}(P_0\|P_\delta) = \|R_0 - R_\delta\|_F^2/(2\tau^2) \le \delta^2/\tau^2 \le 1/8$, hence $\mathrm{TV}(P_0, P_\delta) \le 1/4$. The overlap argument in Theorem D.2, now also summed over the *same* sampled path law, gives

$$\sup_\theta R_\theta(L, q) \ge \frac{3\kappa_2}{8} \sup_{0\le\delta\le a} \sum_k q_k\big(1 - |\cos(k\delta/2)|\big).$$

The absolute value is necessary when $k\delta$ spans several periods; replacing it by $\cos(k\delta/2)$ would be incorrect. Write $g(t) = 1 - |\cos(t/2)|$ and $h(z) = z^{-1}\int_0^z g(t)\,dt$ for $z > 0$, with $h(0) = 0$. For $z \le 1$, $g(t) = 1 - \cos(t/2) \ge t^2/(2\pi^2)$, and hence $h(z) \ge z^2/(6\pi^2)$. For $1 \le z \le 2\pi$, integration

over $[0,1]$ gives $h(z) \geq 1/(12\pi^3)$. For $z \geq 2\pi$, there are at least $z/(4\pi)$ complete periods and each has integral $2\pi - 4$; therefore $h(z) \geq (2\pi - 4)/(4\pi)$. Together, $h(z) \geq c_0 \min\{1, z^2\}$ with $c_0 = 1/(12\pi^3)$. Averaging the last supremum over $\delta \in [0, a]$ and exchanging sum and integral by nonnegativity yields

$$\sup_{0\leq\delta\leq a} \sum_k q_k g(k\delta) \geq \sum_k q_k h(ka) \geq \frac{c_0}{8} \sum_k q_k \min\{1, k^2 s_n^2\}.$$

Thus $c_* = 3\kappa_2 c_0/64$ is valid. For a fixed learner one common δ attains at least this average; it is chosen before the training observations and cannot depend on the realized query. Taking the infimum over learners proves the claim. □

Matched finite-horizon consequence. For all paths of length at most integer $H \geq 2$ in the original two graphs,

$$\mathcal{R}_{n,H}^{\sigma_2,\to} \asymp s_n^2, \qquad \mathcal{R}_{n,H}^{\sigma_2,\mathrm{ret}} \asymp \min\{1, H^2 s_n^2\}. \tag{50}$$

By the edge count of Appendix C.2 (at most one use of the uncertain edge without return, at most H with it), the preceding estimator gives both upper bounds. For the lower bounds, use the uncertain crossing once in the one-way graph, which is the $k = 1$ two-point experiment, and a point mass at $k = \lfloor H/2 \rfloor$ in Theorem D.4 for the return graph, with $\lfloor H/2 \rfloor \geq H/3$. The exponent differs from (10) because both the mass of the disagreement region and the wrong-decision margin shrink for nearby normals.

The tail of future use. If the uncertain-edge count under a fixed path law satisfies $\Pr_q(N_b(p) > t) \asymp t^{-\alpha}$ for $t \geq 1$, $\alpha > 0$, then, as $s = s_n \downarrow 0$,

$$\mathfrak{R}_n(q) \asymp \begin{cases} s^\alpha, & 0 < \alpha < 2, \\ s^2 \log(1/s), & \alpha = 2, \\ s^2, & \alpha > 2. \end{cases} \tag{51}$$

Layer-cake integration gives the exact identity $\sum_k q_k \min\{1, k^2 s^2\} = 2s^2 \int_0^{1/s} t \, \Pr_q(k > t)\, dt$. The integral over $[0,1]$ is at most s^2; the two-sided tail bounds on $[1, 1/s]$ and integration of $t^{1-\alpha}$ give the three cases, with constants depending on α and the tail-comparison constants, and for $\alpha > 2$ the lower bound follows from any finite interval with positive tail mass. For fixed $\rho, \sigma > 0$ the eventual rates are $n^{-\alpha/2}$, $(\log n)/n$ and n^{-1}. A fixed law may thus put mass on arbitrarily long returns, even with infinite mean length, and still permit consistent decisions at a rate set by its tail.

Growing future use. For a sequence of predeclared laws q_n, (48) also characterizes consistency: $\mathfrak{R}_n(q_n) \to 0$ if and only if $s_n k \to 0$ in probability for $k = N_b(p)$, $p \sim q_n$. Writing q_n also for its pushforward to k, necessity follows from $\Pr_{q_n}(s_n k > \epsilon) \leq \mathbb{E}_{q_n} \min\{1, s_n^2 k^2\} / \min\{1, \epsilon^2\}$ for each $\epsilon > 0$. For sufficiency, split that bounded expectation at $s_n k \leq \epsilon$; it is at most $\epsilon^2 + \Pr_{q_n}(s_n k > \epsilon)$ for $0 < \epsilon < 1$. First let $n \to \infty$, then $\epsilon \downarrow 0$. For fixed $\rho, \sigma > 0$ this means $N_b(p) = o_{q_n}(\sqrt{n})$, not that the largest possible query must be $o(\sqrt{n})$.

E Changing deployment laws beyond commuting responses

Calibration and task geometry. This appendix gives the full risk definition and finite-sample theorem behind the changing-deployment criterion of Section 4; it gives no efficient modulus computation, dimension-free rate or neural guarantee. Keep the fixed finite graph, nonempty compact class $\mathcal{K} \subset O(r)^E$ and unit task direction c. Let $\Psi : \mathcal{K} \to \mathbb{R}^D$ be the known continuous mean map of a fixed finite Gaussian calibration experiment, with n independent repeats of each coordinate, for instance (4) or the four-probe experiment of Proposition 4.3. Exact identification by Ψ is not required. With r, D and $\sigma > 0$ fixed, write $u_T(p) = T_p^\top c \in \mathbb{S}^{r-1}$ and $\overline{Y} = \Psi(T) + \tau_n Z$, where $\tau_n = \sigma/\sqrt{n}$ and $Z \sim N(0, I_D)$. An orthogonal change in replicate coordinates separates each probe's mean from Gaussian residuals independent of T, so the means are sufficient and residuals supply only model-independent randomization; the lower bound covers randomized learners using all observations.

Let $\mathcal{Q}_n$ be any nonempty, deterministic, predeclared family of probability laws on finite legal root paths. Both the law family and any $q \in \mathcal{Q}_n$ are independent of the training data; a path $p \sim q$ and a uniform spherical fact $e \sim \sigma_r$ are sampled independently. A learner sees $(\overline{Y}, p, e)$, neither the unknown model nor q. Define

$$\mathcal{R}_n(\mathcal{Q}_n) = \inf_{\mathcal{L}} \sup_{T \in \mathcal{K}} \sup_{q \in \mathcal{Q}_n} \mathbb{E}_{\overline{Y}, p \sim q, e \sim \sigma_r} r\big(\mathcal{L}(\overline{Y}, e, p), u_T(p)^\top e\big), \tag{52}$$

$$\Omega_{\mathcal{Q}}(t) = \sup_{\substack{T,S \in \mathcal{K} \\ \|\Psi(T) - \Psi(S)\|_2 \le t}} \sup_{q \in \mathcal{Q}} \sum_p q(p) \|u_T(p) - u_S(p)\|_2^2. \tag{53}$$

This deterministic class-level modulus lies in $[0, 4]$ and is nondecreasing.

With κ_r as in Appendix D and Φ the standard normal distribution function, for unit u, v,

$$\int r(\operatorname{sign}(v^\top e), u^\top e)\, d\sigma_r(e) = \frac{\kappa_r}{4} \|u - v\|_2^2, \tag{54}$$

$$\begin{aligned} A(u, v) &:= \int \min_{b \in \{-1,1\}} \{r(b, u^\top e) + r(b, v^\top e)\}\, d\sigma_r(e) \\ &= \kappa_r \left(1 - \frac{\|u + v\|_2}{2}\right) \ge \frac{\kappa_r}{8} \|u - v\|_2^2. \end{aligned} \tag{55}$$

Here (54) is (49). Moreover $A = A_{\sigma_r}$ from the proof of Theorem D.2, and $\int |w^\top e|\, d\sigma_r(e) = \kappa_r \|w\|$ gives the equality in (55). For $d = \|u - v\|$, use $\|u + v\|/2 = \sqrt{1 - d^2/4}$ and $1 - \sqrt{1 - x} \ge x/2$ on $[0, 1]$.

Theorem E.1 (A multiscale criterion for moving deployment families)**.** *For every n and every $a > 0$,*

$$\mathcal{R}_n(\mathcal{Q}_n) \ge \frac{\kappa_r}{8} \Phi(-a/2)\, \Omega_{\mathcal{Q}_n}(a\tau_n), \tag{56}$$

$$\mathcal{R}_n(\mathcal{Q}_n) \le \frac{\kappa_r}{4} \Omega_{\mathcal{Q}_n}((1 + 2a)\tau_n) + \kappa_r \Pr(\|Z\|_2 > a). \tag{57}$$

A single deterministic sequence of finite-net estimators, independent of every $\mathcal{Q}_n$, attains the upper bounds. Consequently,

$$\mathcal{R}_n(\mathcal{Q}_n) \longrightarrow 0 \quad \Longleftrightarrow \quad \text{for every fixed } a > 0, \quad \Omega_{\mathcal{Q}_n}(a\sigma/\sqrt{n}) \longrightarrow 0. \tag{58}$$

No assumption of convexity or commutativity is needed. The result concerns predeclared laws, not paths chosen adversarially after observing the calibration noise.

Proof. For the lower bound fix a learner and any pair T, S satisfying the constraint in (53), and let $\Delta = \|\Psi(T) - \Psi(S)\|_2 \le a\tau_n$. The sample-mean laws P_T, P_S have the same covariance, so their common density has mass $1 - \mathrm{TV}(P_T, P_S) = 2\Phi(-\Delta/(2\tau_n)) \ge 2\Phi(-a/2)$, including $\Delta = 0$: with the midpoint at zero and the mean difference along the first coordinate, the smaller density over each halfspace of the log likelihood ratio's zero hyperplane gives one normal tail, and the other coordinates integrate to one. For any fixed q, the overlap argument in the proof of Theorem D.2, with $\nu = \sigma_r$, internal randomization averaged into $g_Y \in [-1, 1]$, this mass in place of $3/4$, and summation over $p \sim q$, bounds the sum of the learner's two expected risks below by $2\Phi(-a/2)(\kappa_r/8) \sum_p q(p) \|u_T(p) - u_S(p)\|_2^2$ by (55), and the larger risk is at least half of it. Since the pair and q are fixed before the data, supremizing over admissible (T, S, q), using approximating triples if necessary, and infimizing over learners proves (56).

For the upper bound use the nearest-net estimator $\widehat{T}_n$ of the proof of Theorem D.3, built for the present Ψ from a finite τ_n-net of the compact set $\Psi(\mathcal{K})$; it is measurable, finite-valued and independent of the deployment law, and $\|\Psi(\widehat{T}_n) - \Psi(T)\|_2 \le \tau_n + 2\|\overline{Y} - \Psi(T)\|_2$. Decode by $\operatorname{sign}(u_{\widehat{T}_n}(p)^\top e)$. On $\{\|Z\| \le a\}$, conditional risk averaged over any $q \in \mathcal{Q}_n$ is, by (54), at most $\kappa_r \Omega_{\mathcal{Q}_n}((1 + 2a)\tau_n)/4$. Off that event it is at most κ_r, as squared unit-row distances are at most four. Averaging over Z, which uses independence of q from Y, proves (57).

If the minimax risk tends to zero, (56) and $\Phi(-a/2) > 0$ force the modulus to vanish at each fixed a. Conversely, if all such moduli vanish, (57) yields $\limsup_n \mathcal{R}_n(\mathcal{Q}_n) \le \kappa_r \Pr(\|Z\| > a)$ for every fixed a. Since D is fixed, this tail tends to zero as $a \to \infty$; Markov's inequality bounds it by D/a^2. □

Relation to all-path and fixed-law calibration. For a fixed family $\mathcal{Q}$, the criterion is equivalent to $\Omega_{\mathcal{Q}}(t) \to 0$ as $t \downarrow 0$; for the reverse implication, use monotonicity and $\tau_{n+1} < t \le \tau_n$. For the bank (4), when $\mathcal{Q}$ contains every path point mass, the criterion, Lemma D.1 and Theorem 4.2 recover the significant-type obstruction. For that same bank, when $\mathcal{Q}$ is uniformly tight in path length, $\Omega_{\mathcal{Q}}(t) \le \omega_H(t)^2 + 4\sup_{q\in\mathcal{Q}} q(|p| > H)$ for each H, where $\omega_H(t) \to 0$ by the proof of Theorem D.3, using (19), and the second term is small for large H; this recovers that theorem's sufficient tail condition. The criterion also covers non-tight families whose long responses cancel.

Statistical moduli and two-point testing are classical (Polyanskiy & Wu, 2026); we use only the testing principle, not their affine-functional, convex-class duality (Appendix A). Nor do (56)–(57) imply a universal sharp rate $\mathcal{R}_n \asymp \Omega_{\mathcal{Q}_n}(\tau_n)$, as the following example shows.

Why one noise radius is insufficient.

Proposition E.2 (A compact single-radius counterexample)**.** *In the original matched return graph, with its unchanged four-scalar experiment and uniform planar facts, there are a fixed compact angle class and predeclared path laws q_n such that*

$$\Omega_{\{q_n\}}(n^{-1/2}) \to 0, \qquad \limsup_n \mathcal{R}_n(\{q_n\}) > 0.$$

Thus the quantifier over all fixed positive multiples in (58) *cannot in general be replaced by the single multiplier one.*

Proof. Set $\sigma = 1$, $N_j = 2^{2^{j+3}}$, $\tau_j = N_j^{-1/2}$ and $\theta_j = 2\tau_j$ for $j \ge 1$. Use angles $\Theta = \{0, \theta_1, \theta_2, \ldots\} \subset [0, 0.1]$ with all other edges fixed. The class is compact since $\theta_j \to 0$. At $n = N_j$, let q_n put all mass on $(bd)^{h_j}$, where d is the known identity return and $h_j = \lfloor \pi/(2\theta_j) \rfloor$. At every other integer n, let q_n put all mass on the empty path. These laws are fixed independently of all data.

By the reduction of Appendix C.2, $\Psi(\theta)$ is $\mathrm{vec}(R_\theta)$ up to an isometry, so $\|\Psi(\theta) - \Psi(\theta')\| = 2\sqrt{2}|\sin((\theta - \theta')/2)|$. The ratios θ_{j+1}/θ_j tend to zero and are at most $1/256$. If two distinct angles have at least one index at most j (the index of zero being infinity), their separation x is at least $\theta_j(1 - 1/256)$; since every angle is below 0.1, $\sin(x/2) \ge x/\pi$, so their mean-vector distance is at least $(4\sqrt{2}/\pi)(1 - 1/256)\tau_j > \tau_j$. Any unequal pair admitted by radius τ_j therefore lies in $\{0, \theta_{j+1}, \theta_{j+2}, \ldots\}$. The output row distance is bounded by $h_j|\theta - \theta'|$, so $\Omega_{\{q_{N_j}\}}(\tau_j) \le h_j^2\theta_{j+1}^2 \le (\pi^2/4)(\theta_{j+1}/\theta_j)^2 \to 0$. Identical pairs contribute zero. At other n the modulus is zero because the empty path has the same row in every model.

In contrast, the pair $(0, \theta_j)$ has mean distance at most $\sqrt{2}\theta_j = 2\sqrt{2}\tau_j < 3\tau_j$, while its squared task-row distance on the chosen path is $4\sin^2(h_j\theta_j/2) \to 2$. Using (56) with $a = 3$ gives $\liminf_{j\to\infty} \mathcal{R}_{N_j}(\{q_{N_j}\}) \ge (\kappa_2/4)\Phi(-3/2) > 0$, so the full sequence does not converge to zero. Being disconnected rather than the interval class of Theorem D.4, this class does not contradict that theorem's sharp single-scale formula. □

Identical counts, different uncertainty.

Proposition E.3 (An exact equal-count order contrast)**.** *On a one-vertex graph let action a have unknown R_θ, $0 \le \theta \le \rho \le 0.1$, and let s have known response $J = \mathrm{diag}(1, -1)$. Supply a known quarter-turn action and the same four scalar rotation-calibration observations as in Proposition 4.3. For $m \ge 1$, the words*

$$p_m = (as)^{2m}, \qquad q_m = a^{2m}s^{2m}$$

have identical length $4m$, endpoint and counts of both actions. Yet $T_{p_m} = I$ for every θ, whereas $T_{q_m} = R_{2m\theta}$. Their spherical-fact minimax risks are respectively

$$0 \qquad \textit{and} \qquad \asymp \min\{1, m^2 \min(\rho, \sigma/\sqrt{n})^2\}.$$

Proof. The identity $JR_\theta J = R_{-\theta}$ gives $(JR_\theta)^2 = I$ and $J^{2m} = I$. With $T_{uv} = T_vT_u$, these prove both word identities. The first task row is known, so choosing its correct sign has zero regret. The second experiment and query are exactly the $k = 2m$ rotation-query experiment analyzed in Theorem D.4; its point-mass upper and common-pair lower bounds give the displayed rate. Except at special angles J and R_θ do not commute. □

F A BOUNDARY: KNOWN RESPONSE TYPE WITHOUT STABLE CALIBRATION

Use the task and regret in (3), with root $h = 0$ on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ and known nominal map $F(h) = h + \alpha$; bad approximability means $\|q\alpha\|_{\mathbb{T}} \geq \gamma/q$ for all integers $q \geq 1$ and some $\gamma > 0$. For a response field T, write $P_T(H; h) = T(F^{H-1}h) \cdots T(h)$. Each of N training queries, with starting point, unit fact e and length $k \leq B_0$ chosen adaptively, observes $c^\top P_T(k; h)e$ plus independent $N(0, \sigma^2)$ noise. Deployment has no further true-model queries.

Proposition F.1 (Unstable calibration with a known response type)**.** *For every badly approximable α, on the known circle rotation $F(h) = h + \alpha$ there exist known $A \in SO(3)$, unit c, and a compact C^∞ class $\{A, T_1, T_2, \ldots\}$, all smoothly gauge equivalent to A and uniformly task observable, with $T_j \to A$ in every C^k norm. Nevertheless, after any finite number of independent Gaussian-noisy scalar calibration queries of bounded length and fixed variance $\sigma^2 > 0$, every learner has worst-model, worst-path expected regret at least $a_0 > 0$, where a_0 depends on neither the number or length bound of the queries nor σ. Queries may be adaptive and start at supplied nominal points; facts are unit vectors and the task is* (3)*.*

Proof of Proposition F.1. Let Z_t and X_t denote rotations about the third and first axes. A response frequency almost resonant with ever higher nominal frequencies comes from increasing integers q_j and $\beta \in (1/8, 3/8)$ with $0 < \|q_j\alpha - \beta\|_{\mathbb{T}} < q_j^{-j}$. Starting from an orbit point in this interval, set $q_{j+1} = q_j + m_j$, where $m_j > q_j$ has positive residue $\varepsilon_j < \min\{\varepsilon_{j-1}/4, q_j^{-j-2}/8\}$, as density of the irrational orbit permits; a small enough first residue keeps the increasing lifts inside the interval, their limit is β, and the geometric tail gives the estimate after discarding finitely many terms. Bad approximability of α is not violated, since the approximated β is nonzero.

Fix $a \in (0, 0.1)$, set $\omega = 2\pi\beta$ and $A = Z_\omega$, and define $Q_j(h) = Z_{2\pi q_j h} X_a Z_{-2\pi q_j h}$ and $T_j(h) = Q_j(Fh)AQ_j(h)^\top$. Every field is smoothly gauge equivalent to A, and the gauges contract to the identity by replacing a with ta, $0 \leq t \leq 1$. Put $\delta_j = 2\pi(q_j\alpha - \beta - \text{nearest integer})$ and $C_a(t) = Z_t X_a Z_{-t} X_{-a}$. Direct multiplication and telescoping give

$$\begin{aligned} T_j(h) &= AZ_{2\pi q_j h} C_a(\delta_j) Z_{-2\pi q_j h}, \\ P_{T_j}(H; 0) &= Z_{H\omega} C_a(H\delta_j). \end{aligned} \tag{59}$$

Since $C_a(0) = I$, $\|T_j - A\|_{C^k} \leq C_{a,k}(1 + q_j)^k |\delta_j| \to 0$ for each fixed derivative order k, so a tail together with A is compact in the C^∞ topology. Set $c = (e_1 + e_3)/\sqrt{2}$. The rows $c^\top, c^\top A, c^\top A^2$ are independent because $\omega \notin \{0, \pi\}$, so by uniform local convergence the three-prefix observation matrices have a common positive lower singular-value bound at every nominal point on a tail, which is uniform task observability. At H_j nearest $\pi/|\delta_j|$, however, $C_a(H_j\delta_j) \to X_{-2a}$. Since $u_j = Z_{-H_j\omega}c$ has third coordinate $1/\sqrt{2}$, $\|(X_{2a} - I)u_j\| \geq \sqrt{2}\sin a$, so on a tail the task normals $P_{T_j}(H_j; 0)^\top c$ and $(A^{H_j})^\top c$ are at distance at least $\sin a$. Their normalized difference is a common unit root fact giving opposite scores of magnitude at least $\sin a/2$. For any starting point and any path of length at most B_0, orthogonality gives $\|P_{T_j}(k; h) - A^k\|_{\text{op}} \leq B_0\|T_j - A\|_\infty$, so the Gaussian chain rule of Appendix B.1 gives

$$\mathrm{KL}(P_{T_j}^{N,B_0} \| P_A^{N,B_0}) \leq \frac{NB_0^2}{2\sigma^2}\|T_j - A\|_\infty^2 \longrightarrow 0. \tag{60}$$

For any finite N, B_0, choose j so this is at most $1/8$. The two-point step of Appendix B.1, with this fact and H_j fixed before the noise, gives every randomized learner expected regret at least $3\sin a/16$ in one model, independently of the training budget. The same fixed compact class works for all finite budgets and all $\sigma > 0$. □

Linearizing $Q = \exp v$ and $TA^{-1} = \exp E$ gives $E = v \circ F - \mathrm{Ad}_A v$, with transverse Fourier divisors $e^{2\pi i q\alpha} - e^{\pm i\omega}$: the obstruction is unstable recovery of coordinates through this classical cocycle resonance (Karaliolios, 2017), not a new response type or an expanding transport norm; the construction adds the fixed-task statistical lower bound and neither contradicts the finite-graph theorem nor asserts this obstruction for the visual experiments.

F.1 THE QUANTITATIVE SIDE OF THE SAME BOUNDARY

For planar rotation responses over the same circle map, the quantities below locate this boundary, and the corollary makes it a sharp regularity threshold with rates. Define the response field and accumulated phase by

$$T_\varphi(h) = R_{\omega+\varphi(h)}, \quad S_k\varphi(h) = \sum_{j<k} \varphi(h + j\alpha), \quad P_\varphi(k) = R_{k\omega+S_k\varphi(0)}. \tag{61}$$

For $s > 1/2$, $M > 0$ and $0 < \kappa < \pi/4$, use the compact C^0 class

$$\mathcal{F}_s(M,\kappa) = \{\varphi \text{ real} : \widehat{\varphi}_0 = 0,\ \sum_{q\neq 0}(1+q^2)^s|\widehat{\varphi}_q|^2 \le M^2,\ \|\varphi\|_\infty \le \kappa\}. \tag{62}$$

Here facts e, c and each e_i are unit vectors in $\mathbb{R}^2$, with e_1, e_2 the standard basis; the $SO(3)$ notation above is not carried over. Training observes $Y_{j,i,t} = c^\top P_\varphi(j)e_i + \xi_{j,i,t}$ for $1 \le j \le L$, $i = 1, 2$ and $1 \le t \le n$, with independent $N(0, \sigma^2)$ noise. For arbitrary randomized or improper learners with unlimited state and exact nominal addresses but no deployment query, let

$$\mathcal{R}_{n,L} = \inf_{\mathcal{L}} \sup_{\varphi\in\mathcal{F}_s} \sup_{e,k} \mathbb{E}_\varphi r(\mathcal{L}(Y, e, k), c^\top P_\varphi(k)e). \tag{63}$$

Recovering an endpoint frame inverts $u \mapsto u \circ F - u$, whose Fourier multiplier $d_q = e^{2\pi iq\alpha} - 1$ can be small even at high frequencies. With r_L the covering radius of $\{j\alpha : 0 \le j \le L\}$, the size and continuity of the inverse are measured by

$$A_s = \sum_{q\neq 0} \frac{(1+q^2)^{-s}}{|d_q|^2}, \qquad \Gamma_s(t)^2 = \sum_{q\neq 0} \frac{\min\{4, (2\pi|q|t)^2\}}{(1+q^2)^s|d_q|^2}. \tag{64}$$

Proposition F.2 (Continuous calibration boundary)**.** *In* (63)*,* $\inf_{n,L<\infty} \mathcal{R}_{n,L} = 0$ *iff* $A_s < \infty$*. In that case a coherent frame estimator gives*

$$\mathcal{R}_{n,L} \le \min\{1, M\Gamma_s(r_L) + 2\sqrt{2}\sigma/\sqrt{n}\}. \tag{65}$$

With probability $1 - \beta$ *for each true model, all its query regrets are at most* $M\Gamma_s(r_L) + 4\sigma\sqrt{\log(4(L+1)/\beta)/n}$*. If* $A_s = \infty$*, a fixed compact subclass of zero and smooth trigonometric polynomials, all continuously gauge equivalent to the known* R_ω*, has risk at least* $3/(8\sqrt{2})$ *for every finite* n, L*. A positive floor survives any finite noisy bounded-length query budget, including adaptive queries at supplied nominal starting points.*

Proof. For sufficiency we estimate one common frame and compose it coherently, rather than fitting a separate product at each depth. If $A_s < \infty$, Cauchy–Schwarz gives the uniformly absolutely convergent Fourier solution

$$u(h) = \sum_{q\neq 0} \frac{\widehat{\varphi}_q}{d_q} e^{2\pi iqh}, \quad u(Fh) - u(h) = \varphi(h), \quad |u(x) - u(y)| \le M\Gamma_s(\mathrm{dist}(x, y)), \tag{66}$$

and dominated convergence gives $\Gamma_s(t) \to 0$. For $Q(h) = R_{u(h)-u(0)}$ the transition is $Q(Fh)R_\omega Q(h)^\top$. Multiplying the averaged two-coordinate row at prefix j by $R_{-j\omega}$ gives $c^\top Q(j\alpha)$ plus isotropic noise of variance σ^2/n. Project it to the unit circle, lift to its unique $SO(2)$ matrix, interpolate by nearest sampled nominal point with fixed ties and $\widehat{Q}(0) = I$, and use

$$\widehat{T}(h) = \widehat{Q}(Fh)R_\omega\widehat{Q}(h)^\top, \qquad \widehat{P}(k) = \widehat{Q}(k\alpha)R_{k\omega}. \tag{67}$$

Every fitted product telescopes, although the nearest-neighbor frame need not be smooth, so a depth-k query interpolates only its endpoint frame, with deterministic error at most $M\Gamma_s(r_L)$. Projection doubles the sampled-row noise, whose expected norm is at most $\sqrt{2}\sigma/\sqrt{n}$; each fixed query uses one row, which proves the expected-risk bound without a union over deployment depths. Gaussian coordinate tails over the $2L$ training entries give the simultaneous guarantee, and density gives $r_L \to 0$.

For necessity we build perturbations that vanish on every fixed training window but keep a fixed phase error at a later depth. If $A_s = \infty$, put $b_{H,q} = (e^{2\pi i q H\alpha} - 1)/d_q$ and $B_{H,Q}^2 = \sum_{0<|q|\le Q} |b_{H,q}|^2 (1+q^2)^{-s}$. The Cesaro mean over H is $2\sum_{0<|q|\le Q}(1+q^2)^{-s}|d_q|^{-2}$, so there are (H_j, Q_j) with $B_{H_j,Q_j} \to \infty$. For $\eta = \pi/2$, choose

$$\widehat{\varphi}_q = \frac{\eta\,\overline{b_{H,q}}}{(1+q^2)^s B_{H,Q}^2}\ (0<|q|\le Q), \qquad S_H\varphi(0) = \eta, \quad \|\varphi\|_{H^s} = \eta/B_{H,Q}, \tag{68}$$

and let the other coefficients vanish. Conjugate symmetry makes the polynomials real, and $s > 1/2$ with Cauchy–Schwarz gives uniform convergence to zero, so a tail together with zero is compact in the declared class; each polynomial has a smooth Fourier gauge. Nevertheless

$$\mathrm{KL}(P_{\varphi_j}^{n,L} \| P_0^{n,L}) \le \frac{n\|\varphi_j\|_\infty^2}{2\sigma^2} \sum_{k=1}^{L} k^2 \to 0. \tag{69}$$

At H_j the normal-difference fact gives scores $\pm\sin(\eta/2)$; with j chosen so that the KL is at most $1/8$, the two-point step of Appendix B.1 gives the floor $3/(8\sqrt{2})$. For N adaptive scalar queries of length at most B, the Gaussian chain rule of that appendix bounds the conditional KL by $NB^2\|\varphi_j\|_\infty^2/(2\sigma^2)$ uniformly over starting points and unit facts, so the same fixed subclass suffices. □

Corollary F.3 (Regularity and decision transfer)**.** *If $\|q\alpha\|_{\mathbb{T}} \ge \gamma/q$ for every integer $q \ge 1$, then $A_s < \infty$ exactly for $s > 1$. For $1 < s < 2$, $\Gamma_s(t) \le C_{s,\gamma} t^{s-1}$ and $r_L \le C_\gamma/(L+1)$. No fixed finite L gives consistency by increasing n alone, even when $s > 1$. For $s > 1$ and sufficiently small fixed M_0, real mean-zero errors with $\|\varphi\|_2 \le \delta$, $\|\varphi\|_{H^s} \le M$ and $0 < \delta \le M \le M_0$ satisfy*

$$\sup_{\varphi} \sup_{k,e} r\big(\mathrm{sign}(c^\top R_{k\omega} e), c^\top P_\varphi(k) e\big) \asymp_{s,\alpha} M^{1/s}\delta^{1-1/s}. \tag{70}$$

The fixed sup-norm cap in (62) *does not change this order for small enough M_0.*

Proof. In $N \le |q| < 2N$, bad approximability separates distinct orbit points and zero by at least $\gamma/(4N)$. Ordering distances to zero on either side, using $|d_q| \ge 4\|q\alpha\|_{\mathbb{T}}$ and $\sum_{j\ge1} j^{-2} < \infty$, gives $\sum_{N\le|q|<2N} |d_q|^{-2} \le C_\gamma N^2$. Dyadic summation proves convergence of A_s for $s > 1$ and, split at $N \asymp 1/t$, the bound on Γ_s for $1 < s < 2$. Convergent denominators have $\|q\alpha\|_{\mathbb{T}} \le 1/q$, so for $s \le 1$ their summands in A_s do not vanish. Dirichlet approximation with a reduced denominator $q \le L+1$ gives $|q\alpha - p| \le 1/(L+1)$ and hence $q \ge \gamma(L+1)$; the first q orbit points lie within $1/(L+1)$ of the q-point uniform grid, proving the covering bound. For fixed L, choose smooth u zero at $0, \alpha, \dots, L\alpha$ but nonzero at a later orbit point; scaling $\varphi = u \circ F - u$ meets the class bounds while leaving identical training means and a later decision conflict.

For (70), Cauchy–Schwarz and the dyadic estimate bound the Fourier solution's part below frequency K by $C\delta K$ and its tail by CMK^{1-s}. Choosing $K \asymp (M/\delta)^{1/s}$ bounds u, every phase difference $u(k\alpha) - u(0)$, and therefore regret. For the reverse bound, bounded ratios of consecutive convergent denominators give $q \asymp_\alpha (M/\delta)^{1/s}$. Set $\varphi(h) = a\cos(2\pi q h)$, $a = c_{s,\alpha}\delta$, with the constant small enough to meet both norm constraints. Since $d = \|q\alpha\|_{\mathbb{T}} \asymp_\alpha q^{-1}$, take $k \asymp d^{-1}$ with a small constant so that the first k cosines are positive and $S_k\varphi(0) \asymp aq$; bounded small q use $k = 1$. Small M_0 keeps this phase below one. The normal-difference fact makes the reference rule choose the wrong sign with regret $|\sin(S_k\varphi(0)/2)| \asymp \delta q$, proving the lower order. □

G Information and cached execution

These information and arithmetic facts support the response interface of Section 3; none is a computational lower bound.

G.1 Nominal prediction and endpoint information

For $d \ge 1$, start at zero in $\mathbb{R}^d$, write $Be = ee_1$, shift $Te_i = e_{i+1}$ for $i < d$ and $Te_d = 0$, and read $C = e_d^\top$. These smooth deterministic maps give zero for every nominal observation (predictive

Hankel rank zero, or at most one with a constant coordinate), whereas $CT^kB = \mathbf{1}[k = d-1]$. The d-by-d response Hankel block is the reverse identity. A linear realization of width a factors it through $\mathbb{R}^a$, so $a \geq d$; the construction attains d.

Replace C by θC, $\theta = \pm 1$, and add a hold action $\bar{a}$ with transport I. At fact $e = 1$, equal-length plans $a^{d-1}\bar{a}$ and a^d have costs $(1, 0)$ or $(-1, 0)$, and their nominal transcripts coincide under every policy. A selector choosing the first with probability p has regrets p and $1 - p$, so minimax expected regret is $1/2$. If the permitted input laws instead have total variation at most δ, the two choice probabilities differ by at most δ, giving lower bound $(1 - \delta)/2$. Known sign or informative intervention data remove this restriction.

G.2 Task readout and its limitations

Fix the fact record and nominal endpoint. An endpoint-only response exists exactly when all compatible histories have the same profile on every legal suffix and probe; a measurable factorization additionally requires measurability, and constancy alone is not efficient decoding. If profiles differ by ω, the triangle inequality forces any common output to err by at least $\omega/2$ on one.

Equality of immediate responses is insufficient, as a specified play actuator shows. For deadband $d = 0.2$, the update is $y_{t+1} = \min\{x_{t+1}+d, \max\{x_{t+1}-d, y_t\}\}$. After a common command $x = 1$ produces $y = 0.8$, the continuations A: $(1, 0)$ and B: $(-1, 0)$ have the same nominal endpoint zero but factual trajectories $(0.8, 0.8, 0.2)$ and $(0.8, -0.8, -0.2)$. Squared terminal errors for target $+0.2$ are 0 and 0.16, so the common innovation -0.2 changes their ranking. The deadband is known; this is an analytic example, not identification from one observation.

A decision lower bound additionally requires conflicting optima: two indistinguishable cases with distinct unique optima and margins at least γ force deterministic worst regret at least γ, or randomized worst expected regret at least $\gamma/2$. A common scalar score shift cannot change a ranking, but a shared latent update can still produce branch-specific scores. When supplied weights and paths determine a response, precompilation moves computation rather than revealing information.

For $k_* \in \arg\min c_k$ and $\widehat{k} \in \arg\min(c_k + \delta_k)$, optimality of $\widehat{k}$ gives

$$c_{\widehat{k}} - c_{k_*} \leq \delta_{k_*} - \delta_{\widehat{k}} \leq \max_k \delta_k - \min_k \delta_k \leq 2\|\delta\|_\infty. \tag{71}$$

This holds for any tie breaking and preserves a unique optimum with margin greater than $2\|\delta\|_\infty$. Task rescoring also need not reconstruct a future: reconstructing a finite correction space requires probes injective on that space, and any common probe kernel remains invisible.

G.3 Conditional matrix compilation

For a zero-origin writer $O_\psi(0) = 0$ and frozen innovation-linear transport, $z_{k,t+1} = A_{k,t}z_{k,t}$, with matrices and reader context q_k known before feedback, store $P_k = A_{k,H-1} \cdots A_{k,0}$ and use

$$\widehat{V}_k(e) = V_k^0 + C_\omega(q_k, P_k O_\psi(e)) - C_\omega(q_k, 0). \tag{72}$$

The subtraction makes the correction zero at zero innovation. This compilation is not the algorithm timed in Table 1.

With nominal-only features $f_{k,t}$, reader context q_k, and a bias-free matrix W, set

$$g_{k,t} = \tanh(W_g f_{k,t} + b_g), \qquad A_{k,t} = I + \mathrm{diag}(g_{k,t})W, \qquad P_{k,t} = A_{k,t-1} \cdots A_{k,0}.$$

For zero-origin writer $z_0 = O_{\mathrm{raw}}(e) - O_{\mathrm{raw}}(0)$, induction gives $z_{k,t} = P_{k,t}z_0$ and hence (72). This is classical cached recurrence composition (Liu & Teitelbaum, 1995; Martin & Cundy, 2018), exact for the learned model and every partition of the same ordered path, not permutations. Affine offsets need an augmented coordinate, and an innovation-dependent gate, including the long-bank model, does not admit this factorization.

Suppose $\|z_0\| \leq R$, writer error is δ_z, stored-product error δ_P, and $\|P_k\|_{\mathrm{op}} \leq \Lambda$. Adding and subtracting $P_k\widehat{z}_0$ gives $\|\widehat{P}_k\widehat{z}_0 - P_k z_0\| \leq \Lambda\delta_z + \delta_P(R+\delta_z)$. With an L_C-Lipschitz reader, physical-model error η and residual arithmetic/context error ν, total score error is $E = \eta + \nu + L_C\{\Lambda\delta_z + \delta_P(R + \delta_z)\}$, and regret is at most $2E$.

At p bits per entry, terminal arrays use $pK(r^2 + r + 2)$ bits besides nominal data and weights. The fact has r scalars and application uses $O(Kr)$ workspace. Dense preparation costs $O(KHr^3)$ operations; application costs $O(Kr^2)$ plus writer/reader work. Prefix reconstruction may store $K(H+1)$ matrices rather than K terminal products. These counts neither certify numerical accuracy nor equal measured wall time.

H EXPERIMENTAL EVIDENCE

The experiments address response identification and composition (Section 5), matched-feedback quality and cost (Section 6), and the structural boundaries of reuse. Reported risks concern finite queries; they neither measure all-path minimax risk nor verify the theorem's assumptions for learned visual connections.

H.1 SUPERVISION, BASELINES, AND SHARED REPORTING CONVENTIONS

Nominal features follow $h^0_{k,t+1} = \Phi_\theta(h^0_{k,t}, a_{k,t})$; simulator forks give teacher states $h^{e,\mathrm{teach}}_{k,t+1}$. With $P_K = I - \mathbf{1}\mathbf{1}^\top/K$, available objectives include

$$\mathcal{L}_\square = \sum_{k,t} \|U_\psi(h^0_{k,t+1}, T_\psi(h^0_{k,t}, a_{k,t}, z_{k,t})) - h^{e,\mathrm{teach}}_{k,t+1}\|^2,$$

$$\mathcal{L}_\circ = \sum_{k,vw} \|T^{vw}_\psi(h, z) - T^w_\psi(\Phi^v_\theta(h), T^v_\psi(h, z))\|^2,$$

$$\mathcal{L}_{\mathrm{resp}} = \sum_{F,e,g} \|P_K(\widehat{\Delta V}_{g,F}(e) - \Delta V_{g,F}(e))\|^2 + \lambda_{\mathrm{rank}}\mathcal{L}_{\mathrm{ranking}}.$$

Ranking uses the same forked values; the direct chunk head is optional, and primitive-only composition omits $\mathcal{L}_\circ$. Not every experiment uses every term: the independent long bank uses binary cross entropy on simulator success, and no ablation shows that $\mathcal{L}_\circ$ improves physical success.

In short contact banks a frozen model prepares $K = 125$ futures; after one common action and its observed image, all candidates fork from that simulator state. Forks give training or evaluation labels, never selector inputs. Each focused task has 64 test banks and 1,792 nonidentity cells, with separate training and test case seeds and candidate realizations; the assembly splits nevertheless share one 50-instance scene pool and are not scene-independent.

Endpoint carries no path state, pooling carries order-invariant path features, broadcast shares one correction, and shuffled binding assigns an innovation to the wrong branch; reroll receives the common fact and regenerates futures; the physical oracle reads simulator outcomes and is not deployable. Unless noted, short-bank means use four seeds with paired bank-bootstrap 95% intervals conditional on those fits, unclustered by scene; public-backbone results use eight seeds; the independent long-bank and six-block screening confirmations resample seeds and roots. Intervals are pointwise unless specified.

Outcome measures. Unless stated otherwise, graded regret uses shaped reward accumulated until first success or scripted closeout, with no terminal success credit. Early success can therefore score below failure. Normalized regret $(r_{\max} - r_{\mathrm{selected}})/\max\{r_{\max} - r_{\min}, 10^{-9}\}$, captured value and library cost measure this success-truncated objective, not necessarily decision quality. The two six-block comparisons in Appendix H.9 instead use return G, which credits success through the horizon; binary success is unaffected. The earlier G study also reports a descriptive success-first rank. Reward-range headroom and pair disagreement alone do not validate reward–success alignment. Throughout, lag, actuator lag and persistent compliance name the same first-order-memory family; actuation latency is distinct, and affine gain error is the default.

Success-truncated reward often ranks failures above successes (Table 2). On the long bank, OEWM regret increases from .483 for failures to .629 for closeout successes and .987 for in-plan successes (1.9/82.1/16.0% of choices); other readers show the same ordering. Under this objective, conditional six-block success is only 16.4/14.2/36.4% for the stored-feature $k = 24$ screen/reroll/corrector, versus 71.5% nominal; selected success was not recorded on registered roots. Even Peg's mostly positive reward–success alignment does not make highest-reward selection outperform random choice

Table 2: Agreement of the success-truncated graded reward with binary success, measured post hoc except in the registered last row; nine nonidentity effects per root unless stated. Mixed: cells containing both successes and failures. Higher: share (%) of mixed cells in which successful candidates have higher mean reward than failed ones. AUC: probability, averaged over mixed cells, that a random success out-scores a random failure in the same cell. The last four columns give success (%) of the highest-reward candidate, the nominal plan, uniform random choice and the physical oracle; — marks values not computed or not defined.

Evaluation bank	Tables	Mixed	Higher	AUC	Top	Nominal	Random	Oracle
Long bank, backlash	13	575/576	38.3	.402	39.9	57.3	32.0	99.8
Registered six-block, lag	16, 17, 24	432/432	0.5	.184	70.6	71.5	53.3	100
Conditional six-block, lag	15, 25	432/432	0.5	.187	71.3	71.5	53.4	100
Peg, six blocks	16	426/432	90.8	.847	62.0	81.2	63.9	100
Door, six blocks	16	10/432	0.0	.005	94.0	96.1	96.0	96.1
Hammer, six blocks	16	430/432	0.0	.001	0.5	100	84.1	100
Second backbone, Assembly lag	16	432/432	1.9	.182	69.0	71.5	53.0	100
Focused Assembly, affine, 28 effects	22, 23	1636/1792	7.6	.152	36.7	74.3	47.1	91.3
New-pool default, 28 effects	22	1630/1792	8.3	.157	39.9	75.0	47.1	91.0
New-pool lag	22	432/432	0.0	.058	62.7	66.7	49.1	100
Three-block lag	22, 25	432/432	0.0	.053	59.5	66.7	49.1	100
Nine-block lag	25	432/432	0.0	.196	42.1	83.3	52.7	100
Three-block backlash	25, 28	349/360	33.0	.340	42.8	33.3	36.0	96.9
Six-block backlash, gate test	28	279/288	30.8	.334	52.4	33.3	37.1	96.9
Nine-block backlash, gate test	28	360/360	31.4	.338	31.4	58.1	41.3	100
Twelve-block backlash, gate test	28	286/288	33.2	.373	35.1	52.1	36.0	99.3
Six-block backlash	23, 24	419/432	30.1	.333	51.2	33.3	37.1	97.0
Six-block latency, ten effects	23	480/480	16.7	.328	53.8	52.1	45.2	100
Six-block affine, 39 roots, 28 effects	23, 24	995/1092	9.2	.219	41.7	77.8	48.7	91.1
Sixteen-chunk backlash, 32 roots	App. H.9	288/288	38.5	.401	35.4	55.9	32.3	100
Assembly, familiar words	19	32/288	100	.985	8.3	—	0.6	11.1
Assembly, unseen words	19	185/288	82.7	.760	35.8	—	6.5	64.2
Door, familiar words	19	181/288	0.0	.037	45.1	—	77.7	98.6
Door, unseen words	19	180/288	0.0	.021	41.7	—	80.3	98.6
Fresh-scene six-block backlash	26	553/576	28.9	—	55.9	33.3	37.1	96.0

in success. Replenished-library higher-reward shares/AUC ranges are 26.6–33.2%/.356–.375 (backlash), 2.3–5.1%/.221–.231 (compliance) and 11.4–11.8%/.233–.241 (affine). These findings motivate the separate G-trained evaluations below.

H.2 Response uncertainty on enumerated graphs

The study exhausts 78 three-vertex and 9,638 four-vertex labeled directed graphs with self-loops. Eligibility requires root reachability and both recurrent and crossing edges on indefinitely extendible root walks. Three response families at two placements give 58,296 structural labels, recorded before measurement. This perturbation study does not evaluate a learner under the theorem's prescribed noisy calibration protocol.

Base edge rotations exponentiate skew matrices with independent standard-normal upper-triangular entries. Perturbations left-multiply either all eligible recurrent or all eligible crossing edges (counts

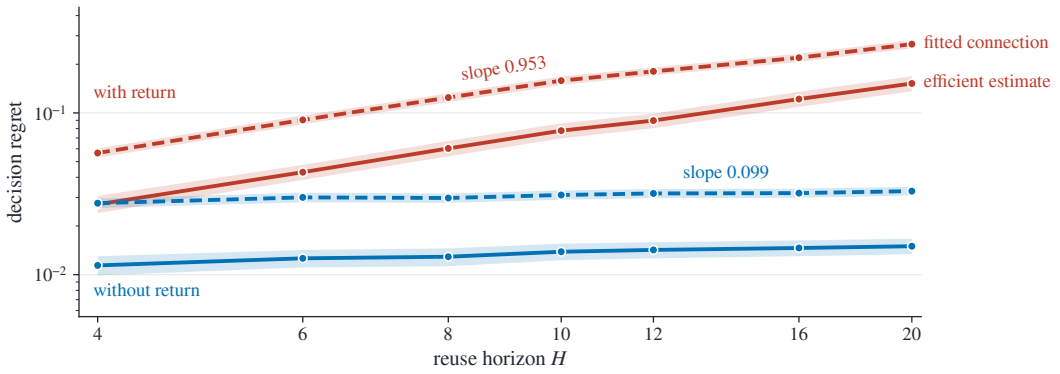


Figure 4: **Measured reuse of the same fit.** The matched graph structures are those of Proposition 4.3. The same estimate is evaluated on both, with mean finite-query sign-decision regret over 200 fitted systems and 95% bootstrap bands. These measurements are distinct from the proposition's minimax risk; this subsection specifies the protocol and aggregation limitations.

may differ) by independent rotations with coefficient standard deviation 0.02. The calibration control replaces selected recurrent edges T_a by $Q_y T_a Q_x^\top$, with a near-identity orthogonal Q_x per vertex, and fixes crossings; it preserves each recurrent component's type, not necessarily one global gauge. A supplied four-rotation finite-dictionary control uses substitutions that need not be locally small.

Each cell averages 96 candidate pairs at three vertices and 64 at four, with common random streams across configurations. At $H \in \{2, 4, 8, 16, 32, 64\}$, 128 legal root walks choose uniformly among outgoing edges permitting indefinite continuation. With $c = e_1$ and $e = e_1 - e_2$, decision damage is

$$d_H = \mathbb{E}_{T,S,p} |c^\top T_p e| \, \mathbf{1}\{\operatorname{sign}(c^\top T_p e) \neq \operatorname{sign}(c^\top S_p e)\}.$$

The fact norm is $\sqrt{2}$, not the theorem's unit convention; there is no training noise or fitted learner. Mean damage weights all 9,716 graphs equally, unfiltered; shared graphs and streams make cells non-independent. Short-response distance uses 64 sampled walks per length $1, \ldots, |V| + 1$ and basis facts, so small discrepancy is not a total-variation certificate.

At $H = 2, 8, 32, 64$, mean damage times 10^3 is $(.180, 1.354, 22.442, 77.571)$ for recurrent perturbations, $(.100, .156, .175, .168)$ for crossing perturbations and $(.213, .351, .388, .365)$ for recurrent vertex calibration.

Local fit does not identify the structural label: fitting unknown orthogonal edges to 288 noisy root scalars ($\sigma = .05$; 250 Adam steps, learning rate .05; known edges exact) in 468 cells of 24 systems gives error–label correlation .2034 and best threshold accuracy 83.33%, equal to the majority baseline. This cohort differs from the pair-substitution study. Sampling can also miss recurrent disagreement: 1,772 of 3,044 single-edge perturbation cells remain below mean damage .05 through $H = 1024$. Neither finite slopes nor sampled agreement certifies all-path learnability. Nonorthogonal distortion changes the gauge relation and is outside the theorem.

H.3 One estimate, two legal-reuse graphs

One draw of the four planar root-query means gives one $\widehat{\theta}$, evaluated on both graphs of Proposition 4.3 without refitting on the return or evaluation path. A sweep of 96 settings ($n \in [8, 8192]$, $\sigma \in [0.1, 1]$, $\rho \in [0.01, 0.1]$), each at 11 horizons $H \in [2, 64]$ (1,056 rows), uses 3,000 trials per row. In 95 of the 96 settings risk is positive at every horizon, and in all 95 the return slope exceeds the one-way slope. Median log–log slopes are 0.083 [IQR $0.038, 0.148$] and 1.051 [IQR $1.007, 1.113$], which are not confidence intervals; the median risk ratio at $H = 64$ is 30.3.

A three-dimensional gradient-fit variant ($\theta \sim \mathrm{Unif}[0, 0.35]$, $n = 64$, $\sigma = 0.5$) fits one free matrix, projected to $SO(3)$, on all training paths of length at most four legal in both graphs, and scores the fitted connection on both families. Over 200 draws, one-way/return finite-query regret is $0.0277/0.0565$ at $H = 4$, $0.0298/0.1244$ at 8, $0.0318/0.1806$ at 12, $0.0319/0.2195$ at 16, and

$0.0328/0.2663$ at 20; fitted slopes are $0.099/0.953$. A maximum taken per fit differs from the theorem's supremum of expected regret.

Separation depends on dimension and aggregation. Across 216 additional fits, its frequency falls from 86% ($n = 16$) to 16% ($n = 4096$) for query maxima, versus $89/80/82/84/82\%$ for means at $n = 16, 64, 256, 1024, 4096$. At $r = 16/32$, one-way slopes are $.28/.31$ and return slopes $1.10/1.13$ through $H = 48$; 2,000 draws at $r = 3$ give $.110/.944$. These results support low-dimensional separation, not a universal exponent.

H.4 Supplied relations and matched probe allocation

Scalar calibration with a shared dictionary. Scalar calibration compares (14) with nonlinear estimators under identical observations and dictionary access, without vector-generator queries. The supplied $SO(3)$ dictionary (41) has angles $4\pi/7, 6\pi/7$. Its 64 type-balanced systems use Haar frames and true angles $\theta_j + \delta$, $\delta \in \{0, .001, .01\}$. The word span has dimension 18: 17 nonempty basis words of length at most six, measured with $c = e_1$ on three basis facts, and a known identity row. Paired sufficient-statistic simulations use Gaussian mean variance $.04^2/n$, $n = 16, 64, 256$; no physical queries are executed.

The estimators are the direct table, joint type/frame least squares, order-only fitting with a free joint angle, and free $SO(3)$ fitting. Joint fitting uses ten starts, including data-derived reconstructions, with 160 evaluations per start. Each relaxation starts from its constrained solution, uses 240 evaluations, and retains no greater training loss. All 2,304 fits are included; convergence does not establish global optimality.

Evaluation fixes $H = 8, 32, 128, 512, 2048$ and 140 queries before collection: alternating $(ab)^{H/2}$, separated $a^{H/2}b^{H/2}$, repeated/truncated motifs (a, b, b, a, b, a) and (a, a, b, a, b, b, b, a), and 24 uniform binary words per horizon. For unit true task row u and estimated row v, exact mean sign-decision regret over sphere facts is

$$\mathcal{R}(u, v) = \frac{1 - u^\top v/\|v\|}{4} \quad (v \neq 0), \qquad \mathcal{R}(u, 0) = \frac{1}{4},$$

This normalizes decision direction, not table output. Reported risks average finite queries; pointwise intervals use 4,000 paired-system bootstrap draws.

Query structure. $A^2 = I$ and $B^3 = I$ hold in truth at every tested deviation and in both relation-constrained fits. With $T_{pq} = T_q T_p$,

$$T_{w_1} = ABAB^2A = (AB)A(AB)^{-1}, \qquad T_{w_1}^2 = I, \qquad T_{w_2} = AB^3ABA^2 = B,$$

and at every tested $H = 8 \cdot 4^j$ the separated query equals B and the truncated motifs equal B^2AB^2A and B. These long queries are thus finite-order controls, not tests of mixed-phase accumulation; free $SO(3)$ fits need not preserve the reductions. Alternating queries expose the joint relation, whose fitted phase error can accumulate even under periodic truth; random words give a different mixed-response test.

Correct relations protect alternating-query decisions, whereas misspecification reverses the benefit (Table 3). Joint fitting is more accurate than the constructive table: their 36-word held-out RMSEs are .001104 and .007548, versus .001211 for order-only and .002282 for free fitting. At $n = 16, 64, 256$, alternating risks are $(25.52, 6.40, 1.60)10^{-6}$ for the table, $(1.734, .433, .108)10^{-6}$ for joint fitting, $(.1402, .0649, .0195)$ for order-only and $(.1260, .0740, .0249)$ for free fitting. At $n = 64$, table/joint intervals are $[5.11, 7.82]10^{-6}/[2.65, 6.36]10^{-7}$ under correct relations and $[.23899, .30195]/[.23704, .29950]$ at $\delta = .01$.

The basis has minimum singular value .2233 and condition number 16.147. At $n = 64$, acquisition costs 3,264 scalars and 10,944 action letters; float64 storage is 432 bytes for the table and 20,160 for 140 query coefficients. Single-thread CPU medians are 1.086 ms for basis construction, 86.902 ms for coefficients and 1.032 μs for applying all rows. The latter excludes preparation and has no equally precompiled fitted-model comparator.

Relation fit and misspecification. A separate vector-generator comparison uses unequal priors: joint fitting knows both dictionary types, polar fitting only orthogonality, and order-only fitting additionally knows angles $\pi, 2\pi/3$. Its 64 CPU trials balance types, with QR-orthogonal

Table 3: Same-data scalar calibration: mean sign-decision regret at $H = 2048$, $n = 64$, over all 64 systems. Every predeclared query family is retained. δ perturbs the true relation while the supplied dictionary stays fixed.

δ	Query family	Direct table	Joint fit	Order-only	Free $SO(3)$
0	Alternating	$6.40\,10^{-6}$	$4.33\,10^{-7}$	.06492	.07396
	Separated	$6.57\,10^{-6}$	$6.67\,10^{-7}$	$6.81\,10^{-7}$	.15793
	Motifs	$3.64\,10^{-5}$	$5.81\,10^{-7}$	$7.50\,10^{-7}$	.04556
	Random	$5.77\,10^{-5}$	$6.12\,10^{-7}$	$5.54\,10^{-5}$	.00086
.001	Alternating	.07608	.07634	.06492	.07397
	Separated	$6.57\,10^{-6}$	$7.09\,10^{-7}$	$6.81\,10^{-7}$	.15795
	Motifs	$3.73\,10^{-5}$	$8.27\,10^{-7}$	$7.50\,10^{-7}$	.04553
	Random	.000119	.000061	.000055	.000856
.01	Alternating	.27044	.26835	.06482	.07396
	Separated	$6.56\,10^{-6}$	$2.18\,10^{-6}$	$6.81\,10^{-7}$	.15807
	Motifs	$7.20\,10^{-5}$	$2.11\,10^{-5}$	$7.47\,10^{-7}$	.04525
	Random	.00613	.00604	.00006	.00087

frames, $\sigma = .04$, $n = 64$, and queries $(ba)^{H/2}$ plus eight fixed-stream random words on 256 sphere facts. Under correct relations, joint/independent/order-only regret at $H = 3584$ is $1.58 \times 10^{-6}/.03027/.01956$; all types are recovered ($\gamma = .5770$, $\tau = .5505$). With .01 axis-angle misspecification, the corresponding risks at $H = 896$ are $.02580/.01899/.01329$, and joint minus order-only is $.01051[.00724, .01393]$ (4,000 paired-trial resamples).

Four-vertex scalar protocol. The root is 0; edges $0, \dots, 7$ are $(0,1)$, $(1,0)$, $(0,0)$, $(1,2)$, $(2,3)$, $(3,2)$, $(2,2)$ and a second $(1,2)$. Each component's internal canonical maps are (I, A, B_{θ_C}), with $\theta_C \in \{4\pi/7, 6\pi/7\} + \delta_C$. Four Haar $SO(3)$ frames and two Haar crossing rotations are independent and unknown; the graph, individual orders and two angle neighborhoods are supplied. Observations are $c^\top T_w e_j + N(0, 0.04^2)$, $c = e_1$, on three calibrated basis facts; intermediate states are unobserved.

The short-versus-return comparison uses 32 systems balanced over the four base-type pairs, at deviations $(0,0)$, $(0.01, 0.01)$ and independent $\mathrm{Unif}[-0.02, 0.02]$. Lexicographic shortest root paths p_x and all legal suffixes through length three give 74 distinct nonempty words $p_x, p_x s, p_x as$ (longest seven), with 64 repeats per fact. The return design uses 32 repeats plus eight words $u_C v_C^k s$, where $(u_1, v_1) = ((), (0,1,2))$, $(u_2, v_2) = ((0,3), (4,5,6))$, $k = 8, 16$ and s is empty or the local B edge, each with 34 repeats and length at most 51. The designs use 14,208 versus 7,920 scalars and 61,632 versus 61,416 action letters: with one reset per observation, the episode ratio is 0.557 but the action ratio 0.996. Wall time and state-dependent reset cost are unmatched.

Fits minimize repeat-weighted scalar residuals: fixed relations search four base pairs; learned relations add two deviations in $[-.04, .04]$; free fitting estimates eight $SO(3)$ edges from root-reconstruction and learned-fit starts, selecting lower training loss. All use analytic Jacobians, 100 evaluations per start and 10^{-10} tolerance. Root reconstructions introduce no extra observations.

At $H = 6k + 2$, $k = 1, 8, 64, 512$, each system is evaluated on 256 fresh sphere facts and five legal words that repeat $(0,1,2)$ or $(2,0,1)$, cross via $(0,3)$ or $(0,7)$, then repeat $(4,5,6)$ or $(6,4,5)$; the first word has fixed order, the others randomize it per repetition. Regret is averaged, not maximized. Learned short/return regrets at $H = 3074$ are $0.013308/0.001001$, $0.012645/0.001086$ and $0.009266/0.001056$ in the three conditions. Free fitting also benefits; the fixed fit excels only with a correct relation.

Length- and exposure-matched ordering. Figure 6 uses 32 independent systems, eight per base pair, with shared short words and repeats. The control uniformly permutes each added word's k two-edge excursions and k self-loops, retaining access and exit segments. Legality, endpoints, length, edge counts, 82 words, 7,920 scalars, 61,416 actions and resets are matched, as are priors, fitting

Table 4: Identical fitted response and block counts, different query order. Mean sign-decision regret at $H = 3074$; paired difference is grouped minus interleaved. All 32 systems per condition are retained.

Condition	Probe design	Interleaved	Grouped	Difference [95% interval]
Exact	Return	.000018	.002011	.001993 [.001357,.002705]
	Random	.000187	.019289	.019102 [.010733,.029506]
Shared .01	Return	.000017	.001718	.001701 [.001159,.002266]
	Random	.000185	.019691	.019506 [.010150,.031539]
Independent	Return	.000019	.001988	.001968 [.001350,.002675]
	Random	.000192	.020241	.020050 [.011090,.031564]

and paired noise. Both designs are fixed before collection. The control permits returns and is not uniform over all words.

Query order with identical block counts. To isolate order, all 192 system/condition/design models are refitted on the same calibration data, and their complete edge matrices are held fixed across paired queries; query outcomes never enter fitting or selection.

Let $L_0 = (0, 1, 2)$, $L_1 = (2, 0, 1)$, $R_0 = (4, 5, 6)$ and $R_1 = (6, 4, 5)$. For each $k = 8, 64, 512$, draw 32 left and right sequences of k independent equiprobable local blocks. Each interleaved query crosses via $(0, 3)$ or $(0, 7)$; its grouped pair sorts the same blocks by type, preserving mixed-block and edge counts, length $6k + 2$ and endpoint. All 96 pairs per fit use exact sphere integration. Since unit task rows have identical marginal score distributions under uniform facts, margins cannot explain the order effect. Table 4 reports the longest horizon; pointwise intervals resample the 32 systems 4,000 times per condition and design, keeping pairs.

Grouping also raises risk for free fits: at shared deviation .01, risk changes from .000208 to .001667 with return probes and from .000278 to .021904 with random probes. Falsely fixed relations give .012339 to .260808 and .012315 to .260780. Thus return probes reduce estimation error but cannot repair misspecification; the order effect holds with the fitted matrices fixed.

Probe-allocation gains depend on query composition. At $H = 3074$, return-minus-random regret is $-.0093[-.0152, -.0047]$ across conditions, versus $-.00014[-.00021, -.00008]$ after removing the fixed-order word. The remaining four-word differences are $-.000003[-.000006, -.000001]$ at $H = 50$, $-.000023[-.000044, -.000008]$ at $H = 386$ and inconclusive at $H = 8$; resampling retains each system's conditions. Exploratory uniform rather than binomial block counts give $-.0318[-.0481, -.0186]$, but change the query law as well as grouping.

Periodicity does not invalidate the fixed-order query. For $C_\theta = B_\theta A$ in (41), $\operatorname{tr} C_\theta = 1 + 2\cos\theta$. At $\theta = \theta_0 + \delta$, $\theta_0 \in \{4\pi/7, 6\pi/7\}$,

$$\|C_\theta^7 - I\|_{\mathrm{op}} = 2|\sin(7\delta/2)|, \qquad \|C_\theta^{512} - C_\theta^8\|_{\mathrm{op}} = 2|\sin(252\delta)|,$$

with residuals .06999 and 1.16466 at $\delta = .01$. Periodic truth need not give periodic regret, because a fitted phase can drift: for $\theta_{\mathrm{true}} = 4\pi/7$, $\widehat{\theta} = 4\pi/7 + .001$ and $c = e_1$, exact single-return regrets at $k = 8, 64, 512$ are .000008013, .000511838, .032058408. Conversely, because $A^2 = I$, the loop-free control $(0, 1)^{\lfloor 3k/2 \rfloor} q_i (4, 5)^{\lfloor 3k/2 \rfloor}$ reduces to the crossing in truth and in relation-constrained fits at these even k; it does not probe mixed-return phase uncertainty, and free fits need not share the reduction.

All 192 trials and 1,920 candidate fits are retained; selected learned types are correct and fits converge, without a global-optimality certificate. At $H = 3074$, 3,000 paired-system resamples assess primary contrasts on 32 systems. Learned fits improve on 25/27/26 systems across conditions, raising decision agreement from $95.73/95.78/95.60\%$ to $98.53/98.65/98.57\%$. These repeated-return endpoints and exploratory query-family comparisons do not test visual transfer.

Prospective query-risk prediction. We next predict finite-query risk magnitude from calibration data and the planned word, rather than merely its ordering effect (Table 4; Figure 5, left).

Table 5: Matched calibration at $H = 3074$, all 32 systems per condition. Difference is random minus return; intervals are paired-system 95%.

Condition	Fit	Random	Return	Difference [interval]
Exact	Fixed	$1.75\,10^{-6}$	$1.99\,10^{-6}$	$-0.24[-0.91, 0.43]\,10^{-6}$
	Learned	.011024	.001148	.009876 [.005099,.015875]
	Free	.012120	.001210	.010910 [.005456,.016974]
Shared .01	Fixed	.037171	.037375	$-.000204[-.000429, .000027]$
	Learned	.010849	.000969	.009880 [.004808,.016526]
	Free	.013099	.001056	.012043 [.006347,.018959]
Independent	Fixed	.067162	.066940	.000222 [−.000099,.000585]
	Learned	.011625	.001311	.010314 [.005552,.016063]
	Free	.013358	.001320	.012038 [.006791,.018315]

Table 6: Query-risk prediction on twelve held-out systems. Every predictor uses one scale constant fitted on twelve development systems; intervals resample systems. Risk spans 5.2 orders of magnitude. Unscaled criterion predictions are evaluated separately in the text.

Predictor	Log RMSE [95% interval]	Median factor	95th percentile
Delta-method criterion	.2704 [.2090,.3353]	1.18	1.74
Word length	1.9240 [1.8783,1.9691]	3.04	57.9
Word length squared	2.7555 [2.6864,2.8217]	5.95	287.6

For the three-dimensional orthogonal responses in this study and fixed unit task direction c, let $u_w(p) = T_w(p)^\top c$ for parameters p. For facts uniform on the sphere, a fit v against truth u has mean sign-decision regret exactly $(1 - u^\top v)/4$, so for a parameter error with covariance Σ and mean b,

$$\mathbb{E}\ \mathrm{Regret}(w)\ \approx\ \tfrac{1}{8}\left[\mathrm{tr}\big(J_w \Sigma J_w^\top\big) + \|J_w b\|^2\right], \qquad J_w = \partial u_w/\partial p.$$

Analytic edge derivatives give J_w by the product rule; $\Sigma = (\sigma^2/n)(M^\top M)^+$ uses the exposure-weighted calibration Jacobian M. This local approximation can fail under type error, aliasing, substantial bias or long-word nonlinear effects; it is not a minimax bound.

The pseudoinverse retains Gram eigenvalues above 10^{-9} of the maximum (singular-value cutoff $\sqrt{10^{-9}}$). Probe exposure weights are $\sqrt{n_i/n}$, with σ^2/n applied once. Across 5,760 predictions, the maximum relative derivative in discarded modes is 5.2×10^{-15}, below tolerance 10^{-3}.

Predictions precede outcomes for 24 systems, three relation conditions, two probe designs and 40 queries (five per family at $H = 6k + 2$, $k = 1, 8, 64, 512$). One fit predicts with $b = 0$; 24 fresh calibration draws estimate risk. Twelve systems train each predictor's single log-scale constant and twelve test it (Table 6, Figure 5, middle and right). Unscaled predictions are evaluated separately.

Held-out risks span 9.3×10^{-7}–.1325. Unscaled prediction gives mean log measured/predicted ratio $-.0539[-.1103, -.0042]$ and log–log slope $.9863[.9760, .9970]$ (Table 6). Arithmetic calibration is less uniform: mean query ratio $.9808[.9374, 1.0244]$, ratio of means $.8640[.7720, .9560]$ and residual $-.000502[-.000808, -.000169]$, largest at $H = 3074$ (4,000 twelve-system resamples). At prospectively fixed noise .02/.08/.16, log ratios are $-.0491[-.1064, +.0017]/ -.0735[-.1278, -.0264]/ - .1341[-.1829, -.0913]$, slopes .9902/.9714/.9271, and log RMSE .270/.287/.410 versus length-only 1.937/1.878/1.757. Accuracy degrades as noise grows, consistent with a local approximation.

Figure 5 offsets conditions at left; at right, families are offset by $\pm.05$ decade with jitter at most .015. Magnitude prediction does not establish superior ranking. Among 2,880 paired queries, 576 have identical words; excluding them, grouping is uniformly riskier at $H = 386, 3074$, where family identity ranks perfectly. At $H = 50/8$, prediction's ranking gains are $+.0625[-.0208, +.1458]/ + .1528[-.0208, +.3056]$, neither separated from zero.

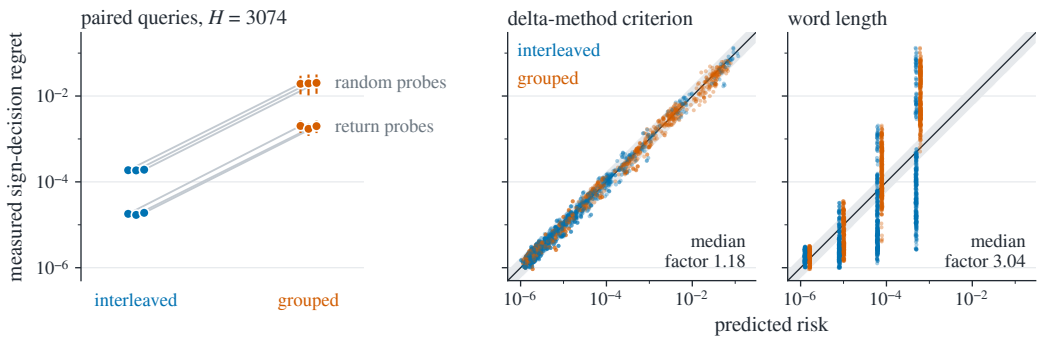


Figure 5: **Query order and prospective risk prediction.** Left: exploratory paired interleaved (blue) and grouped (orange) queries share fits and block counts. Whiskers: interleaved mean plus the pointwise 95% paired-system interval of the difference. Middle, right: separate prediction study, 24-draw mean risks for 2,880 held-out queries; each predictor uses one development-fitted scale constant. Line: exact agreement; band: factor 1.74, the criterion's 95th percentile. Middle: shared words appear once; right: families are offset.

Table 7: Independent 64-system information-design comparison at $H = 3074$. Pilot alone uses a smaller budget; all other designs share the augmented budget.

Design	Mean binary-task regret
Short pilot only	.020937
Uniform random	.013793
Generic D-opt	.002488
Recurrence-targeted D_s	.002032
Fixed returns	.002218

Information-design comparison. A common 74-word pilot with 32 repeats supplies the fitted local linearization for 20 unknown frame, crossing and phase parameters. SVD removes three gauge-null directions at relative threshold 10^{-9}, leaving rank 17; candidate derivatives and both phase coordinates are checked at 10^{-6}. Generic D-opt maximizes log-determinant information gain; targeted D_s maximizes log-determinant reduction of the two phases' marginal covariance, including nuisance uncertainty. These classical criteria are neither new optimal-design guarantees nor complete implementations of GST or task-optimal exploration (Nielsen et al., 2021; Wagenmaker et al., 2021).

On 64 systems disjoint from 32 development systems, random, D-opt, targeted D_s and fixed-return designs each select eight length-26 words from eight returns, six phase-insensitive crossings and up to 128 random legal paths. Greedy covariance updates use no new observations or relinearization. Each word receives 34 repeats per basis fact; including the pilot, every augmented design uses 7,920 scalars and 52,032 actions. Fits share type-specific starts and objective, retaining every system. Edge exposure and design/fitting time are not matched.

Targeted D_s outperforms random allocation ($-.011761[-.015493, -.008370]$) but is not separated from D-opt ($-.000456[-.001058, +.000088]$; 32 wins, 31 losses, one tie) or fixed returns ($-.000186[-.000560, +.000130]$). Intervals use 4,000 paired-system draws and are pointwise for secondary contrasts. All designs use pilot estimates and supplied type neighborhoods.

Two limits of local summaries. Take planar responses $R_{\theta_\pm}$ with $\theta_\pm = .05 \pm \pi/64$. Both calibration designs observe the same one-step row and four long rows on basis facts, with independent noise variance σ^2 per scalar. The long depths are $(64, 64, 64, 128)$ or $(32, 96, 96, 96)$; both have depth sum 320 and squared-depth sum 28,672, hence equal scalar counts, action costs and local Fisher information. Yet the first bank aliases the pair exactly, while the second has squared mean

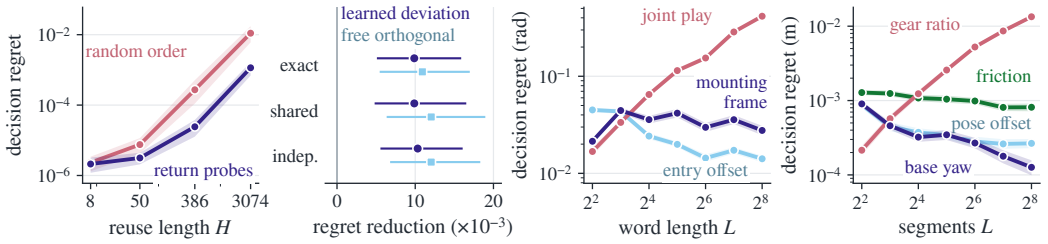


Figure 6: **Calibration design and unresolved uncertainty.** Left: resource-matched probes, 32 systems, five-query means. Center-left: random-minus-return regret at $H = 3074$. Center-right, right: non-adaptive prior averaging on an analytic wrist (152 libraries) and MuJoCo arm (Todorov et al., 2012) (128 scenes); within-panel visibility is matched at 0.30 and 0.010. Means and pointwise 95% bootstrap intervals cluster by system, library or scene. Axes are logarithmic except center-left; units differ. Results concern sampled paths, not all-path suprema (Appendices H.4, H.5).

separation 16. With one anchor and m repetitions of each long row, the total squared separations are

$$s_{\text{alias}} = 4\sin^2(\pi/64), \qquad s_{\text{sep}} = 4\sin^2(\pi/64) + 16m.$$

Under equal priors the Gaussian likelihood test has error $\Phi(-\sqrt{s}/(2\sigma))$, with Φ the standard normal distribution function, and a common fact with opposite scores $\pm\sin(H\pi/64)$ makes decision risk this error times $|\sin(H\pi/64)|$. At $H = 32$, $\sigma = .05$ and $m = 1$, the aliased risk is .1632 and the separating risk below 10^{-300}. Repeating aliased long probes cannot improve the fixed anchor's discrimination; repeating the anchor can. The example separates local precision from global discrimination, not information-based design from all alternatives.

A second example shows why significance in (6) includes exits. With $c = e_1$ and internal response $R_x(t)$, every current row is $c^\top R_x(kt) = c^\top$; after a legal exit $R_z(\beta)$ it becomes

$$c^\top R_z(\beta) R_x(kt) = (\cos\beta, -\sin\beta\cos kt, \sin\beta\sin kt).$$

Parameters separated by Δ then have row distance $2|\sin\beta\sin(k\Delta/2)|$ and, by the unit-row risk formula above, sphere-averaged sign-decision regret $\frac{1}{2}\sin^2\beta\sin^2(k\Delta/2)$. This vanishes without the exit but generally not after it: current invisibility does not imply irrelevance to legal continuations. These analytic examples make no claim about invisible subspaces in trained visual models.

Sensitivity to high-loss fits. Allocation gains are sensitive to high-loss fits. In a separate 64-system sweep, removing the approximately 3% above three times their design/condition median training cost changes the exact-relation regret ratio at $H = 1538/3074$ from $2.44/2.88$ to $17.13/13.44$; unfiltered intervals are $[.87, 85.06]/[1.28, 30.33]$. The sweep uses only $k = 8, 16$ probes and supports neither a universal crossover nor optimal allocation.

H.5 Deployment without calibration

A never-calibrating agent minimizes expected cost under a fixed prior. With m draws this is the m-sample rule; larger m refines the cost estimate, not calibration. Its regret prices this non-adaptive rule, not randomization-trained policies that adapt from deployment history.

The analytic wrist uses signed rotation primitives about unit axes $\{x, y, z, (1,1,0), (1,1,1)\}$ by angles $\pm 2\pi/n$, $n = 2, \ldots, 12$. With command increment u_t, commanded position q_t, loaded position y_t and coordinate-wise play p,

$$q_{t+1} = q_t + u_t, \qquad y_{t+1} = \text{clip}(y_t, q_{t+1} - p, q_{t+1} + p), \qquad R_{t+1} = \exp([y_{t+1} - y_t]_\times) R_t,$$

where $q_0 = y_0 = 0$, $R_0 = I$ and $[v]_\times$ is the cross-product matrix of v. An unknown mounting frame Q gives world orientation $QR_LQ^\top$. Candidate words share commanded joint totals, not necessarily nominal rotations; the goal is world-fixed, so mounting uncertainty does not cancel.

Play is steepest in 149 of 152 libraries. Recurring-play slopes are $+.827, +.710, +.651$ under prior averaging, a nominal point estimate and one sampled world; the mounting-frame arms stay flat (Figures 8 and 7).

Table 8: Uncalibrated deployment on the analytic wrist, 152 action libraries, matched short-word visibility 0.30. Slope is the log–log growth of mean regret against word length over $L \geq 16$; intervals are library-clustered bootstraps.

Unknown	L=16	L=64	L=256	Slope [95% CI]
Mounting frame (one gauge)	.0358	.0298	.0277	-0.104 $[-0.150, -0.058]$
Entry offset (applied once)	.0242	.0144	.0141	-0.129 $[-0.179, -0.077]$
Per-joint play (recurring)	.0649	.1543	.4156	$+0.725$ $[+0.697, +0.755]$

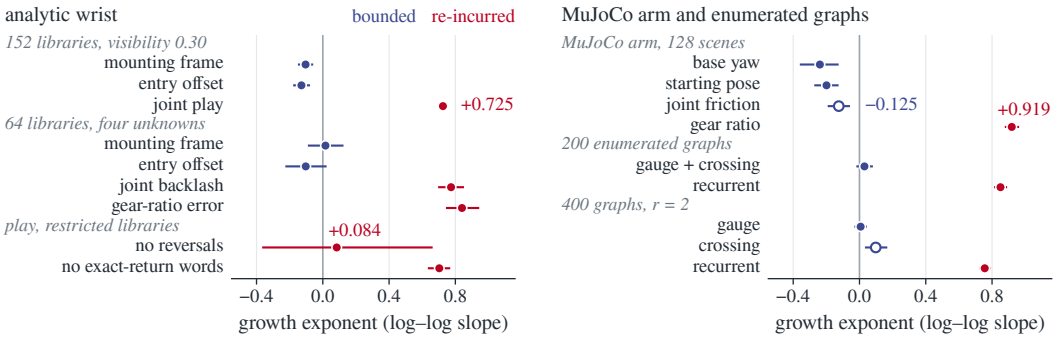


Figure 7: **Finite-range regret growth.** Markers: log–log slopes of mean uncalibrated regret against length, not theorem-predicted exponents. Whiskers: pointwise 95% library/scene/graph-bootstrap intervals, except the 400-graph intervals whose level and resampling unit are unspecified; the 152-library play interval is hidden by its marker. Hollow markers flag post hoc bounded classification of friction (predicted to grow) and nonzero-regret selection for crossings ($r = 2$). Colors denote full-library classes; removing reversals prevents play re-engagement, with a wide slope interval.

Matched short-path visibility. Prior widths are bisected on one fixed probe set to match RMS response divergence on calibration words of length at most four. This controls short-word visibility; otherwise perturbation magnitude alone could drive the comparison.

Short-horizon cost reverses the long-horizon ordering: at $L = 4$, mounting/entry/play regret is $.0214/.0450/.0168$. Matching visibility within 1% strengthens separation (mounting slope $-.104 \to -.142$, play unchanged at $+.725$). At visibility $.15/.60$, play slopes remain positive ($+.803/+.575$); at .60, regret reaches 38.9% of its range and begins saturating.

Finite-sample prior averaging. Increasing draws from 8 to 4096 yields log-regret/log-m slopes $-.106[-.242, +.015]$, $+.029[-.054, +.120]$ and $-.020[-.053, +.013]$ (Figure 8). No gain is detected, but shared randomness and the absence of an integration reference preclude a convergence claim.

Memoryless and hysteretic uncertainty. Across 64 visibility-matched libraries (0.300), backlash and memoryless gear-ratio error give nearly identical $L = 256$ regret, .4289 and .4285, with slopes $+.774[+.696, +.852]$ and $+.840[+.743, +.944]$. Mounting-frame and entry-offset regret is $.0395/.0191$, with slopes $+.016[-.090, +.125]/-.103[-.226, +.023]$. Recurring error, rather than hysteresis alone, distinguishes these arms.

Graph-based replication. On 200 graphs, labels and unknown placement fixed before measurement separate slopes $+0.030$ $[-0.019, +0.081]$ and $+0.851$ $[+0.812, +0.891]$ across vertex gauges and continuous crossing and recurrent responses. The labels group measured exponents; Theorem 4.2, which also depends on model class and task, does not predict them. Recurrent slopes at $SO(5)$ and $SO(8)$ are $+0.975$ and $+0.672$, but matched gauge and crossing cells have zero regret there, so only $r = 2, 3$ allow two-sided slope comparisons.

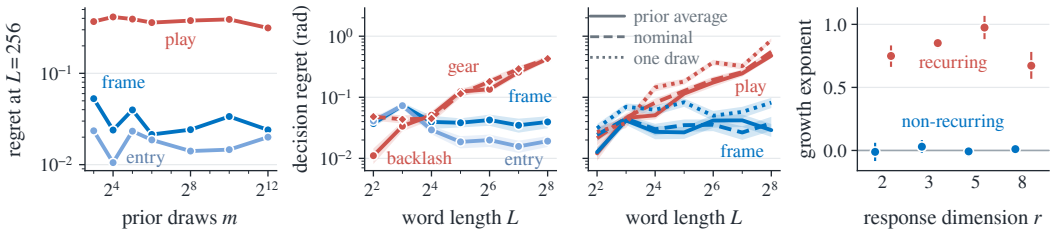


Figure 8: **Controls on recurrence-dependent regret.** Left: prior draws, 8–4096; no detectable gain does not certify convergence, and libraries share one evaluation seed. Center-left: gear-ratio error versus backlash at equal visibility. Center-right: line styles mark prior averaging, nominal estimation and one sampled world; colors mark frame uncertainty and recurring play. Right: graph-based labels fixed before measurement group finite-range exponents across dimensions, not values predicted by Theorem 4.2.

On 400 graphs at $r = 2$, matched recurrent and gauge slopes are $+0.756$ $[+0.722, +0.790]$ and $+0.007$ $[-0.031, +0.045]$; recurrence is steepest in 294 of 314 graphs with all slopes defined. Abelian conjugation does not hide open-path gauge error: with $T \equiv I$, $q_x = I$, $q_y = R_{\pi/2}$, root loops stay at I but $x \to y$ becomes $R_{\pi/2}$, and gauge regret at $L = 16$ is .0446. The additional $SO(2)$ gauge arm in Figure 8 is likewise flat but visible on open paths, with slope $+0.006$ $[-0.121, +0.134]$. Visibility matching succeeds in 62% of gauge/crossing cells at $r = 2$ and 92% at $r = 3$. Crossings occur once per word (step fraction $1/L$) and reach the common visibility target in only a third of graphs; nonzero crossing regret has slope $+0.098$ $[+0.033, +0.168]$.

Rigid-body dynamics. A four-joint MuJoCo arm supplies integrator, actuator dynamics, inertia coupling, joint limits and dry friction; only the uncertain parameter is specified, and co-terminal words command the same nominal configuration. At visibility 0.010 over 128 scenes, $L = 256$ regret (slope) is .00013 (-0.239 $[-0.360, -0.126]$) for base yaw, .00027 (-0.199 $[-0.273, -0.126]$) for starting-pose offset, .00081 (-0.125 $[-0.192, -0.057]$) for joint friction and .01339 ($+0.919$ $[+0.877, +0.963]$) for gear ratio; gear is steepest in 118 of the 125 scenes with all four slopes finite.

Gauge cancellation requires gravity-preserving yaw without floor contact: yaw preserves responses to 1.3×10^{-16}, whereas tilt gives 4.8 mm readout error and contact alters five of twelve words by up to .2 radians. Friction's flat response contradicts the preregistered growth prediction; a thirtyfold deadband sweep (f/k_p) changes the $L = 64$ versus 4 regret ratio only from 1.2 to 2.4.

Restricting direction reversals. Removing reversals reduces $L = 256$ backlash regret from .4708 to .0078, with slope $+.084[-.366, +.663]$, without affecting mounting error. Removing zero-net-command words does not help ($+.703[+.634, +.769]$). Play re-engages at reversals: the slope changes from near zero to near $+.7$, rather than increasing monotonically with reversal rate; regret magnitude varies hundredfold (Figure 9).

Selection and capability loss. These regrets are scored inside the given library, so they price selection error, not lost capability. For restricted and full libraries $\mathcal{B} \subseteq \mathcal{A}$,

$$C_T(\widehat{p}_{\mathcal{B}}) - \min_{p \in \mathcal{A}} C_T(p) = \underbrace{C_T(\widehat{p}_{\mathcal{B}}) - \min_{p \in \mathcal{B}} C_T(p)}_{\text{selection error inside } \mathcal{B}} + \underbrace{\min_{p \in \mathcal{B}} C_T(p) - \min_{p \in \mathcal{A}} C_T(p)}_{\text{capability given up}}.$$

The sweep above measures only the first term; a separate experiment measures both.

Common-target evaluation. To measure lost capability, all strategies use targets from one full nominal library, including targets outside the restricted library. The full pool contains the restricted pool, ensuring a nonnegative capability gap. We test 64 libraries, four seeds, lengths $8, 32, 128$ and three noise levels. Calibration fits play magnitude by least squares on geodesic residuals of N length-six words; intervals resample libraries.

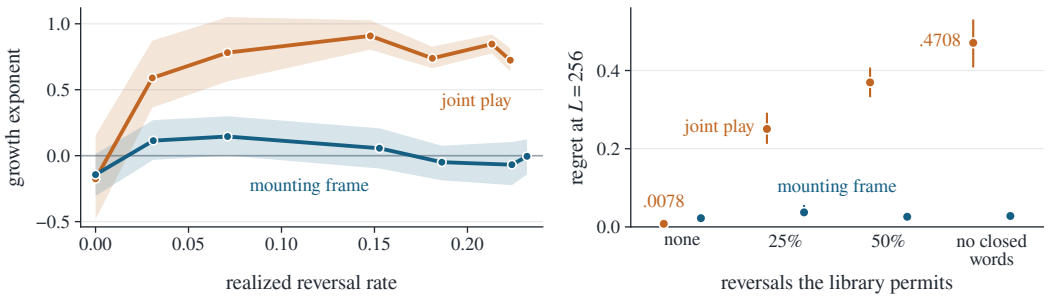


Figure 9: **Reversal restrictions and wrist regret.** Left: measured growth exponent versus realized reversal rate. Right: the same backlash under four libraries; intervals narrower than a marker are hidden by it. Means and library-clustered bootstrap intervals measure selection error against each library's own optimum, not capability lost by restricting the library (Table 9). Removing reversals, unlike forbidding exact-return words, prevents play re-engagement.

Table 9: Selection error and capability loss, in radians. Total regret uses the full-library optimum; restricted selection error uses its own optimum. All strategies are uncalibrated, so values are independent of observation noise. Intervals resample libraries.

L	capability given up	restricted, no calibration	its selection error	full, no calibration
8	.4531 [.3664, .5486]	.4646 [.3784, .5607]	.0115 [.0065, .0173]	.0197 [.0130, .0269]
32	.2039 [.1692, .2399]	.2153 [.1790, .2528]	.0114 [.0063, .0176]	.0530 [.0369, .0708]
128	.1605 [.1323, .1901]	.1733 [.1443, .2038]	.0128 [.0072, .0190]	.1092 [.0828, .1347]

Calibration versus restriction. Restricting the library reduces selection error but sacrifices twelve to forty times as much capability (Table 9). Calibration beats restriction at all tested lengths and noise levels with eight words (Figure 10). The smallest paired winning budget is one word in six of nine settings, two at noise 0.10, length 128 ($-.0607[-.1022, -.0169]$) and noise 0.30, length 32 ($-.0519[-.0915, -.0133]$), and eight at noise 0.30, length 128 ($-.0459[-.0851, -.0055]$). At noise .02, length 8, one word gives $-.4538[-.5489, -.3674]$; at noise .30, length 128, 64 words give $-.1435[-.1756, -.1120]$.

Low-budget calibration. Noisy, low-budget fitting can be worse than no calibration: one-word least-squares excess cost is $+.0309[+.0012, +.0620]$ at noise .10, length 32; $+.1259[+.0721, +.1864]$ at .10, length 128; and $+.3980[+.3118, +.4784]$ at .30, length 128. In the last setting restriction also beats four-word calibration by $.0859[.0350, .1404]$, but loses at eight words. Restriction can therefore help under limited calibration, without replacing it generally.

Retaining the prior. A third selector reweights the same 256 prior draws by the same calibration words instead of committing to a least-squares estimate, using self-normalized importance weights with log likelihood $-d^2/(2\sigma^2)$ for geodesic residual d (a near-identity Gaussian-tangent approximation). Thirty-two libraries, four seeds, three lengths and three noise levels give 54 cells; intervals resample libraries.

No posterior-minus-uncalibrated interval among 54 cells is strictly positive; least squares is worse in eleven cells, all with noise $\geq .10$ and at most four words (Table 11, Figure 11). Posterior averaging is worse than least squares in twelve cells with ESS $\leq 7.2/256$, but not in 24 cells with ESS ≥ 20. ESS diagnoses neither causation nor posterior accuracy (Elvira et al., 2022); particle convergence is untested. Exact conditioning reduces prior-averaged risk, not every parameter's risk or necessarily finite-importance-sample risk.

At noise .30, length 128 and one word, posterior averaging beats least squares by $-.4255[-.5416, -.3151]$; it also beats restriction at every tested length/noise. This benefit assumes the supplied three-dimensional prior. Prior misspecification and general task-aware library design

Table 10: Full-library regret after N executed calibration words, in radians, against the full-library optimum. Intervals resample libraries. Noise-independent restricted counterparts appear in Table 9; paired crossover budgets, including unplotted budgets, are reported below.

noise	L	$N = 1$	$N = 4$	$N = 16$	$N = 64$
0.02	8	.0108 [.0045, .0187]	.0012 [.0000, .0033]	.0001 [.0000, .0001]	.0000 [.0000, .0000]
0.02	32	.0236 [.0098, .0422]	.0003 [.0000, .0007]	.0001 [.0000, .0003]	.0000 [.0000, .0000]
0.02	128	.0468 [.0285, .0700]	.0023 [.0008, .0044]	.0007 [.0001, .0017]	.0003 [.0000, .0009]
0.10	8	.0323 [.0212, .0448]	.0085 [.0042, .0137]	.0008 [.0002, .0018]	.0002 [.0000, .0004]
0.10	32	.0839 [.0585, .1120]	.0105 [.0057, .0168]	.0033 [.0011, .0061]	.0007 [.0002, .0017]
0.10	128	.2352 [.1807, .2953]	.0374 [.0244, .0533]	.0099 [.0056, .0151]	.0021 [.0007, .0040]
0.30	8	.1165 [.0874, .1505]	.0612 [.0453, .0781]	.0139 [.0087, .0196]	.0008 [.0003, .0016]
0.30	32	.2344 [.1853, .2861]	.0990 [.0737, .1250]	.0253 [.0171, .0341]	.0050 [.0023, .0085]
0.30	128	.5072 [.4234, .5882]	.2592 [.2078, .3137]	.0830 [.0625, .1037]	.0298 [.0187, .0425]

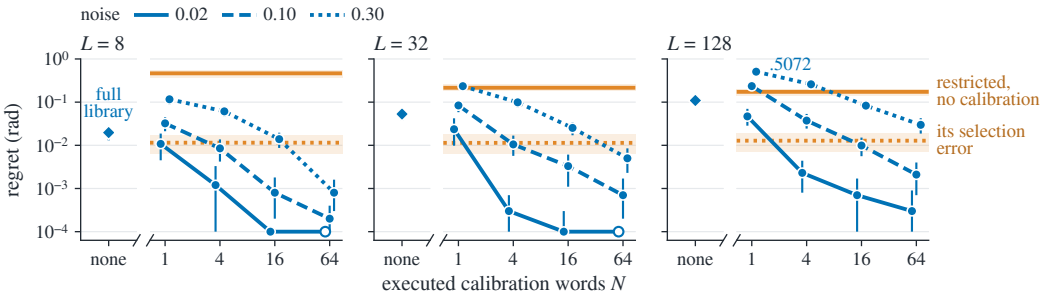


Figure 10: **Calibration versus library restriction.** Regret in radians; facets fix length L. Warm: uncalibrated restricted-library regret (solid) and within-library selection error (dotted). Blue: full-library regret before calibration (diamond) and after N words at noise $0.02/0.10/0.30$ (solid/dashed/dotted), offset along N. Except for selection error, all use the full-library optimum. Bands/whiskers: pointwise library-resampled intervals, sometimes hidden by markers. Hollow circles and clipped bounds at 10^{-4} denote values rounded to .0000. The eight-word crossover is tested between plotted budgets.

are untested; without reversals, co-terminal words have identical realized orientations and restricted selection is degenerate.

Finite-range growth versus learnability. Theorem 4.2 concerns worst-model, worst-query risk as training grows, not a non-adaptive selector's horizon slope. For one self-loop with $c = e_1$, $e = e_2$ and class $\{R_\delta, R_{-\delta}\}$, true scores are $\mp \sin(k\delta)$ and the indifferent prior average has regret $|\sin(k\delta)|/2$. For $0 < \delta < \pi/(2H)$ this rises to $\sin(H\delta)/2 < 1/2$, approaching $1/2$ only as $H\delta \to \pi/2$, with a fitted slope that depends on δ and the horizon range. Yet one length-one probe separates the models by $2 \sin \delta$, so risk decays exponentially in its repeat count, satisfying both branches of Theorem 4.2. Growth does not imply nonlearnability, nor sampled flatness learnability.

Scope. Finite-range growth is not uniformly greater cost: on the enumerated graphs, the recurring arm remains below the bounded arm at $L = 256$. Evidence comprises analytic rollouts and a contact-free rigid-body simulator with a fixed co-terminal task, not robot policies. Separation weakens at high visibility/prior width as regret saturates. Three preregistered predictions fail: growth normalized at $L = 4$, large recurring-arm cost at short horizons, and declining bounded-arm cost with more prior draws.

H.6 Decision-critical composition and visual replication

Each three-action block permutes one action multiset: identical nominal endpoints, different non-commuting response products. Structured transport learns one matrix per action; endpoint, GRU and

Table 11: Prior averaging versus least squares on matched libraries, targets, prior draws and calibration words. Differences (radians) are after one word; positive means worse than never calibrating. ESS is mean effective sample size out of 256 draws at that budget. The smallest winning N requires the library-resampled interval to lie strictly below the uncalibrated reference.

		never	difference at $N{=}1$			smallest winning N	
noise	L	calibrates	least squares	posterior	ESS	squares	posterior
0.02	8	.0149	$-.0053\,[-.0151, +.0048]$	$-.0111\,[-.0209, -.0012]$	8	2	1
0.02	32	.0629	$-.0438\,[-.0750, -.0159]$	$-.0386\,[-.0676, -.0103]$	8	1	1
0.02	128	.1123	$-.0586\,[-.1005, -.0214]$	$-.0651\,[-.1058, -.0283]$	9	1	1
0.10	8	.0149	$+.0140\,[+.0023, +.0267]$	$-.0027\,[-.0109, +.0056]$	87	4	4
0.10	32	.0629	$+.0067\,[-.0285, +.0422]$	$-.0285\,[-.0537, -.0071]$	90	4	1
0.10	128	.1123	$+.0773\,[+.0034, +.1642]$	$-.0533\,[-.0974, -.0112]$	92	4	1
0.30	8	.0149	$+.1062\,[+.0712, +.1428]$	$+.0021\,[-.0028, +.0073]$	192	64	8
0.30	32	.0629	$+.1861\,[+.1303, +.2423]$	$-.0181\,[-.0432, +.0032]$	192	64	4
0.30	128	.1123	$+.4197\,[+.3030, +.5427]$	$-.0058\,[-.0264, +.0121]$	199	64	2

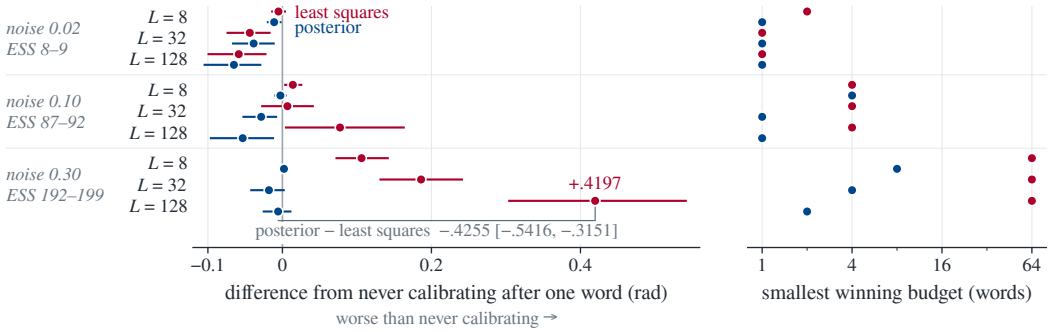


Figure 11: **Prior averaging versus a point fit.** Matched libraries, 256 prior draws and calibration words. Upper/lower marks: least squares/importance-weighted posterior averaging. Left: one-word regret minus uncalibrated regret (radians; positive is worse). Whiskers: pointwise library-resampled intervals, sometimes hidden by markers; bracket: paired posterior-minus-least-squares contrast at noise 0.30, $L = 128$. Right: smallest tested budget with an interval strictly below the uncalibrated reference. Group labels show the range of mean effective sample size (ESS) after one word.

evaluation-only path shuffle share supervision. Training uses one and two blocks, testing four, eight and twelve, with 512 independent roots, 64 candidates and aperture 0.14. The structured hypothesis class contains the true generator.

The visual model trains and freezes a convolutional encoder/decoder, state readout and factual observer, then learns a bounded residual nominal transition over 36 steps; correction trains on one and two blocks only. Nominal-model selection excludes unstable rollouts and representation-scale drift before correction evaluation. The three-seed diagnostics are a subset of the twelve-seed study, not an independent replication.

Structured transport succeeds in every cell across twelve seeds; attention, pooling and recurrent baselines fail to maintain this extrapolation (Table 12, Figure 12). All 720 root-clustered comparisons are positive, with minimum pointwise 95% lower bound $+21.48$ points. For endpoint/GRU/shuffle alone, the minimum bounds are $72.27/28.32/71.48$ points in analytic state and $72.85/22.27/72.46$ in the learned representation. Four-block GRU success ranges from 18.16 to 73.83% across seeds. A separate visual re-evaluation gives twelve-block endpoint/GRU/shuffle means $11.34/10.30/11.15\%$; it is distinct from the table's evaluation.

Attention reads innovation and action tokens using two pre-norm causal layers, four heads, four-times feed-forward width and last-action readout. Learned-position variants change initialization/parameterization; sinusoidal, rotary, ALiBi and position-free variants avoid untrained test-

Table 12: Twelve-seed mean decision success (%). Attention geometry and supervision are fixed across positional schemes; pooling is order-invariant.

Method	Parameters	4 blocks	8 blocks	12 blocks
Analytic state				
Structured	366	100.00	100.00	100.00
Attention, learned (variant A)	81,603	22.87	19.92	10.90
Attention, learned (variant B)	81,699	22.75	20.23	11.31
Attention, sinusoidal	57,123	22.28	18.42	10.94
Attention, rotary	57,123	20.65	16.16	12.01
Attention, ALiBi	57,123	25.21	18.65	10.97
Attention, none	57,123	24.04	18.13	11.10
Pooling	65,427	21.01	13.64	11.13
GRU	3,843	43.72	16.57	11.38
Endpoint	4,739	21.11	14.47	10.61
Path shuffled	—	20.90	13.44	11.26
Learned convolutional world model				
Structured	366	100.00	100.00	100.00
Attention, learned (variant A)	81,603	23.01	19.92	10.60
Attention, sinusoidal	57,123	22.71	18.08	11.91
Attention, rotary	57,123	20.28	16.28	11.28
Attention, ALiBi	57,123	24.37	19.04	10.60
Attention, none	57,123	23.91	18.68	10.79
Order-invariant pooling	65,427	20.75	13.31	10.86
Recurrent GRU	3,843	43.77	15.97	10.63
Endpoint editor	6,659	20.57	16.03	11.17
Path-shuffled	—	21.63	13.82	11.20

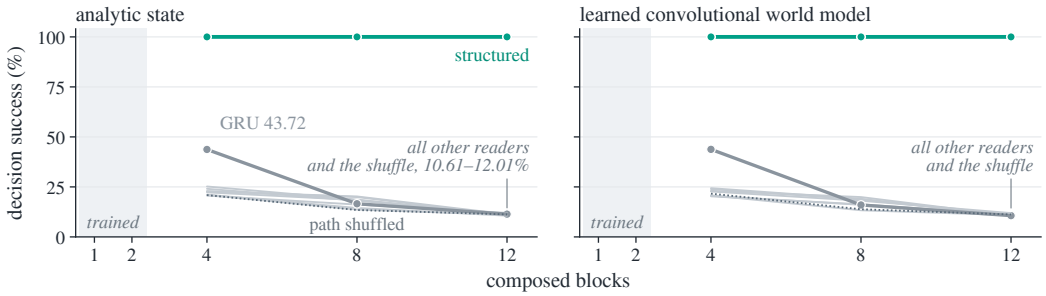


Figure 12: **Composition beyond training depth.** Twelve-seed mean decision success for analytic state (left) and a learned convolutional world model (right); correction trains on one and two blocks (shaded). Gray: GRU (labeled), fixed-configuration attention variants, pooling and endpoint; dotted: evaluation-only path shuffle. The structured class contains the true generator: this tests well-specified extrapolation, not limits of generic sequence models. Means only, without intervals; values in Table 12.

position embeddings. A 180-cell sweep (learning rates $10^{-3}, 3 \times 10^{-3}, 10^{-2}$; 4,000/16,000 steps; five seeds) attains twelve-block means up to 12.62% and individual cells up to 16.21%. Width, depth and head count remain fixed; this does not bound other sequence models.

In the three-seed subset, response RMSE is 0.00406–0.00680, operator relative error 0.70–1.96%, reconstruction MSE 0.001801–0.001802, observer MSE 0.000112–0.000139 and nominal rollout MSE 0.00173–0.00404. These are staged-representation diagnostics. Response-model timing excludes the pixel-to-action path measured separately below.

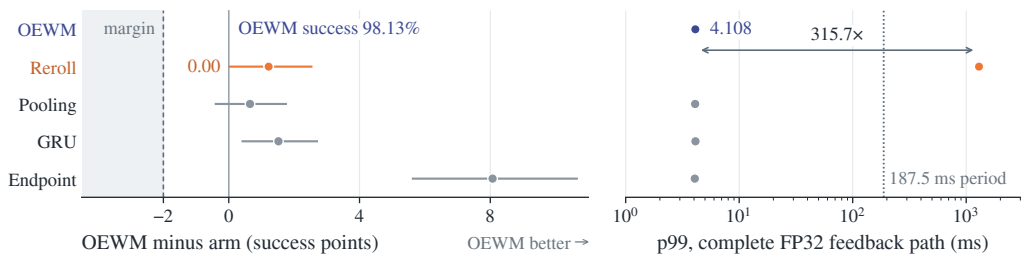


Figure 13: **Sixteen-chunk, success-trained comparison.** OEWM is success-noninferior to reroll at about 300-fold lower FP32 p99 latency, but is not separated from pooling. Left: pointwise crossed seed/root 95% intervals; dashed: predeclared margin. Right: complete-feedback-path p99; no intervals. This protocol differs from the six-block, return-trained comparison in Table 1.

H.7 Independent banks: quality, screening and combination transfer

The sixteen-chunk study uses Assembly-v3, persistent backlash, frozen DINO-WM and $K = 125$ candidates. Its $64/16/64$ train/validation/test roots have independent scenes, initial states and factual images but share nine nonidentity effects; no headroom filtering is applied. The 576 test cells represent 64 physical roots. Nominal/oracle success is $57.29/99.83\%$.

All arms train on simulator-fork success with fixed closeout. Full uses a path GRU with an innovation gate; pooling averages path features; endpoint reads terminal features; response GRU is recurrent; reroll additionally receives regenerated futures. Widths 256/304/380/520/444 match parameters within 1.3%, not FLOPs. BCE training uses AdamW (weight decay .01), batch 16, unit clipping and 60 epochs; validation success, then BCE, selects learning rate (3×10^{-4} or 10^{-3}) and five-epoch checkpoint. Normalization uses training roots.

Success and uncertainty. OEWM meets the predeclared -2-point noninferiority margin against reroll, but neither superiority nor separation from pooling (Tables 13–14; 10,000 crossed seed/root resamples retaining effects). Pooling has only 2.35 points of oracle headroom; 58 versus OEWM's 43 failures among 2,304 choices limit discrimination. A post hoc fixed-dispersion calculation (SE .565 points, two-sided .05) reaches 80% power near a 1.6-point gap. Post hoc pooling/GRU–reroll differences, $+.56[-1.00, +2.17]$ and $-.30[-1.69, +1.13]$, also meet the margin; endpoint's $-6.86[-9.68, -4.21]$ does not. These success endpoints differ from the return-based six-block study.

Table 13: Independent long banks. Crossed seed/root 95% intervals; Δ is OEWM minus the row in percentage points. p99 is the complete FP32 feedback path.

Method	Success (%)	Δ (pp)	p99 (ms)
OEWM	98.13 [97.18,99.00]	—	4.108
Pooling	97.48 [96.35,98.48]	0.65 [−0.43,1.78]	4.091
Endpoint	90.06 [87.15,92.80]	8.07 [5.60,10.68]	4.063
GRU	96.61 [95.40,97.70]	1.52 [0.39,2.73]	4.118
Reroll	96.92 [95.49,98.18]	1.22 [0.00,2.56]	1296.873

Success-trained readers perform poorly on success-truncated reward (Appendix H.1): regrets are .6832 for full, .6938 for pooling, .7251 for endpoint, .6799 for GRU and .6810 for reroll, all above random's .627. Success out-scores failure in only 220 of 575 mixed cells. Full-minus-reroll, pooling, endpoint and GRU differences are $.0022[-.0446, .0429]$, $-.0106[-.0417, .0196]$, $-.0419[-.0800, -.0097]$ and $.0033[-.0251, .0309]$. These reward comparisons do not determine success noninferiority.

End-to-end timing. Execution is fixed on 16 validation roots (144 cells): FP32 encoding/prediction, TF32 disabled, and regenerated projections rounded to the training targets' FP16 rep-

Table 14: Independent long-bank success (%) by training seed on the same 576 test cells. Columns do not represent independent test splits.

	Seed 1	Seed 2	Seed 3	Seed 4
OEWM	98.44	98.26	97.57	98.26
Pooling	97.40	97.74	97.57	97.22
Endpoint	92.71	90.28	87.15	90.10
GRU	96.53	96.18	97.22	96.53
Reroll	97.40	96.18	96.18	97.92

resentation before readout. Reroll retains its context and prepays action embeddings. Cached readers precompute fact-independent branch features; GRU also prepays zero-innovation recurrence. Full's feedback-dependent gate is not the matrix compiler of Appendix G.3.

RTX 5090 timing spans host images/proprioception to host action, including preprocessing, encoding, innovation, execution, readout, transfers and synchronization; acquisition, communication, loading and pre-feedback preparation are excluded. Five processes each measure 2,000 serial calls with rotated arm order, cycling roots/effects and four frozen seeds (first repeated), after 20 warmups plus two per root.

Full/reroll p50 is $4.051/1293.887$ ms and p95 $4.089/1296.290$ ms; process-wise p99 ranges are 4.094–$4.112/1295.707$–1297.164 ms. Cached arms meet 187.5 ms in all 10,000 calls, whereas FP32 reroll misses it in every call. Quality and timing use the same implementations.

Matched FP32 median/p99 speedups are $319.4/315.7$. Reduced-precision timings based on an unverified checkpoint, a different reader and component-level correction measurements are excluded from quality–latency comparisons. On 32 separate sixteen-chunk roots, 288 cells and four reward-MSE seeds, BF16/FP16 changes regret from .2491 by $-.0037[-.0087, +.0004]/+ .0013[-.0004, +.0034]$. This cohort differs in objective and is not paired with the complete-path timing above; its intervals do not establish equivalence.

Screening with complete-path regeneration. Screening uses 125 candidates, 64 three-block training roots and 48 six-block test roots under nine effects, optimizing normalized success-truncated reward. Four seeds train for 150 epochs (width 256, AdamW 3×10^{-4}, weight decay .01, batch 16, unit clipping). Full/pooling/endpoint receive three-frame innovation; reroll receives the last frame/proprioception. Parameter counts are 1,088,001/1,088,001/529,921/427,009. Test roots informed configuration and budget $k = 24$, so this is a conditional comparison.

Screens rank by full correction, pooling, endpoint, zero-innovation full correction, nominal cost or a fixed random order. The per-seed task-supervised reader scores the k regenerated candidates independently, without subset-dependent normalization or physical outcomes. Budgets $1, 24, 125$ include the correction-only and full-reroll endpoints. With true costs c_j, $D = \max\{\max_j c_j - \min_j c_j, 10^{-9}\}$ and selected subset S, total regret splits exactly into

$$\frac{c_{\widehat{j}} - \min_j c_j}{D} = \underbrace{\frac{\min_{j \in S} c_j - \min_j c_j}{D}}_{\text{coverage loss}} + \underbrace{\frac{c_{\widehat{j}} - \min_{j \in S} c_j}{D}}_{\text{ranking loss}}.$$

True costs and the shared full-bank denominator enter only evaluation. A better shortlist need not give a better action if the reader misranks it.

Screening does not robustly outperform correction alone: a root-only interval $[-.0408, -.0083]$ becomes $[-.0694, +.0224]$ under crossed seed/root resampling (difference $-.0248$; 4,000 draws; seed differences $-.0553/-.0751/+.0066/+.0245$). Crossed contrasts with pooling/endpoint/zero-innovation/nominal/random screening are $.00015[-.01434, +.01652]/-.0421[-.0696, -.0097]/-.0200[-.0415, -.0009]/-.0212[-.0426, +.0009]/-.0359[-.0550, -.0158]$. Better coverage offsets worse ranking against pooling; the trade-off reverses against nominal and zero-innovation screens. Ranking difficulty depends on the subset, so the decomposition is descriptive.

Table 15: Six-block conditional re-evaluation on the same 48 roots and four fitted seeds. Each screen regenerates 24 of 125 candidates. Regret intervals resample seed and root axes; p99 measures the matching complete FP32 feedback path. Coverage and ranking are evaluated costs, not selector inputs.

Method	Coverage	Ranking	Total regret [95% interval]	p99 (ms)
Correction only	—	—	.6471 [.5997,.6930]	3.946
Full reroll	.0000	.6383	.6383 [.6145,.6638]	456.090
Full screen	.2341	.3882	.6223 [.6019,.6439]	97.497
Pooling screen	.2516	.3705	.6222 [.6012,.6422]	97.504
Endpoint screen	.2874	.3770	.6644 [.6354,.6887]	97.473
Zero-innovation screen	.2071	.4352	.6423 [.6149,.6704]	97.480
Nominal screen	.2108	.4328	.6435 [.6209,.6659]	96.451
Random screen	.2646	.3936	.6582 [.6357,.6825]	96.445

Quality and timing use the same image-to-action computation: feedback-time FP32 without TF32 and projections rounded to the FP16 cache representation; nominal computation uses cuDNN TF32. All 432 root/effect cells are rescored at these settings, unlike the stored-feature depth sweep.

Timing spans host image to host action with nominal futures, paths, costs and prefix features prepaid. Correction screens encode three frames and reuse the last for reroll; nominal/random/full-reroll arms encode only the last. Only full/pooling screens therefore match observation use. One RTX 5090 process measures 432 warmed, randomized calls per arm over six roots, three effects and four seeds, versus 48 roots for quality; subset/full-bank scores agree.

Median latency is 3.914 ms for correction, 97.223 ms for screening and 455.061 ms for full reroll; p95 values are 3.932, 97.436 and 455.895 ms. Nominal preparation costs mean 7.943 s (7.832–8.029 across 48 roots), including setup, candidate generation and world-model execution. Shared tensors occupy 2.630 MB, excluding weights, reader states and host data. These prepaid costs are excluded from feedback latency.

The stored-feature sweep is not paired with this timing. Its seed-conditional reroll-minus-correction regret is $+.1167[+.0977, +.1351]$, $-.0440[-.0609, -.0273]$ and $-.0010[-.0225, +.0205]$ at three, six and nine blocks; all nine-block budget intervals include zero. At six blocks, $k = 24$ gives $-.0509[-.0680, -.0346]$ versus $-.0214$ for random screening. Only the corrector extrapolates beyond training depth, and reduced-precision regeneration is not quality-matched here.

Training depth and decision quality. Table 16 holds test banks, twelve seeds, candidates and reader settings fixed, comparing three-block with evaluation-depth training. Each training bank has 64 roots; newly sampled matched-depth roots confound depth with training data. Both readers receive longer sequences, but only reroll receives regenerated futures.

Both readers improve across five families. Matched-depth budget-24 results meet the $.03$ margin on Peg and Door where shallow training fails; these are exploratory comparisons, not replacements for registered outcomes.

Without reroll ranking (budget one), matched-depth correction outperforms regeneration on all five families ($-.0289$ to $-.1082$, intervals excluding zero), whereas shallow training favors regeneration on three. Assembly complete-path medians are $3.7/455$ ms. The contrast establishes training-protocol sensitivity, not an isolated effect of depth.

Matched-depth results remain conditional on the selected task families.

Decision quality at trained depth. At the trained three-block depth, existing predictions give correction regret $.2907[.2573, .3246]$ and correction-minus-reroll $-.1167[-.1436, -.0903]$ (four seeds, 48 roots, nine effects, 4,000 crossed draws; actions selected before aggregation). Uniform/nominal regret is $.7156[.6992, .7323]/.7216[.7027, .7409]$ with root-only intervals. All costs use negative normalized reward; this is a reanalysis, not an independent trial.

Table 16: Training depth on fixed evaluation banks: shallow readers train at three blocks; matched readers at evaluation depth. Differences are from full regeneration, with crossed intervals over 48 roots and twelve seeds. Budget one is correction alone. Shallow budget-24 endpoints retain 10,000 registered resamples; other entries use 4,000 and do not revise those verdicts (Appendix H.7).

Family	Readers	Full regeneration	Budget 1 minus full	Budget 24 minus full
Assembly	Shallow	.6526	$-.0053$ [$-.0342, +.0280$]	$-.0190$ [$-.0341, -.0044$]
	Matched	.4933	$-.0803$ [$-.1161, -.0449$]	$-.0048$ [$-.0096, -.0003$]
Peg	Shallow	.5775	$-.0139$ [$-.0542, +.0278$]	$+.0000$ [$-.0284, +.0306$]
	Matched	.3840	$-.0722$ [$-.0974, -.0479$]	$-.0305$ [$-.0421, -.0203$]
Door	Shallow	.2167	$+.0841$ [$+.0398, +.1325$]	$+.0556$ [$+.0290, +.0841$]
	Matched	.1976	$-.0502$ [$-.0732, -.0289$]	$-.0004$ [$-.0012, +.0002$]
Hammer	Shallow	.6324	$+.0420$ [$+.0150, +.0776$]	$-.0003$ [$-.0033, +.0026$]
	Matched	.2087	$-.0289$ [$-.0444, -.0133$]	$-.0014$ [$-.0046, +.0012$]
Backbone	Shallow	.6366	$+.0408$ [$+.0112, +.0713$]	$-.0022$ [$-.0171, +.0108$]
	Matched	.5148	$-.1082$ [$-.1364, -.0813$]	$-.0062$ [$-.0115, -.0020$]

Table 17: Six-block confirmation on 48 roots unused for configuration or budget choice, with twelve independent training seeds. Success-truncated regret (Appendix H.1) and differences from full reroll use crossed seed/root intervals. Exploratory p99 times complete calls, not component sums: 1,080 per arm on six roots, three effect levels and one seed's readers.

Method	Normalized regret	Difference from full reroll	$p99$ (ms)
Full screen k=24	.6336 [.6115,.6551]	$-.0190$ [$-.0341, -.0044$]	99.6
Pooling screen k=24	.6350 [.6132,.6567]	$-.0176$ [$-.0335, -.0036$]	99.5
Correction alone	.6473 [.6262,.6696]	$-.0053$ [$-.0343, +.0274$]	4.1
Full reroll	.6526 [.6304,.6747]	—	456.1
Random screen k=24	.6653 [.6433,.6868]	$+.0127$ [$+.0003, +.0250$]	—
Nominal	.8060 [.7922,.8202]	—	—

Independent-root confirmation. Quality endpoints, margin, seeds and stopping were fixed before evaluation, with analysis revised before training. Registration is local, without third-party timestamping; timing is exploratory. Confirmation uses 48 new six-block roots and twelve new reader seeds, holding training roots, generation settings and world-model weights fixed: test roots and optimization vary, not training-data draws.

Confirmation requires the crossed 95% upper bound for screening-minus-reroll regret to be below .03, resampling seeds and roots with nine effects retained. This margin follows a precision calculation, not practical tolerance: SD .04788 for the more variable screening–correction contrast gives half-width .03 near ten seeds. Complete-path p99 is reported separately (Figure 14).

Screening meets the margin and improves regret (Table 17); nine of twelve seeds favor it, with differences $-.0711$ to $+.0113$. Exploratory complete-path p99 is $99.6/456.1$ ms for screening/reroll and 99.5 ms for pooling screening. These are pooled call-level quantiles, not registered timing endpoints.

At the same 24-of-125 budget, learned screening improves regret whereas random screening worsens it. Full/pooling coverage and ranking losses are $.2187/.4149$ and $.2550/.3801$, respectively, yielding similar totals. The result supports learned selection, not transport-specific superiority; correction alone is not separated from reroll.

The fixed reroll reader faces different coverage and ranking difficulty across subsets; correction-only and pooling contrasts are secondary (Table 17).

The same margin, twelve seeds and 10,000-resample analysis apply to subsequent tasks (Table 16, shallow budget 24). Selection is outcome-dependent: Peg/JEPA-WM follow Assembly; Door/Hammer follow those results. JEPA-WM and Hammer pass, Peg misses, and Door is worse in

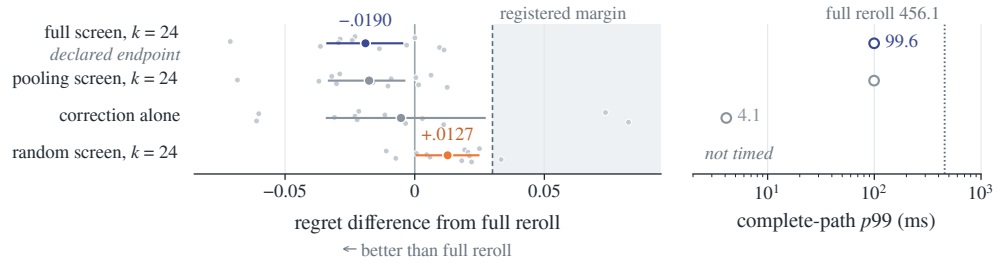


Figure 14: **Learned screening on independent roots.** Left: success-truncated regret minus unconditioned full reroll, with crossed seed/root 95% intervals; gray: vertically jittered seed differences. Noninferiority requires the upper bound below the dashed .03 margin (failure region shaded); correction alone is secondary. Right: exploratory complete-path p99, without intervals. Hollow markers and dotted reroll line use 1,080 calls per arm in one process on six roots, three effect levels and one seed's readers (Table 17).

Table 18: Long-bank memory and pre-feedback preparation; decimal MB. Peak allocation is the maximum across processes; preparation excludes nominal-bank generation.

	Reader	Extra cache	Peak allocation	Median (ms)	p99 (ms)
OEWM	3.871	.128	235.917	1.254	84.840
Pooling	3.862	.152	236.868	1.071	93.949
Endpoint	3.900	.190	237.916	1.052	89.296
GRU	3.869	4.168	246.393	1.208	96.843
Reroll	3.852	.000064	2082.075	1.105	81.840

every seed. Peg's exploratory 4,000-draw upper bound .0291 does not replace the registered .0306. Door/Hammer have weak reward–success alignment; their nominal-selection prediction fails under latency. Only Assembly preregistered cost; Peg's exploratory p99 is $99.53/455.88$ ms. Root-subset single-process timing does not establish worst-case or joint quality/cost bounds.

Reconstructed innovations and visual features differ from stored values by at most 7.7×10^{-3} (all innovations) and 7.60×10^{-3} (two roots, three effects); rounding native features to FP16 matches exactly. Rounding intervals admit one nominal prefix across effects, but unavailable unrounded states prevent unique attribution or a guarantee of identical decisions.

All arms also hold 44.720 MB of shared cache and the 173.440 MB frozen model; full/reroll peak reserved memory is 268.435/3468.689 MB. Preparation quantiles cover 320 preparations per arm, cold starts included, exclude nominal-bank generation, and are not end-to-end speedups including generation.

Task selection used a shared-scene, one-seed compliance result ($-2.43[-4.86, -.69]$ transport success points despite nominal/oracle success 1.0). New-scene tests remain conditional on that selection and use fixed-depth simulator decisions with scripted closeout.

Unseen combinations with matched factual history. Fixed-depth transfer is tested on 90 permutations of $AABBCC$, using frozen DINO-WM and 48/12/32 independently pooled training/validation/test roots for Assembly and Door, without outcome filtering. Three five-action primitives use Cartesian controls $0.85b_j + 0.15e_j$ (nominal chunks b_j at 0, 2, 4; basis direction e_j; unchanged gripper). Primitives depend on nominal context, so matched counts need not imply matched endpoints. A pre-outcome random 54/18/18 word split supplies training/validation/unseen words; test roots compare 18 familiar and 18 unseen words. All test primitive-position and adjacent-pair combinations occur in training. Nine nonidentity compliance conditions provide a three-frame factual prefix.

Interchanger, pooling, GRU and residual MLP share nominal futures, paths and three-frame innovation; pooled nominal features may encode order. History-conditioned reroll adds the same

Table 19: Fixed-depth combination generalization on independent test scenes. Top: normalized regret on familiar and unseen words (lower is better), and success (%) on unseen words. Bottom: paired unseen-word regret differences, Interchanger minus comparator, with 95% crossed seed/root intervals. All seven methods and both tasks are retained.

	Assembly			Door		
Method	Familiar	Unseen	Success	Familiar	Unseen	Success
Interchanger	.0387	.2152	24.74	.3067	.3225	42.97
Pooling	.1114	.2460	20.75	.3544	.3727	42.10
GRU	.0385	.3118	19.53	.2776	.2919	42.01
Residual MLP	.1433	.3799	16.67	.3409	.3922	38.54
History reroll	.0370	.2649	25.43	.2721	.2793	44.70
Original reroll	.1013	.2242	29.77	.2987	.2900	44.79
No-feedback scorer	.2520	.5317	21.70	.3461	.4015	38.02

Comparator	Assembly difference [95% CI]	Door difference [95% CI]
Pooling	$-.0308$ $[-.0695, +.0037]$	$-.0502$ $[-.1077, +.0124]$
GRU	$-.0967$ $[-.1612, -.0429]$	$+.0306$ $[-.0137, +.0736]$
Residual MLP	$-.1648$ $[-.2202, -.1094]$	$-.0697$ $[-.1408, -.0082]$
History reroll	$-.0497$ $[-.0783, -.0203]$	$+.0432$ $[-.0037, +.0894]$
Original reroll	$-.0091$ $[-.0395, +.0196]$	$+.0325$ $[-.0369, +.1016]$
No-feedback scorer	$-.3165$ $[-.3468, -.2904]$	$-.0790$ $[-.1495, -.0090]$

innovation to futures regenerated from the last frame; original reroll omits it. No-feedback scoring uses nominal futures/paths only. Active parameter counts are $.968/.966/.981/.961/.981/.427/.969$ million, respectively: all except the original-reroll reference match Interchanger within 1.4%, not in arithmetic cost.

Four seeds per arm train with normalized candidate-reward MSE, AdamW (weight decay .01), batch 16, unit clipping and 150 epochs. Validation regret selects learning rate (3×10^{-4} or 10^{-3}) and checkpoint every ten epochs. Within each 18-candidate bank, regret uses the success-truncated objective of Appendix H.1; success is separate. Successful candidates out-score failures in 82.7% of 185 mixed unseen Assembly cells and none of 180 Door cells. Pointwise paired intervals use 4,000 crossed seed/root draws, retaining nine conditions per root.

Transfer depends on task and candidate set (Table 19): unseen-minus-familiar regret is $+.1765[+.1472, +.2051]$ on Assembly and $+.0158[-.0620, +.0938]$ on Door. Assembly favors Interchanger over GRU, MLP and history reroll; Door does not reproduce the GRU/history gains. Neither separates it from pooling or original reroll. Adding reroll history changes regret by $+.0406[+.0160, +.0667]/ -.0107[-.0651, +.0518]$. Because reward–success alignment differs, these are fixed-depth reward-selection results, not success-transfer or quality-matched latency claims.

Perfect-condition headroom bounds the value available from an observation. For raw cost $C_{e,k}$ and uniform weights over nine conditions, define

$$V_{\text{perf}} = \min_k \frac{1}{9} \sum_e C_{e,k} - \frac{1}{9} \sum_e \min_k C_{e,k},$$

Both terms use the same raw scale. Unseen-bank headroom is $64.95[57.60, 72.93]$ on Assembly and $54.32[36.26, 73.78]$ on Door (root resampling); Interchanger's gain over the hindsight optimal static choice is $-3.59[-18.27, +9.82]$ and $-93.97[-131.98, -64.41]$ (crossed seed/root). Beating a learned no-feedback scorer therefore need not capture this oracle gap. Negative capture includes ranking error relative to an unavailable hindsight reference, not intrinsically harmful feedback. Image-usable value remains unknown (Appendix H.9); binary-success headroom is $.1111/0$. These are 18-word banks, not the 125-candidate studies.

Table 20: Unseen combinations with G supervision. Four-seed means on the same independent scenes and unseen words as Table 19; regret is normalized within each bank and success is a percentage.

	Assembly		Door	
Reader	G regret	Success	G regret	Success
Interchanger	.3484	32.29	.0528	98.61
Pooling	.3038	34.38	.1141	98.52
GRU	.3617	30.64	.0590	98.52
Residual MLP	.4692	23.00	.1153	98.61
History reroll	.3664	28.30	.0639	98.61
Original reroll	.4362	26.82	.0510	98.61
No-feedback scorer	.4646	24.65	.1421	98.61

Table 21: Assembly on the two released Meta-World world models, which share a DINOv2 ViT-S/14 encoder; eight paired seeds, scene pools shared across splits. Spread is maximum minus minimum seed success, not a confidence interval.

	DINO-WM			JEPA-WM		
Method	Success	Gain	Spread	Success	Gain	Spread
Nominal	74.33	—	—	74.83	—	—
Reroll	88.07	13.74	.73	86.99	12.16	1.51
OEWM	88.20	13.87	.73	88.69	13.86	.45
Pooling	88.72	14.39	.67	88.83	14.00	.67
Broadcast	78.96	4.62	2.29	78.26	3.43	6.14
Branch shuffle	74.61	.28	1.28	74.76	$-.08$	.28
Physical oracle	91.29	—	—	91.13	—	—

Objective-aligned combination transfer. An exploratory target-only comparison uses absorbing-success return G (Appendix H.9, $T = 280$). All seven readers, splits, seeds, capacities, optimization and validation selection are unchanged. First-success times are recovered from the same trajectories without changing actions or success labels. The test banks were already inspected.

Under G, Interchanger beats pooling on Door ($-.0613[-.1251, -.0180]$) but not Assembly ($+.0447[-.0337, +.1139]$), and Door's no-feedback scorer has equally high success (Table 20). Binary-success MSE also gives no consistent advantage: Interchanger-minus-pooling success is $-7.12[-14.15, -.69]$ points on Assembly (35.07/42.19%); Interchanger/no-feedback reaches 98.44/98.61% on Door. Hindsight-static G capture remains negative ($-204.21/-11.55$).

H.8 Public backbones and independent scene-pool checks

DINO-WM and the released Meta-World JEPA-WM (Terver et al., 2025) share a DINOv2 ViT-S/14 encoder and keep their published predictors, so this compares predictor designs, not independent visual representations. Each uses 64 training and 64 test banks (1,792 cells per split) and eight paired training seeds. Splits reuse scenes from one 50-instance pool, with different candidates, so Table 21 is not independent-scene evidence.

Both backbones support branch-specific correction, not superiority over pooling (Table 21). Shuffle removes nearly all gain, broadcast retains some, and pooling matches or exceeds full. JEPA-WM gains over nominal and same-seed reroll are $13.86[13.12, 14.63]$ and $1.70[1.20, 2.22]$ points; DINO-WM's reroll contrast is $.13[-.40, .66]$. Intervals resample banks after averaging eight seeds, not scenes. The DINO-WM eight-seed and four-seed means, 88.20% and 86.5%, are different estimands.

Table 22: Graded regret differences (success-truncated reward, Appendix H.1), full minus competitor, with 95% bank-bootstrap intervals. Negative favors full. The scene-pool change also changes realized test banks.

Family	Comparison	Shared pool	New pool
Default	vs. pooling	$.014[.010, .019]$	$-.009[-.020, -.001]$
Default	vs. GRU	$-.031[-.036, -.026]$	$-.033[-.040, -.026]$
Lag	vs. pooling	$-.020[-.035, -.005]$	$-.029[-.042, -.017]$
Lag	vs. GRU	$.009[-.003, .023]$	$.018[.006, .031]$

Device-resident timing uses an unfitted corrector and latent-distance reroll reader, so it does not match decision quality. DINO-WM BF16 p99 is $71.20/71.44$ ms under graph capture/eager execution; JEPA-WM uses eager execution. Configuration differences and component-wise quantiles preclude a complete-path comparison. Independent long-bank timing instead evaluates quality and cost jointly.

New scene pools. New-pool DINO-WM tests use 64 default and 48 lag banks from a 50-instance pool with randomization vectors disjoint from training. Training banks, four seeds, 150 epochs and candidate generation are unchanged. Intervals bootstrap banks after seed averaging, without scene clustering or seed resampling. This excludes graded backlash and JEPA-WM and tests pool separation, not every stored state.

New-pool results preserve endpoint separation, not uniform superiority over path readers (Table 22). Full-minus-endpoint regret is $-.072[-.083, -.062]$ under default and $-.061[-.077, -.047]$ under lag. Default regret increases by .006 for full and .029 for pooling, confounded with the changed banks. Under lag, GRU beats full while reward opposes success (20.7–27.4% across readers versus 66.7% nominal). Eight-seed default success gains are 13.41 points for pooling and 13.23 for full.

H.9 Feedback information, decision quality and length transfer

The matched-information comparison jointly evaluates quality and feedback cost (Table 1); the short-to-long study evaluates objectives and training budgets (Table 29). The accompanying controls distinguish feedback information, reader capacity and execution choices.

Four-seed short-bank success for nominal/endpoint/full/reroll/oracle is 76.3/87.4/88.0/81.6/96.4% on Peg and 74.3/83.6/86.5/86.1/91.3% on Assembly. Assembly's first-seed gain, $12.39[11.83, 12.95]$ points, is distinct from its four-seed mean gain of 12.19; a four-seed ensemble gains $3.91[2.62, 5.19]$ over endpoint. Full-minus-reroll gains are $6.40[4.24, 8.78]$ on Peg and $.42[-.32, 1.17]$ on Assembly.

Branch shuffling and broadcasting reduce one-seed Assembly success from 86.72% by $14.17[13.17, 15.23]$ and $13.95[12.78, 15.35]$ points. These ablations remove branch-specific information that pooling retains; they do not isolate sequential transport.

A fractional actuator-delay queue of 0–2.75 environment steps, hidden from all methods, persists through prefix and continuation and is restored before every candidate; all 32 train and 32 separately generated test banks pass the complete-reroll teacher check. Four-seed mean success for nominal/broadcast/endpoint/reroll/full is $68.44/66.87/83.44/82.03/86.72\%$ (oracle 100%), with endpoint/reroll/full seed ranges 69.1–88.4, 74.7–88.4 and 85.6–87.8%. Every full seed has a positive bank-clustered gain interval, yet endpoint remains competitive.

Short-bank timing and objective mismatch. Short-path p99 is $3.736/3.709/205.012$ ms for full/endpoint/FP32 reroll. A separate full-bank probe gives $204.92/74.12/74.58$ ms for FP32/BF16/FP16, so reduced precision meets 187.5 ms at three chunks. FP16/BF16 alter 10/76 of 896 FP32 choices (98.88/91.52% agreement), yet every precision succeeds only 11.05%. This probe uses terminal-latent norm plus action penalty, not task-supervised readout; it establishes neither preserved task quality nor OEWM–reroll agreement.

Task-supervised reduced-precision checks. To isolate precision under task supervision, all 125 candidates are regenerated at each precision and read by the same frozen task readers, without refitting.

Reduced-precision regret changes are small but do not establish equivalence. Relative to FP32, BF16 and FP16 differences are respectively $+.0006[-.0048, +.0064]$ and $+.0006[-.0011, +.0025]$ on Assembly; $-.0020[-.0111, +.0061]$ and $-.0008[-.0043, +.0020]$ on Peg; $-.0013[-.0081, +.0057]$ and $+.0000[-.0012, +.0018]$ under backlash; and $+.0049[+.0003, +.0106]$ and $+.0004[-.0017, +.0030]$ under latency. Crossed seed/root intervals lie within Peg's registered $\pm.02$ band, whose practical relevance is unspecified. The objective remains success-truncated reward; choice agreement is 88–94% for BF16 and 97–99% for FP16.

BF16 reduces median complete-path regeneration latency from 455 to 161 ms on Assembly and 488 to 177 ms under latency, versus 3.7 ms for cached correction. These single-process feedback-cost ratios do not imply equal quality or worst-case deadline guarantees.

On the focused banks (64 roots, 1,792 cells, four seeds), BF16/FP16 regret differences from FP32 (.3914) are $-.00080[-.00516, +.00283]/\ +\ .00021[-.00110, +.00152]$, with choice agreement $.869/.974$. Median regeneration latency is $219.5/79.9/80.1$ ms for FP32/BF16/FP16; p99 over 160 calls is $444.1/165.3/165.7$ ms. No equivalence margin was predeclared for this cohort.

Other physical mechanisms. Matched-depth testing fixes all settings across three prespecified families: backlash, actuation latency and affine gain error.

At matched depth, unconditioned-reroll regret is .3091 under backlash, .4962 under latency and .4393 under affine error. Cached correction improves it by $-.0507[-.0681, -.0335]$, $-.0787[-.0958, -.0619]$ and $-.1577[-.1752, -.1399]$, respectively; budget-24 differences are $-.0185[-.0287, -.0095]$, $-.0080[-.0129, -.0038]$ and $-.0570[-.0683, -.0467]$. Full-budget identities hold exactly in two families and within 5.7×10^{-6} for latency. With feedback matched, cached-minus-conditioned-reroll differences become $+.0024[-.0106, +.0153]$, $+.0090[-.0045, +.0227]$ and $+.0125[+.0004, +.0237]$ (Table 23, Figure 15). The first two are inconclusive and affine error favors regeneration; intervals spanning zero do not establish equivalence.

A prespecified screening rule fails under latency. It used the sign of nominal-minus-chance regret to predict no screening gain, yet screening attains .4176 versus .6842 nominal regret. Earlier agreement on mostly same-sign banks does not support this rule.

Perfect-condition headroom and observation-usable value. Fix a root, candidate bank and fitted reader; let $C(e, k)$ be raw branch cost under hidden condition e and Z the reader's declared observation. Define

$$R_0 = \min_k \mathbb{E}C(e,k), \qquad R_Z = \mathbb{E}_Z \min_k \mathbb{E}[C(e,k) \mid Z], \qquad R_{\mathrm{perf}} = \mathbb{E} \min_k C(e,k).$$

Then $R_{\mathrm{perf}} \le R_Z \le R_0$: the conditional expectation of a minimum is at most the minimum of conditional expectations, and a constant choice is feasible with Z. Hence

$$0 \le V_Z := R_0 - R_Z \le V_{\mathrm{perf}} := R_0 - R_{\mathrm{perf}}, \qquad V_{\mathrm{cap}}(m) := R_0 - \mathbb{E}C(e, \hat{k}_m(Z)) \le V_Z.$$

The last inequality follows by minimizing conditional cost, also for randomized choices whose randomness reveals nothing more about e. A no-observation reader thus has $V_{\mathrm{cap}} \le 0$, not a prescribed negative floor. Equiprobable cost rows $(0, 1)$ and $(1, 0)$ with identical observations give $V_{\mathrm{perf}} = 1/2$ but $V_Z = 0$: perfect-condition headroom need not be available from images.

We measure V_{perf} and unclipped V_{cap}, not the Bayes quantity V_Z. Costs share one raw scale within a root, which per-cell normalization would destroy. The static reference R_0 needs all fact-averaged branch costs, unavailable to a deployed reader. Capture shares are ratios of sums across roots, not averages of unstable per-root ratios.

Excluding the no-feedback identity condition, perfect-condition headroom V_{perf} is positive on every bank: $247.45[230.21, 265.89]$ on focused Assembly, $251.06[236.85, 264.98]$ on registered six-block roots, $130.47[104.25, 157.45]$ on Peg, $49.14[30.62, 69.49]$ on Door, $156.27[143.46, 168.95]$ under backlash and $226.30[207.15, 244.16]$ under latency.

Table 23: Conditioning and cache reuse: graded normalized-regret differences (Appendix H.1) with crossed seed/root intervals. Conditioning is conditioned minus unconditioned regeneration; cached vs. regen. is cached minus conditioned regeneration; screen vs. cond. is screen minus conditioned regeneration. Negative favors the first arm. Depth denotes training–evaluation blocks; conditioning adds the same three-frame innovation. All columns share six seeds per bank. This exploratory comparison does not recompute a registered endpoint.

Bank, depth	Conditioning	Cached vs regen.	Screen vs cond.
Focused assembly, 3	$-.1644$ $[-.1841, -.1465]$	$+.0006$ $[-.0063, +.0075]$	$+.1225$ $[+.1108, +.1347]$
Backlash, 6	$-.0514$ $[-.0822, -.0257]$	$+.0024$ $[-.0106, +.0153]$	$+.0290$ $[+.0140, +.0448]$
Actuation latency, 6	$-.0559$ $[-.0723, -.0399]$	$+.0090$ $[-.0045, +.0227]$	$+.0712$ $[+.0515, +.0940]$
Affine gain, 6	$-.1728$ $[-.1910, -.1560]$	$+.0125$ $[+.0004, +.0237]$	$+.1195$ $[+.1057, +.1333]$
Registered six-block, 3–6	$-.0237$ $[-.0777, +.0189]$	$+.0452$ $[+.0114, +.0815]$	$-.0098$ $[-.0337, +.0151]$
Peg, 3–6	$+.0119$ $[-.0299, +.0579]$	$-.0090$ $[-.0504, +.0283]$	$-.0030$ $[-.0423, +.0329]$

The best branch changes across facts: focused Assembly uses $23.42[22.98, 23.86]$ distinct optima over 28 conditions, versus 7.38–8.25 over nine or ten conditions elsewhere. The most-often-best branch serves only $9.4[8.7, 10.0]\%$ of focused conditions and 19.9–24.3% elsewhere. On focused banks, committing to it costs $342.27[321.21, 363.13]$: it is not the best-on-average static choice defining R_0.

On focused Assembly, captured value is $102.98[91.27, 115.12]$ for correction, $100.56[88.74, 113.37]$ for pooling, $73.97[64.38, 84.30]$ for GRU, $58.96[52.30, 66.44]$ for endpoint and $8.72[3.77, 13.83]$ for unconditioned reroll. Correction captures 41.6% of total headroom and is positive on all 64 roots; reroll captures 3.5% and is positive on 41. The feedback-blind scorer captures $-6.48[-9.15, -4.19]$, negative on every root, while nominal selection costs 240.25 more than the static reference. These are observed values, not theorem-imposed floors.

Beyond training depth, every arm has negative capture on registered six-block roots, including correction ($-187.19[-211.82, -163.18]$) and reroll ($-172.31[-191.76, -153.14]$); none is positive on more than one of 48 roots. Peg's seven arms range from -194.30 to -174.43 with overlapping intervals. These comparisons establish neither the Bayes value of images nor depth as the cause.

The unseen-combination banks have headroom 64.95 and 54.32, yet Interchanger capture is -3.59 $[-18.27, +9.82]$ on Assembly (inconclusive, not evidence of zero capture) and -93.97 $[-131.98, -64.41]$ on Door. Positive headroom alone cannot distinguish an inadequate reader from an uninformative observation.

Raw captured value and normalized regret rank the seven arms identically on focused Assembly; on six-block roots only the last two positions differ, with overlapping intervals. This does not identify depth as the cause of between-bank differences.

Separating conditioning, reader class and cache reuse. An exploratory six-seed factorial comparison separates feedback conditioning from cached versus regenerated futures within the corrector's reader class. It also includes the original unconditioned reader and a feedback-blind training/evaluation control. These contrasts supplement, rather than replace, the twelve-seed registered endpoint.

Changing reader class without innovation lowers latency-family regret by $-.0261[-.0492, -.0065]$; five other intervals include zero. Adding innovation improves all four matched-depth banks, but not conclusively the two extrapolated banks (Table 23).

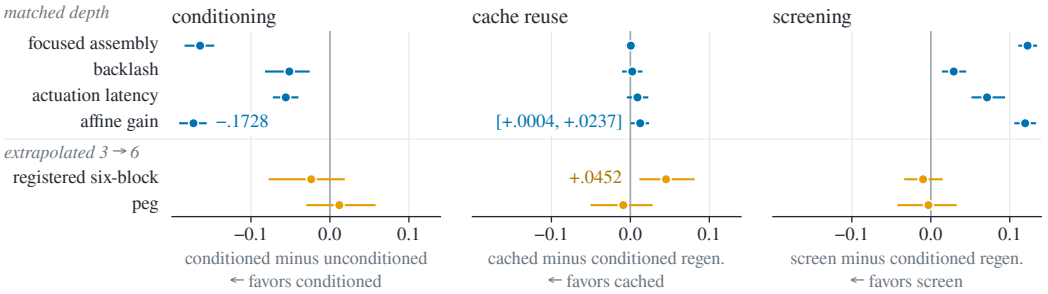


Figure 15: **Feedback conditioning versus cache reuse.** Exploratory normalized-regret differences with pointwise crossed seed/root intervals, six seeds per bank. Left: conditioned minus unconditioned regeneration; middle/right: cached/screening minus conditioned regeneration. Negative favors the first arm. Blue: matched training depth; orange: three-to-six-block extrapolation. Intervals including zero do not establish equivalence. Values in Table 23.

With conditioning fixed, three matched-depth cache/reroll intervals include zero, while affine gain and extrapolated six-block banks favor regeneration. Thus gains over unconditioned reroll combine information and computation differences; nonsignificance does not establish equal quality.

Feedback blinding increases regret from .2326 to .4418 on focused Assembly and from .6830 to .8415 at six blocks. Focused captured value is 109.16, versus 102.98 under separately fitted readers above; these are not paired estimates.

Conditioned/original reroll regret is $.0370/.1013$ on familiar Assembly and $.2721/.2987$ on familiar Door. On unseen Assembly, conditioning instead adds $+.0406[+.0160, +.0667]$ regret: a transfer failure of the fitted reader, not proof of uninformative observations.

Six-seed screening is not separated from original reroll ($-.0080[-.0228, +.0065]$) or conditioned reroll ($-.0098[-.0337, +.0151]$); these do not recompute the twelve-seed endpoint. On focused Assembly, conditioned full reroll improves over the original screen by $.1225[.1108, .1347]$. Using that reader on the shortlist gives .2305 versus .2320 on the full bank, without establishing equivalence.

Restricting the candidate library in the visual pipeline. Motivated by the controlled-plant result, we compare low-reversal, short-path and random-acceptance libraries. Each accepts a quarter of proposals on the same roots and factual conditions and replenishes to 125 candidates (3.7–4.1 proposals per accepted plan; equal rollout counts). Unlike the controlled-plant action restrictions, the banks need not be nested.

Let A be the unrestricted sampled bank and B a replenished bank; raw cost decomposes as

$$C(\widehat{k}_B) - \min_{k\in A} C(k) = \underbrace{C(\widehat{k}_B) - \min_{k\in B} C(k)}_{\text{selection error}} + \underbrace{\min_{k\in B} C(k) - \min_{k\in A} C(k)}_{\text{signed optimum difference}}.$$

Only the first term is necessarily nonnegative; the second compares sampled optima and is a capability loss only if $B \subseteq A$.

Reversal restriction lowers median reversed travel from .411 to .258 at unchanged path length 36.8; length restriction gives length 31.3 and reversed travel .776.

Reversal restriction sacrifices sampled optimum relative to random acceptance: excess cost is $72.6[55.5, 89.1]$ under backlash and $63.9[45.1, 82.8]$ under compliance, contrary to prediction. Relative to the unrestricted bank, random acceptance improves optima by $93.0[71.1, 114.5]/61.5[46.5, 77.2]$, while reversal restriction changes cost by $-20.4[-38.0, -3.2]/ + 2.4[-11.6, +16.1]$. These finite-sample differences do not imply a broader underlying proposal distribution.

Despite worse optima, reversal restriction lowers selection regret relative to random acceptance by $.1195[.0918, .1470]$ under backlash and $.0448[.0192, .0738]$ under compliance; its mean total costs are also lower (Table 24).

Table 24: Visual library comparison in raw cost units, the negative success-truncated reward of Appendix H.1. Signed optimum difference compares each replenished bank's minimum with the unrestricted sampled minimum; it may be negative because banks are not nested. Selection error is relative to the bank's own minimum; total is their sum. Random acceptance matches the acceptance rate independently of the plan. Six seeds; intervals resample roots.

Restriction	Signed optimum diff.	Selection error	Total
Backlash			
none	0	$+163.3$ [$+150.9, +176.2$]	$+163.3$ [$+150.9, +176.2$]
reversal	-20.4 [$-38.0, -3.2$]	$+189.9$ [$+174.2, +206.0$]	$+169.5$ [$+153.0, +186.7$]
random	-93.0 [$-114.5, -71.1$]	$+293.3$ [$+273.3, +313.3$]	$+200.3$ [$+179.6, +221.5$]
length	-69.5 [$-93.5, -45.3$]	$+368.0$ [$+341.8, +393.4$]	$+298.5$ [$+275.4, +322.4$]
Compliance			
none	0	$+410.3$ [$+391.5, +430.5$]	$+410.3$ [$+391.5, +430.5$]
reversal	$+2.4$ [$-11.6, +16.1$]	$+417.3$ [$+399.6, +434.1$]	$+419.7$ [$+403.5, +436.3$]
random	-61.5 [$-77.2, -46.5$]	$+492.7$ [$+468.3, +516.9$]	$+431.2$ [$+409.4, +453.0$]
Affine gain			
none	0	$+191.5$ [$+182.4, +201.6$]	$+191.5$ [$+182.4, +201.6$]
reversal	$+18.7$ [$+7.4, +30.9$]	$+212.2$ [$+196.3, +230.0$]	$+230.9$ [$+214.1, +249.1$]
random	-19.9 [$-31.3, -7.6$]	$+393.4$ [$+373.6, +414.8$]	$+373.5$ [$+356.0, +391.2$]

Length restriction worsens both mean optimum and selection error relative to random acceptance; reversal restriction improves regret over it by $.2521[.2186, .2868]$. This ordering alone identifies no physical mechanism.

The largest regret improvement over random acceptance occurs under memoryless affine gain, $.2270[.1929, .2616]$, despite optimum cost worsening by $38.6[27.8, 48.7]$. This refutes mechanism specificity without isolating physics from candidate-distribution effects.

The unrestricted bank nevertheless has the lowest mean total cost in every family, without separation in every paired comparison. Restriction therefore makes selection easier without establishing net benefit; only the length-control prediction among four prespecified predictions holds.

Path information at trained depth. Lag tests with sixteen roots and 125 candidates give 8,928 pair–fact comparisons. Tight endpoint deciles are near matches, not collisions: paired assembly-success disagreement is 16.0% overall and 17.2% in the tight decile. Single-anchor banks widen endpoint spread, and full and pooling then reach the oracle on all 128 held-out banks. At eight of ten lag strengths all candidates share one binary outcome, a ceiling that cannot rank editors.

Dense reward removes this ceiling in headroom, not in agreement with success (Appendix H.1). Four seeds train at three blocks for 150 epochs on 48–64 banks per split, with all ten lag strengths. Only one candidate is best, reward-range headroom is 74–97%, and 57–59% of near-matched pairs differ. At three blocks under lag, full minus pooling is $-.020[-.035, -.005]$ and equal-compute GRU minus pooling $-.029[-.047, -.013]$; recurrent arms converge sooner, so the 40-epoch ordering differs. Pooling is best at six blocks; at nine all methods lie within .045 (Table 25). Backlash has zero measured drift over twenty held steps, whereas lag's per-edge retention is at most $0.70^5 = 0.168$. Under backlash at trained depth, differences between successive arms in the order full, GRU, pooling, reroll, endpoint, shuffle are $-.029[-.045, -.015]$, $-.023[-.042, -.004]$, $-.039[-.064, -.013]$, $-.079[-.137, -.028]$, $-.303[-.348, -.257]$: full is favored at that depth, not universally.

Fresh-scene evaluation with absorbing-success return. This study replaces success-truncated reward with $G = R + (T - \tau)r_\star$: first success at τ receives credit through $T = 280$ steps (30 plan, 250 closeout), with $r_\star = 10$; failures retain full-horizon reward R. Protocol and predictions were fixed before test-root generation, without third-party timestamping; subsequent changes concerned execution only. This cohort remains separate from shared-scene and binary-success studies.

Table 25: Graded simulator regret (success-truncated reward, Appendix H.1), four-seed means. Training uses three blocks; six and nine are unseen. Backlash is a separate trained-depth experiment, not the independent long-bank study.

	Actuator lag			Backlash
Method	3 blocks	6 blocks	9 blocks	3 blocks
Full	.291	.684	.731	.240
GRU, equal compute	.281	.669	.742	.269
Pooling	.311	.635	.741	.292
Endpoint	.355	.709	.698	.410
Reroll	.407	.640	.730	.331
Path shuffle	.642	.727	.716	.713

Six-block Assembly uses DINO-WM, $K = 125$, nine backlash levels .10–1.00 and the existing 64 training roots, relabeled with first-success times without changing trajectories. Two disjoint new pools provide 64 test roots and 576 cells. Only peg-goal position varies, within $[-.100, .097] \times [.751, .843]$ m; object start is fixed. Nearest training-scene distance has median 9.1 mm (range .93–24.8), versus 10.0 mm within training, with no coincident scenes. This tests new positions within range, not new physics or depth.

Six seeds fit OEWM/pooling/GRU/endpoint/unconditioned reroll and a within-cell permuted-target control for 150 epochs using AdamW (3×10^{-4}), batch 16, width 256, unit clipping and no tuning or early stopping. Active parameter counts are .97/.73/1.06/.53/.43M, respectively. Separate readers minimize MSE to normalized G or truncated reward, with frozen checkpoints and lowest-index tie-breaking. The primary OEWM-minus-pooling G-regret contrast uses 10,000 crossed seed/root bootstrap draws within pool, retaining nine effects per root. Success requires an upper bound below zero and point estimate $\leq -.02$ (2% of return span). Four secondary contrasts (OEWM–GRU, GRU–pooling, OEWM–endpoint/reroll) use two-sided bootstrap tests with Holm correction at .05; success, success-first rank, truncated regret and plan-end distance are descriptive.

OEWM meets the primary rule (Table 26), with all seeds favoring it by .020–.047 and pool-specific differences $-.031[-.046, -.017] / -.034[-.055, -.014]$; the interval does not establish a minimum .02 gain. Post hoc removal of the strongest effect preserves the sign but misses the point threshold (Table 27). Only the weakest level does not favor OEWM ($+.016[-.008, +.064]$); regrets are $\leq .064$ at the three weakest levels and the descriptive strength–difference correlation is $-.78$, not monotone. Holm-adjusted comparisons favor GRU over pooling and OEWM over endpoint/reroll, but do not separate the ordered readers. Their 33%/46% parameter advantages motivate the capacity control below. OEWM gains 2.0 success points and reduces plan-end distance by $10[3, 18]$ mm, without improving success-first rank ($-.006[-.028, +.017]$), which favors slower successes within the success class.

Truncated-reward training favors OEWM over pooling on its own regret ($-.053[-.083, -.028]$), but yields poor G regret (.559–.653 versus random .589) and success (27.7–42.9% versus random 37.1%, oracle 96.0%) across non-control readers. Training on G instead improves every non-control arm over random by .42–.48, with separated intervals; the permuted-target control does not ($+.106[-.042, +.246]$).

Among 576 cells, 553 contain both outcomes and 23 no success; successes occur after the plan, at steps 49–280. Mean G favors success in all mixed cells, versus 28.9% for truncated reward; strict ordering holds in 95.5% versus 5.1%. Highest-G selection succeeds whenever possible (96.0%), versus 55.9% for highest truncated reward, with no tied maxima or zero spans. This is not strict success consistency: sufficiently late success can rank below near-goal failure. The 36-cell pre-test check gives 100% mean alignment, 85.7% strict ordering and 97.2% highest-G success.

Pre-test predictions correctly anticipate endpoint and truncated-regret separation, an inconclusive OEWM–GRU contrast, and poor success under truncated-reward training. The primary estimate lies within the predicted 80% interval $[-.040, +.015]$. Two predictions fail: OEWM's success gain is

Table 26: Fresh-scene Assembly with backlash: 64 roots, 576 cells and six seeds. Both training objectives are evaluated by cell-normalized absorbing-success regret and selected-candidate success (%). Differences use G-trained readers with crossed seed/root 95% intervals. The primary regret contrast and four Holm-corrected secondaries were preregistered; success contrasts are descriptive. Exploratory rows are later capacity-matched fits on the same opened test bank, outside the registered test.

	Trained on G		Trained on truncated reward	
Arm	Regret	Success	Regret	Success
OEWM	.105	92.6	.576	41.2
Pooling	.138	90.6	.612	36.7
GRU	.113	92.0	.595	37.8
Endpoint	.168	87.4	.653	27.7
Reroll	.129	91.3	.559	42.9
Permuted-target control	.694	25.9	.688	25.6
Pooling, width 304 (exploratory)	.142	90.4	—	—
Random choice	.589	37.1	.589	37.1
Nominal plan	.608	33.3	.608	33.3
Oracle (highest G)	0	96.0	0	96.0

Difference, G-trained	Regret	Success (points)	Holm p
OEWM − pooling (primary)	$-.032\,[-.046, -.019]$	$+2.0\,[+0.7, +3.5]$	—
OEWM − GRU	$-.008\,[-.016, +.001]$	$+0.6$	.063
GRU − pooling	$-.024\,[-.041, -.009]$	$+1.4$	.003
OEWM − endpoint	$-.062\,[-.077, -.048]$	$+5.2$	$< .001$
OEWM − reroll	$-.023\,[-.035, -.011]$	$+1.3$	.001
OEWM − pooling, width 304 (exploratory)	$-.036\,[-.049, -.024]$	$+2.2\,[+0.9, +3.7]$	—

separated, and the permuted-target control does not beat random. These forecasts are not additional confirmatory tests.

Exploratory capacity-matched check. This post-test capacity check fixes its analysis before fitting. Width-304 pooling has 965,505 active parameters, .22% fewer than OEWM's 967,681; six seeds share the original G data and training settings. OEWM's regret advantage is $-.036[-.049, -.024]$, positive in every seed (.026–.048) and both pools ($-.033[-.047, -.019]$/ $-.040[-.061, -.020]$). Widening pooling changes regret by $+.004[-.004, +.013]$ and success by $-.2[-1.2, +.8]$ points; GRU beats widened pooling by $-.028[-.044, -.015]$ regret. Success gains are $2.2[.9, 3.7]$ points for OEWM and $1.6[-.03, 3.5]$ for GRU. This matches total parameters, not capacity allocation: pooling's encoder/gate/head widen together, GRU remains 10% larger, and no matched reader uses truncated-reward training. Width-256 reproduction matches the original fits.

Secondary success intervals exclude zero only for OEWM minus endpoint; Table 26 reports their point estimates. Reference arms require no training. Level-wise analyses were specified after both evaluations but before computing the breakdown.

Matched information, return and execution cost. The confirmation retains the six-block Assembly training bank, DINO-WM, backlash, $K = 125$ and return G. Two further disjoint pools provide 64 new roots and 576 unfiltered cells at the same depth/effect levels. Analysis and checkpoints are frozen before collection; all scene vectors differ from training and development. Results are not pooled with earlier cohorts.

OEWM, width-304 pooling, GRU and unconditioned reroll retain six frozen G-trained readers. Conditioned reroll shares OEWM's width-256 architecture, three-frame innovation and action tokens, but reads regenerated futures. A matched no-feedback arm zeros residuals during training/testing while retaining prefix actions. New arms use the same seeds, 150 epochs, AdamW (3×10^{-4}, weight

Table 27: Post hoc breakdown by backlash level: 64 cells per level and six seeds. Entries are normalized-G regret differences between G-trained readers; negative favors the first. Brackets: crossed seed/root 95% intervals with shared resamples, unadjusted for multiplicity and not supporting per-level claims. Width-304 pooling is the exploratory capacity-matched comparison.

Backlash	OEWM − pooling	OEWM − pooling, width 304	GRU − pooling
0.100	$+.016$ $[-.008, +.064]$	$+.018$ $[-.009, +.065]$	$-.008$ $[-.018, +.002]$
0.213	$-.005$ $[-.013, +.002]$	$-.003$ $[-.007, +.002]$	$+.008$ $[-.005, +.024]$
0.325	$-.004$ $[-.013, +.003]$	$.000$ $[-.007, +.007]$	$+.008$ $[-.010, +.035]$
0.438	$-.031$ $[-.057, -.009]$	$-.028$ $[-.053, -.007]$	$-.023$ $[-.048, -.003]$
0.550	$-.016$ $[-.047, +.007]$	$-.023$ $[-.054, -.002]$	$-.005$ $[-.034, +.016]$
0.663	$-.066$ $[-.118, -.024]$	$-.069$ $[-.122, -.027]$	$-.070$ $[-.117, -.031]$
0.775	$-.043$ $[-.080, -.007]$	$-.057$ $[-.091, -.024]$	$-.027$ $[-.065, +.010]$
0.888	$-.007$ $[-.044, +.029]$	$-.012$ $[-.053, +.032]$	$+.026$ $[-.016, +.067]$
1.000	$-.135$ $[-.214, -.059]$	$-.154$ $[-.230, -.082]$	$-.127$ $[-.206, -.053]$
All but 1.000	$-.0197$ $[-.033, -.006]$	$-.0217$ $[-.035, -.009]$	$-.0116$ $[-.026, +.002]$
All nine	$-.032$ $[-.046, -.019]$	$-.036$ $[-.049, -.024]$	$-.024$ $[-.041, -.009]$

decay .01), batch 16, unit clipping and normalized-G MSE, without tuning. Parameter counts are 967,681 for OEWM/conditioned reroll/no feedback, 965,505 for pooling, 1,062,913 for GRU and 427,009 for unconditioned reroll; the last comparison changes both capacity and information.

Table 1 reports matched-reader comparisons. The smaller unconditioned reroll achieves FP32 regret $.1154[.1029, .1278]$, success 92.10% and p99 491.70 ms; BF16 gives $.1169[.1042, .1295]$, 92.01% and 181.29 ms. Quality intervals cross seeds and roots; timing uses 4,320 calls per arm across three processes with the evaluated checkpoints and runtime.

The primary OEWM-minus-conditioned-FP32-reroll regret difference is $-.0025[-.0166, +.0111]$, using 10,000 crossed seed/root draws within pool and all nine effects. Descriptive success/raw-G differences are $-.03[-1.48, +1.39]$ points and $+4.22[-19.32, +26.83]$; equivalence and noninferiority are unestablished. Six secondary regret tests use Holm correction: OEWM against pooling, GRU and no feedback; conditioned against unconditioned reroll; and BF16 against FP32 for each reroll reader.

OEWM beats matched pooling and GRU on regret in both pools and every seed (Table 1); success gains are $3.36[1.24, 5.93]$ and $1.56[-.17, 3.59]$ points. Against no feedback, gains are $21.41[18.34, 24.36]$ success points and $.2495[.2270, .2713]$ regret. Conditioned reroll also beats its smaller unconditioned counterpart ($-.0239[-.0387, -.0066]$, Holm $p = .0366$), but that contrast mixes information and capacity. The matched results establish benefits in this setting, not universal architectural necessity.

Seed-wise regret ranges are $[.0840, .0983]$ for OEWM, $[.1200, .1531]$ for pooling, $[.0975, .1230]$ for GRU and $[.0708, .1199]$ for conditioned reroll. Their first-pool means are .0944, .1342, .1074 and .0957; second-pool means are .0836, .1315, .1106 and .0873. The OEWM–conditioned-reroll binary-success contrast changes direction between pools; the regret point estimate favors OEWM in both.

Only regeneration uses BF16; encoding, projection and readout remain FP32 with training-matched FP16 storage. Conditioned/unconditioned regret changes by $-.0006[-.0021, +.0005]/+$ $.0014[-.0009, +.0042]$ relative to FP32, altering 2.84/5.03% of choices. These are descriptive precision comparisons, not equivalence guarantees.

One RTX 5090 measures CPU-image/proprioception-to-CPU-action latency, including transfers, encoding, innovation, regeneration and readout but excluding acquisition and nominal preparation. Four roots per pool, three effects and six seeds run in three fresh processes with five warmups, 30 repetitions, randomized method order and no observed competing GPU compute. Reconstructed FP32 features agree on all 576 cells; timed and quality-evaluated choices match.

Table 28: Gate variants: graded regret minus pooling (Appendix H.1). Negative favors the variant. Training mixes two-, three- and four-block banks, so three blocks is an in-distribution test depth; six, nine and twelve are unseen. Brackets are reported paired intervals.

Variant	3 blocks	6	9	12 blocks
Normalized	$-.026[-.050, -.004]$	.081	.226	.203
Original	$-.023[-.049, .002]$	.052	.183	.170
Length augmented	$.011[-.009, .031]$	.026	.055	.061
Additive	$-.007[-.032, .018]$	.017	.050	$.024[-.014, .062]$
Ungated	$.029[.005, .055]$	.013	.024	$.024[-.014, .063]$

Median latency is 4.43 ms for OEWM, 491.18 ms for conditioned FP32 reroll and 182.20 ms for BF16; p95 values are 4.64, 493.98 and 184.22 ms. Process-specific p99 ranges are $[5.00, 5.11]$, $[494.98, 495.20]$ and $[184.84, 185.03]$ ms. All 4,320 cached and BF16 calls meet 187.5 ms; all FP32 calls miss it. This establishes lower feedback cost, not exclusive deadline feasibility, and does not replace the sixteen-chunk BCE study.

Cache preparation remains separate: simulator setup and nominal-bank construction take 7.79–9.02 s per root (mean 8.05); shared tensors occupy 2,599,856 bytes and OEWM adds 128,000 bytes per reader, excluding weights. Branch preparation takes median .093 ms, maximum 33.79 ms including cold start; no-feedback selection can be prepaid. Feedback-time ratios exclude these costs.

Reconstruction and decision error. With a frozen reader trained on regenerated rather than physical futures, three-seed multiplicative/Householder/fact-reset/GRU transports reduce relative reconstruction error to 2.6–2.7% (pooling 3.9%, endpoint 4.0%) without separating regret (.438–.442) from pooling. At nine blocks, state-growth/error pairs are $1.348/55.4$, $.994/.92$, $.976/.91$, $.958/.72$; none improves extrapolated regret. The Householder–identity blend is nonexpansive, not isometric; two exact reflections preserve norms but do not outperform pooling (regret .6901 versus .6829; $+.0072[-.0077, +.0244]$). Teacher features give regret .7291. Reconstruction and norm control alone do not ensure decision reuse under this objective.

Stability and length extrapolation. Under success-truncated reward, low gains approach unchanged-cache reconstruction, while unrestricted propagation can explode. With two/three/four-block training and six/nine/twelve-block testing, normalization meets trained-depth criteria against pooling and ungated recurrence (latter $-.055[-.079, -.034]$), not the twelve-block criterion. Augmentation, additive and ungated variants also fail the joint criteria (Table 28). Augmentation masks suffix inputs but retains full-plan targets, rather than supplying correctly labeled short plans.

Objective alignment and short-to-long transfer. We retain the preceding two/three/four-block training banks and compare success-truncated reward with $G = R+(T-\tau)r_\star$, where $T = 5d+250$ for depth d and $r_\star$ is the reward at first success. This credits an absorbing successful state; it does not measure additional executed steps. The 48 training banks contain only sixteen distinct task geometries, reused across depths. Historical unseen-geometry subsets motivate, but do not replace, a new-scene evaluation.

Four seeds fit each target with cell-normalized MSE, AdamW (3×10^{-4}, weight decay .01), batch 16, unit clipping and a fixed 150 epochs, without validation selection. Models are frozen before collecting 32 new Assembly scene pools, one root per pool, each with 125 twelve-block candidates and all nine nonidentity backlash levels (288 cells). No root is selected for headroom, and no task vector coincides with training or historical tests. The prospective primary contrast is G-trained OEWM minus width-256 pooling regret. Intervals use 4,000 crossed seed/root bootstrap draws, retaining all effects per root, and are pointwise rather than multiplicity-adjusted.

Secondary capacity-control protocols, fixed before these outcomes, use 967,681 active parameters for OEWM, 965,505 for width-304 pooling and 981,121 for width-240 GRU. The same-sized no-feedback reader zeros residuals during training and testing but retains prefix actions. Later exploratory controls use OEWM's architecture and innovation on regenerated features, or train all

Table 29: Short-to-long transfer on 32 fresh Assembly scenes. Training uses two/three/four blocks; testing uses twelve, with four seeds. Every row uses G supervision except the explicitly marked original-target control. Dashes denote configurations not run.

	150 epochs		450 epochs	
Reader	G regret	Success (%)	G regret	Success (%)
OEWM, original target	.7304	24.31	—	—
OEWM	.3082	77.69	.2576	81.34
Normalized gate	.2739	80.47	—	—
Additive correction	.3043	77.26	—	—
Pooling, width 256	.5114	50.09	—	—
GRU, width 256	.4280	61.63	—	—
Pooling, capacity-matched	.4952	52.08	.6021	39.67
GRU, capacity-matched	.4060	62.59	.2867	78.99
No feedback	.5318	52.34	—	—
Conditioned reroll	.2914	79.17	—	—

three capacity-matched readers for a fixed 450 epochs; neither selects checkpoints on the opened test bank. Epoch budgets, not FLOPs, are matched.

Objective alignment raises twelve-block OEWM success from 24.31 to 77.69% (Table 29). The primary regret advantage is $-.2032[-.2948, -.1222]$; capacity matching preserves $-.1869[-.2911, -.0896]$ and a success gain of $25.61[15.28, 37.50]$ points. Feedback adds $25.35[11.11, 39.58]$ success points. Optimization changes the GRU comparison: regret differences are $-.0978[-.1526, -.0379]$ at 150 epochs and $-.0291[-.0932, +.0319]$ at 450, with success difference $2.34[-5.73, +11.20]$. Training MSE falls for every arm/seed, but pooling test regret rises. Longer training removes GRU separation, not pooling separation.

Cached-minus-conditioned-reroll regret is $+.0168[-.0115, +.0448]$ at 150 epochs, without equivalence or noninferiority evidence. Hindsight-static capture remains negative ($-416.96[-539.80, -294.71]$ raw G units), with positive seed-mean capture on only one root. A success-optimal static reference reaches 92.36% but uses all test effects and is undeployable (Appendix H.9). Results support finite-length transfer, not the original joint gate criterion, unique architecture necessity or all-path reliability. Twelve-block latency and online MPC are untested.

H.10 Task coverage

All fifty Meta-World (Yu et al., 2020) v3 tasks pass a prespecified two-case collection check with 125 candidates and forked targets. Twenty-seven have less than two points of oracle headroom at 48 banks and are not extended. The other 23 use 72 banks: 48 training, eight validation for width/fallback-margin selection and sixteen held-out evaluation. This headroom selection limits the population studied, not the possible effect of correction.

Each task selects one configuration, unlike the four-seed focused study, using success cross entropy, AdamW 3×10^{-4}, batch 16, 40 epochs and unit clipping. Widths 64/128/256/512 are selected on 3/8/5/7 tasks. Gain intervals bootstrap evaluation banks.

Mean/median success gains are 4.1/3.1 points: twenty of 23 are positive, thirteen with intervals excluding zero; losses are .45/.45/.22 points. Assembly reproduces the focused magnitude (13.6 versus 12.19), but Peg does not (1.1 versus 11.68) despite nominal/oracle baselines only 2.9/.9 points higher. Peg's gain is therefore protocol-specific.

Table 30 includes every selected task. The prespecified training-only predictor, excess response over endpoint rank at 95% spectral energy after common-mode removal, fails to predict gain (Spearman $-.214$, task-clustered 95% CI $[-.608, .215]$).

Table 30: Held-out results on all 23 tasks with oracle headroom, sorted by gain. Each uses 72 banks and a validation-selected width w. Success and gain are percentages and percentage points; intervals are bank clustered.

Task	w	Nominal	OEWM	Oracle	Gain	95% interval
assembly	256	74.3	87.9	90.6	13.6	[11.8,15.8]
soccer	512	82.8	95.3	98.9	12.5	[7.6,18.5]
basketball	512	71.4	81.0	95.8	9.6	[4.9,15.4]
button-press-wall	64	83.0	91.3	98.9	8.3	[−0.0,19.2]
coffee-push	512	84.2	91.7	99.8	7.6	[4.7,10.7]
handle-pull-side	512	78.3	84.6	96.4	6.2	[1.8,10.3]
stick-push	128	82.1	87.9	97.5	5.8	[2.9,8.5]
lever-pull	512	90.8	95.1	100.0	4.2	[2.2,6.2]
push-wall	128	82.4	86.6	95.1	4.2	[2.2,6.0]
sweep	256	87.9	91.7	99.3	3.8	[2.0,5.4]
box-close	128	81.2	84.6	99.1	3.3	[−0.4,9.4]
pick-place	512	84.4	87.5	94.0	3.1	[0.9,5.6]
stick-pull	128	85.7	88.8	99.8	3.1	[0.4,5.8]
pick-place-wall	128	83.7	86.6	92.6	2.9	[1.8,4.0]
hand-insert	256	85.5	88.2	96.4	2.7	[−0.9,6.5]
shelf-place	64	69.9	71.2	97.8	1.3	[0.4,2.5]
sweep-into	512	76.8	78.1	84.4	1.3	[−1.1,4.0]
peg-insert-side	256	79.2	80.4	97.3	1.1	[0.0,2.7]
coffee-pull	256	92.2	93.1	98.4	0.9	[−1.3,3.1]
push-back	128	88.8	89.5	99.1	0.7	[0.0,2.0]
push	64	84.4	84.2	98.0	−0.2	[−0.7,0.0]
disassemble	128	90.4	90.0	96.9	−0.4	[−1.1,0.0]
peg-unplug-side	128	94.9	94.4	100.0	−0.4	[−2.2,1.3]