CROSS-STATE ACTION-EFFECT SUPERVISION: WHICH CORRESPONDENCE MATTERS?

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-state action-effect supervision compares the same action substitution at two starting states. When learners receive the same branch labels, which state correspondences contribute to its training benefit? We decompose the auxiliary loss into common and differential effect residuals and use controls that preserve action identity and match auxiliary-gradient strength. In a controlled dynamical system, reassigning position and velocity within actuator modes retains improvements over paired supervision; exact alignment establishes no advantage over this control. We then vary whether mode and gate affect motion while preserving their visibility and matching permutation cycle lengths. Under the original physics, mode-preserving re-pairing reduces common-active effect error by 23.21% relative to cross-effect-class re-pairing. A paired endpoint comparison establishes dependence on the physical condition, and independent intermediate-condition cohorts show a 14.26% reduction. Cross-class re-pairing also improves over the paired baseline, so preserving mode changes the size of the benefit rather than determining whether any benefit exists. Lower differential error can coexist with greater false response or longer-horizon effect error. Native-system studies also show that response and absolute prediction gains need not coincide, and establish no general control advantage. These findings motivate action-preserving correspondence controls alongside absolute-error and readout checks for relational supervision.

1 INTRODUCTION

World models predict the consequences of candidate actions for planning and behavior learning (Ha and Schmidhuber, 2018; Hafner et al., 2019; 2020; Hansen et al., 2024). The effect of an action substitution can depend on the starting state. Supervising this variation couples predictions across states, but a training gain alone does not identify which part of that coupling helps. When all learners receive the same branch outcomes, an auxiliary loss changes how existing errors are weighted and combined.

Within-state action-effect supervision has direct precedents in Counterfactual Quotient Models and WorldSync (Chen et al., 2026a;b). We compare the same action contrast at two states using four branches. For numerical predictors retaining absolute outcomes, the resulting auxiliary loss emphasizes effect errors that differ across states and leaves their shared component unconstrained. This decomposition motivates separate controls for auxiliary-gradient strength, action identity and state correspondence. The empirical question is which correspondences improve fitting after these quantities are controlled.

We distinguish exact state matching from preserving actuator modes under a physical intervention. Reassigning position and velocity within a mode retains local training benefits even though it breaks exact state alignment. We then change whether mode and gate affect motion while keeping them visible to the predictor. Mode-preserving re-pairing outperforms cycle-matched cross-effect-class re-pairing under the original physics; a paired endpoint interaction and an independent intermediate condition establish physical-condition dependence. Both re-pairing operations keep predictions with their correct labels. They change which residuals are coupled, rather than introducing false branch targets.

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.

The resulting distinction concerns the tested correspondence operations. Mode identity, effect class and input geometry are not independently varied in every comparison. Cross-class re-pairing itself still improves over the paired baseline. A midpoint intervention further shows that improving differential residuals can worsen false responses from the same saved training state. Native-system and readout studies delimit the reach of these findings: response improvements can coexist with absolute prediction failures, and a general control advantage remains unestablished. We use these results to specify what correspondence controls reveal and which further interpretations require different evidence.

2 Related work

Supervising action differences. Counterfactual Quotient Models (CQM) learn action-centered successor features directly (Chen et al., 2026a). Their finite-feature decision guarantee concerns rewards linear in the modeled features. WorldSync aligns differences between action-conditioned video branches sharing the current observation and instruction; its intervention loss acts on flow predictions, while an auxiliary decoder separately predicts trajectories (Chen et al., 2026b). Its full system also expands action coverage. These are direct precedents for within-state effect supervision. We retain absolute numerical outcome predictions and compare the *same action substitution across two starting states*. Action centering removes candidate-shared content within one state; our extra subtraction removes effect residual shared across the two states. Our numerical predictors retain absolute outputs, so this subtraction changes auxiliary error weights within an absolute-prediction objective.

Consistency and derivative information. CoCo imposes inverse-action cycles, zero-action stability, and spatial equivariance (Shi et al., 2026). Our four-branch target instead uses observed outcomes and requires neither an inverse action nor a supplied symmetry. Sobolev training supplements function values with derivatives (Czarnecki et al., 2017). In smooth dynamics, the cross-state target integrates a mixed state–action derivative over a finite rectangle; it also remains defined for discrete or nonsmooth transitions. This interpretation does not turn a small finite-contrast error into a bound on pointwise derivative error.

Predicting auxiliary-loss effects. Differentiation through optimization and forward/reverse hypergradients are established methods (Maclaurin et al., 2015; Franceschi et al., 2017); gradient cosine similarity also guides auxiliary-loss adaptation (Du et al., 2018). Our native study tests whether these local derivatives predict the effect of a finite auxiliary-loss dose.

Supporting prediction controls. Task grouping and robot/environment prediction branches have established precedents (Standley et al., 2020; Anonymous authors, 2025), as do direct action-sequence prediction and multi-horizon supervision (Lin et al., 2025; Benechehab et al., 2023). Our separate numerical grouping study fixes the loss and compares specified partitions with dense prediction. These controlled partition comparisons differ from evaluations of the cited complete systems (Appendix A15).

3 Cross-state action-effect supervision

We define a four-branch target and derive the residual weights that guide our comparisons.

3.1 Absolute outcomes and action effects

Let $F_t(s, u)$ denote the observed outcome after applying action sequence u from a fully specified simulator state s. The predictor receives its native observation $o(s)$; this need not make the observation itself a Markov state. A training family contains two states s_1, s_2 and $K \geq 2$ shared candidate sequences. Write $y_{mi,t} = F_t(s_m, u_i)$ and $\hat{y}_{mi,t}$ for the observed and predicted outcomes in fixed training-standardized coordinates. For an unordered candidate pair $i < j$, define

$$\Delta_{mij,t} = y_{mj,t} - y_{mi,t}, \qquad e_{mij,t} = \hat{y}_{mj,t} - \hat{y}_{mi,t} - \Delta_{mij,t}. \tag{1}$$

The cross-state target is $\Delta_{2ij,t} - \Delta_{1ij,t}$, with residual $e_{2ij,t} - e_{1ij,t}$. A zero target expresses an effect invariant to the state change; it does not require either action effect to vanish. In Meta-World, two candidate sequences are applied from two saved states on the same behavior trajectory; the controlled tasks use eight shared sequences from a gate-flipped state pair. The pairing distribution determines which relationships receive supervision: the loss requires neither proximity nor a known symmetry. Collecting these labels requires exact simulator-state restoration and additional branch queries beyond ordinary offline trajectories. All loss arms within a comparison receive the same collected labels.

3.2 What an auxiliary loss actually weights

Let $\mathcal{L}_{\text{pred}}(t)$ be mean squared state-prediction error per coordinate. Averaging over the same families, action pairs, and selected coordinates, define

$$G_t = \tfrac{1}{2d_e}\,\text{mean}(\|e_{1ij,t}\|^2 + \|e_{2ij,t}\|^2), \qquad D_t = \tfrac{1}{d_e}\,\text{mean}\,\|e_{2ij,t} - e_{1ij,t}\|^2,$$
$$\mathcal{J} = \sum_t w_t(\mathcal{L}_{\text{pred}}(t) + \alpha_t G_t + \beta_t D_t). \tag{2}$$

Here mean averages families and action pairs, while d_e is the number of supervised effect coordinates: six in the fresh continuous controlled studies and seven in the earlier Meta-World coefficient studies. The native forecast study instead uses each system's effective dynamic coordinates (Appendix A14). Prediction alone sets $\alpha_t = \beta_t = 0$; within-state supervision has $\beta_t = 0$. Scales are derived from training data and fixed across the compared loss arms. These coefficients apply to the auxiliary terms; the prediction loss also penalizes branch errors.

3.3 Residual weighting and its required controls

For one action pair and time, set

$$c = (e_1 + e_2)/2, \qquad q = (e_2 - e_1)/2. \tag{3}$$

The common component c is effect error shared across states; q is their differential error. Their mean squared per-coordinate magnitudes are C and Q, with $G = C + Q$ and $D = 4Q$.

Proposition 1 (Residual weighting)**.** *For fixed $\alpha, \beta \geq 0$ and matched nonnegative pair weights,*

$$\alpha G + \beta D = \frac{\alpha}{d_e}\,\text{mean}\,\|c\|^2 + \frac{\alpha + 4\beta}{d_e}\,\text{mean}\,\|q\|^2. \tag{4}$$

Proof sketch. Substitute $e_1 = c - q$ and $e_2 = c + q$. The same weight must apply to both states and their cross-state term within each pair; different pairs may have different weights. The cross terms cancel in G, while $\|e_2 - e_1\|^2 = 4\|q\|^2$.

The added term cannot identify a shared effect error: if $e_1 = e_2 = v \neq 0$, then $D = 0$ while $G = \|v\|^2/d_e > 0$. We retain absolute-prediction and within-state losses as anchors. Although the output residuals decompose orthogonally, their parameter gradients can interact. Shared parameters and optimization can increase either fitted component (Appendix A13).

On the same two-state, K-candidate branch population, using identical coordinates, all unordered candidate pairs and uniform weights, let B_{abs} be the per-coordinate energy of each state's candidate-mean prediction residual. Then $\mathcal{L}_{\text{pred}}^{\text{branch}} = B_{\text{abs}} + (K-1)G/(2K)$ and

$$\mathcal{L}_{\text{pred}}^{\text{branch}} + \alpha G + \beta D = B_{\text{abs}} + \left(\frac{K-1}{2K} + \alpha\right) C + \left(\frac{K-1}{2K} + \alpha + 4\beta\right) Q. \tag{5}$$

The auxiliary terms reweight residual components already supervised by the branches. B_{abs} is shared *across candidate actions within a state*; C is shared *across the two states' action effects*. Readout midpoint error is the error in the mean of two candidate outcomes; its role in distance scores is derived in Appendix A10. Equation 5 requires the matched population and weights: the earlier native prediction loss also covers ordinary trajectories and more coordinates, while executed effect losses use the training-energy normalization specified below.

	Action plan a_i	Action plan a_j
State s_1	$r_{1i} = \hat{y}_{1i} - y_{1i}$	$r_{1j} = \hat{y}_{1j} - y_{1j}$
State s_2	$r_{2i} = \hat{y}_{2i} - y_{2i}$	$r_{2j} = \hat{y}_{2j} - y_{2j}$
	$e_1 = r_{1j} - r_{1i}, \qquad e_2 = r_{2j} - r_{2i},$	$D = \text{mean}\, \|e_2 - e_1\|^2 / d_e$

Operation	Which second-state residual is coupled to e_1?
Aligned A	The original second state in the same family.
Within-mode M	Another second state with the same actuator mode and action library.
Cross-class $M_\times$	Another second state from a different effect class, with the same action library and matched cycle lengths.

Figure 1: Four-branch information and correspondence operations. Every moved prediction retains its own true label; action indices and branch marginals remain fixed. A, M and $M_\times$ change which effect residuals enter the cross-state difference. Cross-class re-pairing therefore still supervises a correctly labeled effect difference.

What must be controlled. Increasing β at fixed α raises only the explicit Q coefficient; scaling G changes both. Common-coefficient controls separate these changes. Auxiliary-gradient norm matching tests strength, but not direction, total gradient or optimizer history. A state-correspondence control must also preserve action identities: permuting candidates answers a broader question. These controls test how the changed weights affect fitted models.

What a state-only correspondence control changes. Within one mode and shared-action block, let x_k, y_k be the two states' effect residuals for a fixed action pair, and let n be the number of families. A uniformly random single-cycle permutation of the second states preserves their marginals and has

$$\mathbb{E}_\pi[D_\pi - D] = \frac{2n}{n-1}\Gamma, \qquad \Gamma = \frac{1}{nd_e}\sum_k \langle x_k - \bar{x}, y_k - \bar{y}\rangle. \tag{6}$$

Writing within-stratum residual variances as V_c, V_q gives $\Gamma = V_c - V_q$. At equal raw coefficients, random re-pairing thus penalizes common variation and weakens the differential-variation penalty; neither term measures the common stratum mean. A fixed cycle also has a signed common–differential cross term (Appendix A3). These identities hold at fixed parameters. During training, the gradient-matched arm also changes its detached multiplier with the current full-batch gradients (Eq. 7), so its fitted errors require empirical evaluation. We therefore test subsequent objectives from the same saved training state, separately from parameter and optimizer history.

3.4 Boundary of a response metric

Exact position differences need not preserve a squared-distance choice. With goal zero, true candidates $(1, 2)$ and predictions $(-2, -1)$ have identical responses but reversed distance scores. The expanded feature $\phi(y) = (y, \|y\|^2)$ admits the linear readout $(-2g, 1)^\top \phi(y)$ up to a candidate-independent constant, so its exact action difference does preserve this score. This distinction is consistent with CQM's decision-sufficiency result for rewards linear in the modeled features. We therefore measure absolute prediction and each specified candidate readout separately from response RMSE. Appendix A10 derives how midpoint and response errors jointly affect a distance margin.

4 Testing which correspondence matters

Within each study, we vary auxiliary-gradient strength, state correspondence and normalization while fixing branch labels, predictor and update count. Figure 1 separates the branch information from the correspondence operation. A cohort is one independently generated dataset and model initialization; its training arms form paired comparisons.

4.1 Action-preserving controls on fresh cohorts

The controlled system contains position, velocity, gate and actuator mode. Paired states initially share position, velocity and mode, with gates zero and one. Five modes implement independent, re-

dundant, gated, saturated and weak actuation (Appendix A13). Each cohort independently generates 2,048 training and 1,024 evaluation families, action libraries and initialization. Each family has two starts and eight plans; a complete 32-family batch shares one action library.

A fixed permutation π rearranges second states within each mode and library. It changes position/velocity correspondence while preserving the complete prediction-loss and within-state-effect populations. At a fixed mode and gate, true effects in these dynamics are invariant to base position and velocity. The experiment asks whether breaking this exact correspondence changes fitting under an invariant true effect.

Seven arms receive 2,000 updates. Executed $\widetilde{G}, \widetilde{D}$ average $d_e G_t / Z_{G,t}, d_e D_t / Z_{D,t}$ over steps 1–4, with training-only true-energy divisors. The primary arms are paired P ($\mathcal{L}_{\text{pred}} + \widetilde{G}$), aligned A (P plus $\lambda\widetilde{D}$), raw mismatch U, gradient-matched mismatch M, and gradient-matched added-strength S. P is an arm name; $\mathcal{L}_{\text{pred}}$ always denotes prediction loss. M and S multiply their added term L by

$$\gamma_L = \text{stopgrad}\left(\lambda \frac{\|\nabla_\theta \widetilde{D}\|}{\|\nabla_\theta L\|}\right), \qquad L \in \{\widetilde{D}_\pi, \widetilde{G}\}. \tag{7}$$

This matches A's auxiliary norm at each arm's current model and batch before total-gradient clipping. Gradient direction and the resulting Adam trajectory remain free to differ. A zero target norm sets $\gamma_L = 0$; the matched mismatch is therefore not an always-active common-error penalty. Sixteen training blocks fix initial λ. Two normalization sensitivities replace steps 1–4 with 1–16 for training statistics, including the effect-energy divisors, while retaining labels, initialization, batches and short-window λ. Every arm is scored in the same short-window reference scale (Appendix A1).

4.2 Intervening on the physical relevance of state

A new study changes acceleration to $b_\eta = (1-\eta)(u_1 + u_2, 0) + \eta b_{\text{original}}$ for command (u_1, u_2). Mode and gate remain observed but do not affect motion at $\eta = 0$; $\eta = 1$ recovers the original physics. Eighteen fresh cohorts pair these endpoints; eighteen independent cohorts test $\eta = 0.5$. Each uses four arms: paired P, aligned A, mode-preserving M, and cross-effect-class $M_\times$. The 36 independent cohorts produce 216 fits, with cohort-level inference rather than treating fits as independent repetitions.

M permutes within each mode. $M_\times$ crosses the second-state ($g = 1$) effect classes $\{0\}, \{1,2\}, \{3\}, \{4\}$, with identical permutation cycle lengths and action/label marginals. Both match A's local auxiliary norm by Eq. 7. Modes 1 and 2 share an effect class at this gate, so preserving mode is more restrictive than preserving effect class. Mode 4 also has a narrower initial horizontal-position range, $[-1, .5]$ rather than $[-1, 1]$; the mode/input-geometry association remains at every η. Changing the physics with the same input design identifies dependence on physical condition, while leaving the separate effects of mode identity, effect class and input geometry unresolved. All worlds use the cohort's $\eta = 1$ training scales and calibration. Appendix A2 specifies reference access, eligibility and all comparisons.

4.3 Cross-system scope and measurement conventions

Separate native studies use twelve complete Cartpole/Reacher/Push cohorts and an earlier ten-task Meta-World coefficient bridge with five complete repetitions. Their observations, pairing and architectures differ from the controlled experiment. They test training tradeoffs and optimizer-path forecasts, rather than repeating the mode-preserving versus cross-effect-class intervention (Appendices A5 and A8). A further grouping study fixes prediction loss and varies output partitions.

The algebra uses per-coordinate mean energies. Controlled C, Q endpoints instead sum four motion-coordinate energies, whereas the native forecast study uses means over effective dynamic coordinates. Ratios retain their own study's convention. Appendix A14 lists these coordinate choices and recorded computational costs. Equal 2,000-update budgets define the fitting comparison; they do not imply equal wall time or convergence.

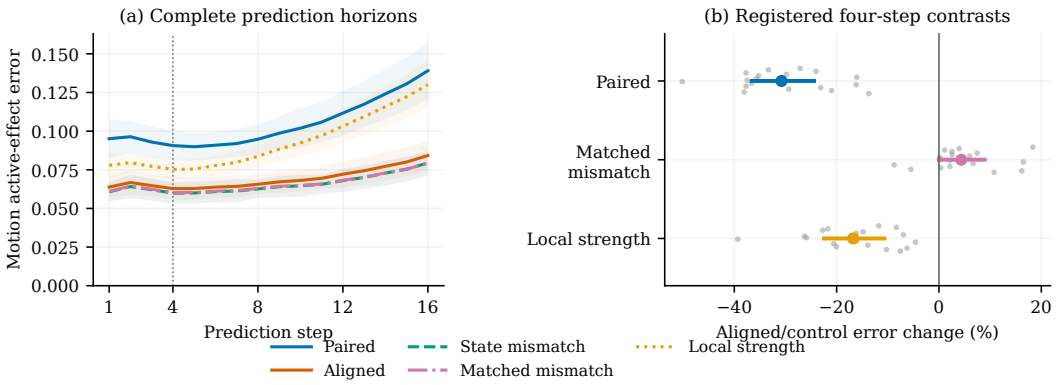

Figure 2: Fresh action-preserving correspondence study. (a) Motion active-effect error through every prediction step; lines and shading are arithmetic means and sample standard deviations over all eighteen cohorts. Loss and normalization use steps 1–4 only. (b) All three primary log-ratio comparisons transformed to error change, with corrected t_{17} intervals and all repetition points. Negative favors aligned supervision. The matched-mismatch comparison does not resolve an alignment advantage.

5 Experiments

Appendix Table 1 distinguishes populations and inference units.

5.1 Action-preserving correspondence on fresh cohorts

Active-effect error E_{active} is the root ratio of squared response residual to true nonzero-effect energy. False response F_{equiv} is the RMS predicted difference between distinct command prefixes with identical native paths. Both are lower-is-better and use the four motion coordinates in the fresh study. Eligibility is computed from native truth; static gate/mode outputs do not contribute to these reported errors (Appendix A13).

Eighteen independently regenerated data/model cohorts are paired across seven arms. Three primary four-step active-error log ratios use corrected t_{17} intervals; secondary and normalization tests retain their separate corrections in Appendix A1. The cohort count and comparisons are fixed before results.

Aligned supervision lowers four-step active error by 30.77% [24.42%,36.59%] relative to paired supervision and 16.78% [10.67%,22.47%] relative to per-update strength. Against gradient-matched state mismatch, its error is 4.37% higher, with a corrected interval of [−0.0155%,8.94%]. This comparison resolves neither an alignment advantage nor equivalence. Figure 2 retains all eighteen repetitions and all five short-normalization arms.

The differential component improves without a corresponding response advantage. Relative to matched mismatch, aligned supervision lowers four-step Q by 17.34% [11.51%,22.78%], yet increases four-step false response by 22.32% [4.48%,43.20%] and sixteen-step active error by 5.99% [1.14%,11.08%], all with the secondary-family correction. Thus lower differential error need not improve the effects themselves. The reported motion Q_4 differs from the full six-coordinate, time-weighted training objective.

These gains use only four training steps for normalization. Longer normalization raises P's four-step error by 11.51% [8.12%,15.00%]; A's change is unresolved. The normalization interaction is resolved at four steps but not sixteen. Appendix A1 retains all 52 contrasts. These are local fitting gains with paired branch access; M still uses both the paired base loss and the aligned auxiliary norm reference. Earlier candidate-permutation evidence changed action identity and used longer normalization (Appendix A15).

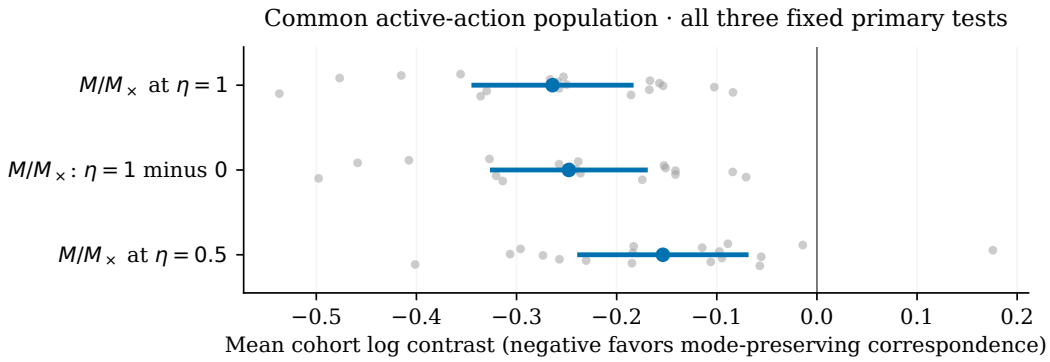

Figure 3: All three physical-correspondence primary tests. Points show all eighteen cohort contrasts per test; bars are simultaneous 95% intervals. Negative favors mode-preserving M over cycle-matched cross-class $M_\times$. The middle contrast is the paired difference of endpoint log ratios; the intermediate condition uses different cohorts. The common-active population is fixed across physical conditions.

5.2 The benefit of mode preservation changes with physics

The second study fixes the evaluated action pairs across physical conditions: $E_{\cap,4}$ uses the intersection of pairs active at both $\eta = 0$ and 1, with each world's true-effect energy as denominator. This denominator cancels in the within-world M/$M_\times$ log ratio. The three primary contrasts use simultaneous t_{17} intervals corrected within their own family (Appendix A2).

Mode-preserving M lowers $E_{\cap,4}$ by 23.21% [16.93%,29.01%] relative to cross-class $M_\times$ at $\eta = 1$. Its advantage is greater than at $\eta = 0$: the paired log-ratio difference is -0.24775 $[-0.32415, -0.17136]$, an interaction rather than a direct error reduction. At $\eta = 0.5$, eighteen independent cohorts give a 14.26% [6.82%,21.10%] reduction. All three predeclared contrasts are resolved (Figure 3), establishing physical-condition dependence for this correspondence operation. Monotonicity and equivalence at $\eta = 0$ remain untested.

The separately corrected secondary comparisons reveal tradeoffs among error components. M improves world-specific active error over P in all three worlds, so exact position/velocity alignment *in the auxiliary term* is unnecessary for these gains; paired data, base P and A's norm reference remain. $M_\times$/P also improves in all worlds. At $\eta = 1$, A/M raises active error by 5.71%, false response by 27.61% and C_{16} by 17.67%, yet lowers Q_4 by 14.36%, all resolved. At $\eta = 0.5$, A/M active and differential differences remain unresolved. Class preservation is therefore not necessary for every benefit over P. The observed advantage of M over $M_\times$ concerns the size of the gain under these operations, with different effects on active, false-response and common errors.

5.3 Separating inherited state from the future objective

At update 1,000, an eight-cell intervention crosses parameter source, Adam-history source and future objective, each A or M. All eighteen original cohorts continue the remaining 1,000 updates. These are post-result interventions; all 36 unchanged AAA/MMM paths exactly reproduce their originals. Labels follow parameter/Adam/objective order, and final-update comparisons retain separate primary and secondary correction families (Appendix A4).

Changing parameter source from M to A raises final C_{16} by 19.16% [9.51%,29.67%], averaged over the other assignments; changing Adam source lowers it by 1.28% [0.82%,1.74%]. All four future-objective effects remain unresolved, averaging -0.42% [-5.24%,4.64%]. The prescribed swaps reveal history dependence, including possible effects of parameter/optimizer incompatibility; they do not estimate natural mediation shares.

Continuing with A lowers Q_4 by 12.73% [10.10%,15.27%] but raises F_4 by 11.53% [6.53%,16.76%]; all four starting assignments have the same resolved directions. The parameter-source increase in E_{16} is 8.25% [2.71%,14.09%], with future-objective effects unresolved. Resolved

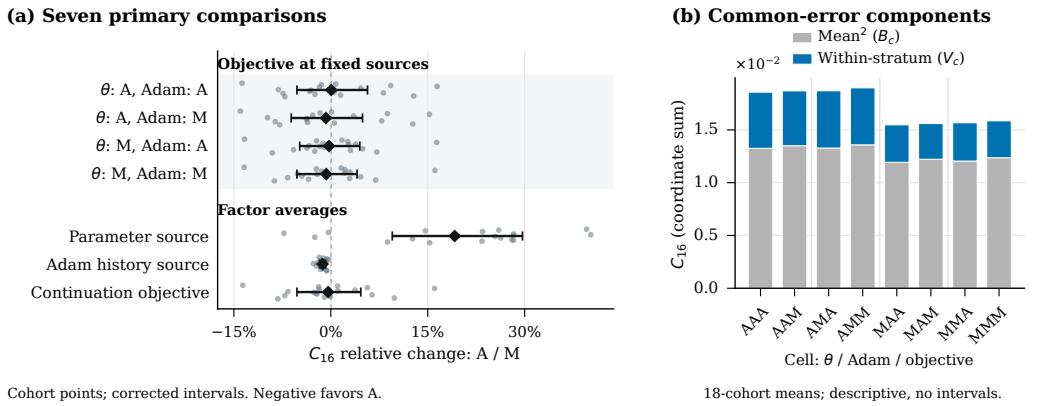

Figure 4: Fixed-midpoint training-state intervention. (a) All seven primary C_{16} contrasts, with their seven-test corrected intervals and all eighteen cohort points. Positive means higher error for the A level. Four simple effects change only the future objective; the final three average one factor over the other assignments. (b) All eight final cells' common mean-squared and within-stratum variance components, averaged over eighteen cohorts. This saved-prediction decomposition is descriptive; components sum to coordinate-summed C_{16}.

Adam–objective and three-factor interactions for C_{16}, E_{16} and a parameter–objective interaction for F_4 preclude an additive account. The appendix reports all 44 comparisons.

5.4 Cross-system evidence and measurement limits

The twelve-cohort native study uses Cartpole, Reacher and Push with four arms: paired B, aligned A, state-reassigned M and added-strength S. Here B is the native paired arm, distinct from the residual energy B_{abs}. Descriptive A/B response-RMSE changes are -5.49%, -39.70% and $+1.82\%$, respectively, while absolute-state RMSE changes are $+30.82\%$, $+13.04\%$ and $+13.25\%$. A/M response changes are $+0.69\%$, -1.00% and -0.59%. These measurements use unadjusted descriptive intervals and do not constitute a confirmatory family of training-efficacy tests (Appendix A6). They show why a response improvement and an absolute-prediction improvement must be distinguished.

The earlier ten-task coefficient bridge isolates adding differential weight while retaining common weight. SAdd/SG reduces held-out response RMSE by 0.133% [0.087%,0.179%]. Weakening common weight afterward instead gives a reduction of -0.354% [-0.604%,-0.106%]. SAdd's reductions against strength and mismatch controls are 0.055% [$-0.012\%, 0.122\%$] and 0.020% [$-0.179\%, 0.219\%$]. This small coefficient-specific gain does not establish an additional correspondence advantage. Appendix A8 retains all five corrected comparisons and repetitions.

Separate readout and input-scale studies locate practical limits of response metrics. None of five focal method contrasts establishes a reward benefit. A known Reach cell combines 3.077 mm first-block response RMSE with 1109.417 mm absolute error; a fixed input-scale intervention reduces the latter to 8.847 mm. A direct predictor also fails before recursive feedback. These selected-cell diagnostics reveal preprocessing-sensitive tails and residual failures; they neither replace full-suite results nor identify a general architecture failure (Appendices A10 and A11).

Prospective optimizer-path forecasts provide a further interpretation boundary. None of five corrected tests establishes an advantage for extrapolating a zero-dose derivative to a full auxiliary-loss intervention. Large unit-dose error persists on validation data in double precision, while small-dose numerical exceptions remain. This result concerns the tested linear extrapolation rule; it does not identify a unique failure mechanism for sensitivity-based prediction. The complete forecast comparisons and precision analysis are in Appendices A6 and A7.

6 DISCUSSION

The controlled studies distinguish exact position/velocity correspondence from the state relations preserved by mode-restricted re-pairing. Exact matching is unnecessary for a local auxiliary-loss benefit in dynamics where true effects are invariant to those coordinates. Comparing mode-preserving with cross-effect-class re-pairing then reveals dependence on the physical role of observed state variables. Correct labels and matched branch marginals make this a result about residual coupling under fixed information.

The physical interaction supplies empirical information beyond the residual identity, but its interpretation retains a specific boundary. Mode identity and effect class are not fully separated, and mode remains associated with input geometry. A fixed learner, reference-world normalization and a 2,000-update budget further define the result's scope. The experiments establish the behavior of these operations, rather than a universal effect-class mechanism or a deployment rule for choosing pairs. Paired labels and the aligned norm reference remain required by the tested controls.

Error components also respond differently. The midpoint intervention preserves a tradeoff between differential error and false response from the same saved training state. Native response gains can coexist with worse absolute prediction, and readout studies establish no general reward benefit. These observations motivate reporting active effects, false responses, absolute errors and the intended decision readout together. The additional grouping, precision and input-scale analyses retain their distinct populations in the appendix.

7 CONCLUSION

Cross-state supervision changes how correctly labeled branch errors are coupled. In the tested controlled system, within-mode re-pairing retains gains without exact position/velocity alignment, and its advantage over cross-effect-class re-pairing changes with the physics. The accompanying error tradeoffs show why correspondence controls must preserve action identity and account for auxiliary strength while evaluating more than the optimized differential residual. Together, these results provide a controlled way to assess which state relations matter for a particular learner and training budget.

AI USE STATEMENT

The authors proposed the initial research ideas and further refined them with AI assistance. AI tools also assisted mathematical derivations and proofs, experimental design, method and data-pipeline implementation, result interpretation, literature discovery, language polishing, and figure and artifact preparation. Synthetic targets were generated by the specified simulator or closed-form dynamics; reported measurements were computed from saved experimental outputs. Recorded checks include independent saved-statistic reconstructions, targeted algebra and implementation checks, and primary-source citation checks. The verification covers these recorded computations and source checks. The authors are responsible for the final text, claims, code, and citations.

REPRODUCIBILITY STATEMENT

The anonymous code and data supplements supply sources, frozen configurations, complete endpoint populations, model states and executable reconstruction commands. The main CPU entry reconstructs each study's original statistical family; separate commands replay final models or rerun training. Appendix A16 describes these reproduction levels. Complete-population final-weight inference was replayed for the new controlled studies and forecast diagnostic, separately from training. Native forecast inference is self-contained, whereas full native training and simulator recollection retain separate runtime requirements. A companion archive supplies the four-cell, 24-ensemble Reach scale/feedback experiment. Historical raw data and weights remain separate; original native final weights lack Adam state. The study READMEs specify inputs, commands, tested scope and limitations.

REFERENCES

Anonymous authors. Pre-training robo-centric world models for efficient visual control. Anonymized ICLR 2025 submission on OpenReview, 2025. URL https://openreview.net/forum?id=DJw1JBTmuk. Project and released implementation: https://robo-centric-wm.github.io/.

Abdelhakim Benechehab, Giuseppe Paolo, Albert Thomas, Maurizio Filippone, and Balázs Kégl. Multi-timestep models for model-based reinforcement learning. *arXiv preprint arXiv:2310.05672*, 2023. URL https://arxiv.org/abs/2310.05672.

Junlin Chen, Ruijie Wang, and Jianxin Li. Counterfactual quotient models: Learning what actions change, not what the world does. *arXiv preprint arXiv:2608.22092*, 2026a. URL https://arxiv.org/abs/2608.22092.

Sixiang Chen, Jiaming Liu, Jixian Wu, Yichen Guo, Tinghao Wang, Siyuan Qian, Hao Chen, Jiajun Cao, Jian Tang, and Shanghang Zhang. Do robotic world models really follow actions? diagnosing and aligning action-conditioned generation for policy learning, 2026b. URL https://arxiv.org/abs/2608.24885.

Wojciech M. Czarnecki, Simon Osindero, Max Jaderberg, Grzegorz Swirszcz, and Razvan Pascanu. Sobolev training for neural networks. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/758a06618c69880a6cee5314ee42d52f-Paper.pdf.

Yunshu Du, Wojciech M. Czarnecki, Siddhant M. Jayakumar, Mehrdad Farajtabar, Razvan Pascanu, and Balaji Lakshminarayanan. Adapting auxiliary losses using gradient similarity. *arXiv preprint arXiv:1812.02224*, 2018. URL https://arxiv.org/abs/1812.02224. Revised 2020.

Luca Franceschi, Michele Donini, Paolo Frasconi, and Massimiliano Pontil. Forward and reverse gradient-based hyperparameter optimization. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1165–1173. PMLR, 2017. URL https://proceedings.mlr.press/v70/franceschi17a.html.

David Ha and Jürgen Schmidhuber. Recurrent world models facilitate policy evolution. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/2de5d16682c3c35007e4e92982f1a2ba-Paper.pdf.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 2555–2565. PMLR, 2019. URL https://proceedings.mlr.press/v97/hafner19a.html.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1lOTC4tDS.

Nicklas Hansen, Hao Su, and Xiaolong Wang. TD-MPC2: Scalable, robust world models for continuous control. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://arxiv.org/abs/2310.16828.

Haoxin Lin, Yu-Yan Xu, Yihao Sun, Zhilong Zhang, Yi-Chen Li, Chengxing Jia, Junyin Ye, Jiaji Zhang, and Yang Yu. Any-step dynamics model improves future predictions for online and offline reinforcement learning. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/b0506debbf49e31d25690fbd1e69cd2f-Abstract-Conference.html.

Dougal Maclaurin, David Duvenaud, and Ryan Adams. Gradient-based hyperparameter optimization through reversible learning. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2113–2122. PMLR, 2015. URL https://proceedings.mlr.press/v37/maclaurin15.html.

Yuhong Shi, Zhenhao Chu, Jie Wei, Jun Hao, Jianyi Liu, and Jingwen Fu. Overcoming statistical bias in action-controllable world models. *arXiv preprint arXiv:2608.04653*, 2026. URL https://arxiv.org/abs/2608.04653.

Trevor Standley, Amir Zamir, Dawn Chen, Leonidas Guibas, Jitendra Malik, and Silvio Savarese. Which tasks should be learned together in multi-task learning? In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9120–9132, 2020. URL https://proceedings.mlr.press/v119/standley20a.html.

Table 1: Which measurement supports each conclusion? Studies keep their own populations and scales. Task branches and ensemble members are never counted as independent training repetitions.

Evidence	Population / unit	Reported quantity	Inference
Fresh state correspondence	Eighteen independent data/model repetitions	Motion errors, all 16 steps; primary step 4	Three-test primary; 37 secondary; 12 normalization, separate t_{17} families
Physical correspondence	36 new cohorts: 18 paired endpoints / 18 intermediate	Common-active E_4 primary; world-specific secondary	Three primary / 60 secondary tests in separate t_{17} families
Midpoint path intervention	The same eighteen cohorts; eight continuations each	Final-update C_{16} primary; Q_4, F_4, E_{16} secondary	Post-result follow-up; seven primary / 37 secondary tests in separate families
Prospective native forecasts	Twelve new three-system cohorts; four arms	Step 4 log-metric forecasts; h1–16 descriptive	Five-test family; simultaneous t_{11} intervals
Forecast decomposition	Same twelve cohorts; 396 FP64 fits	Unit-dose, V/E and precision discrepancies; both signed probes	Post-result descriptive analysis; no new efficacy test
Original controlled effects	Five seeds; 1,024 state families	Original full-state / new motion-only errors, steps 4 and 16	Paired seed/family bootstrap; original 28 / new 30 tests
Native coefficients	Ten fixed tasks; five complete repetitions	Geometric response-RMSE reduction, four horizons	Five-test family; 99% intervals
Output grouping	Nine fixed tasks; five complete repetitions	Geometric response-RMSE reduction, four horizons	Eight-test family; 99.375% intervals
Readout diagnosis	All 300 loss / 180 recursive grouping models	Physical MSE and geometry/reward regret; step 20 display	Repetition-level descriptive 95% intervals
Input-scale sensitivity	Four Reach cells; twelve matched fits	Absolute and response RMSE in millimeters	Fixed-cell intervention; no suite-level confirmation

EVIDENCE AND REPRODUCTION MAP

The appendix gives the designs, complete central contrast families and interpretation needed to assess the paper. The accompanying anonymous data/code package retains every original comparison, cohort and failure. Earlier exploratory studies are summarized separately rather than pooled as evidence for one mechanism. All table signs and units are stated locally.

For each eighteen-cohort contrast in these controlled studies, $\bar{d} \pm t_{17,1-.05/(2m)} s_d/\sqrt{18}$ is the corrected interval for a paired log contrast. A direct metric change is $100(e^{\bar{d}} - 1)$; interactions remain log contrasts. Correction families are separate: 3/37/12 in the fresh study 7/37 in the midpoint study, and 3/60 in the physical study. They do not give a campaign-wide 95% guarantee. The native forecast study instead uses twelve complete three-system cohorts and five tests. All original data, negative results and undefined qualifications remain in their respective populations. A confidence interval containing zero does not establish absence or equivalence.

A1 ACTION-PRESERVING CORRESPONDENCE: DESIGN AND COMPLETE RESULTS

Population and controls. Each of eighteen independent repetitions regenerates 2,048 training and 1,024 evaluation families, action libraries, permutations, initialization and batch stream. Every family contains two six-coordinate states sharing base position, velocity and actuator mode, with gates zero and one, and eight action plans. Each 32-family block shares its plans. A fixed single-cycle permutation moves second-state predictions together with their true labels within each mode/library

Table 2: All three primary four-step motion comparisons. Entries are relative metric changes of A against the indicated control; negative values favor A. Intervals are transformed paired t_{17} intervals with three-comparison Bonferroni correction. The near-zero endpoint for A/M is retained at four decimals.

Ratio	Metric	Change (%)	Corrected interval (%)
A/P	E_{active}	-30.77	[-36.5916, -24.4188]
A/M	E_{active}	+4.37	[-0.0155, +8.9429]
A/S	E_{active}	-16.78	[-22.4684, -10.6697]

stratum; it preserves action identities, gate direction, all marginal branches and the original $\mathcal{L}_{\text{pred}}, G$ population. One update uses a complete block, identically across arms. In these prescribed dynamics, true action effects at fixed mode/gate are independent of the re-paired base position/velocity. Thus this tests residual correspondence in an invariant regime, not destruction of every physical state relation or identification of a gate mechanism.

P denotes paired $\mathcal{L}_{\text{pred}}+\widetilde{G}$; A adds $\lambda\widetilde{D}$; U adds $\lambda\widetilde{D}_\pi$; M and S add respectively $\widetilde{D}_\pi$ and $\widetilde{G}$, matching the auxiliary-gradient norm of $\lambda\widetilde{D}$ at their own parameters and batch. The detached multipliers in Equation 7 precede total clipping; they do not match direction, Adam state or actual parameter displacement. Zero target norm sets the multiplier to zero; a positive target with zero denominator is an explicit failure, with no cap or epsilon substitute. All 126 fits completed without that failure. The maximum logged all-update scalar norm-match relative residual is 1.15×10^{-7}, rounded upward.

The tilded terms sum six-coordinate squared residuals, divide by training-only true-energy scales and average supervised steps 1–4. $\mathcal{L}_{\text{pred}}$ averages absolute branch error. Each repetition fixes λ to the median of sixteen initial training-block values $0.25\|\nabla(\mathcal{L}_{\text{pred}} + \widetilde{G})\|/\|\nabla\widetilde{D}\|$. The residual MLP has two width-128 SiLU layers and is applied recursively. All seven arms use 2,000 FP32 AdamW updates, no AMP/TF32, learning rate 10^{-3}, weight decay 10^{-3}, betas $(.9, .999)$, epsilon 10^{-8}, norm clipping at one, one-percent warmup and cosine decay: 252,000 total updates.

Normalization and metrics. The primary five arms fit normalization on training initial states and steps 1–4. P16/A16 use steps 1–16, with the same initialization, batch sequence, four-step supervision and original short-window λ. They also recompute the four supervised energy denominators in their changed standardized coordinates; this is the complete normalization treatment. Standard deviations at most 10^{-8} become one; energy divisors are floored at 10^{-4}. Each model is unscaled and measured using the same short-window reference, preventing a changed evaluation ruler.

Every formal endpoint uses four motion coordinates and original aligned evaluation pairs. E is qualified relative active-effect error; F is RMS false response on distinct prefixes with identical true paths (Appendix A13). C, Q, G sum the four coordinate energies, average all 28 unordered action pairs and retain $C + Q = G$; they are four times the per-coordinate energy convention. Full-six-coordinate and per-mode results are descriptive. Each inferential unit is one complete regenerated data/model repetition, not a family, branch or checkpoint.

The eighteen-cohort size was fixed before results: the earlier five-seed maximum log-ratio SD .11962 gave a one-sided 90% planning bound .23197 and a planned primary log half-width .1452. This neither guarantees realized precision nor powers the new mismatch contrast. The three separate corrected families have 3, 37 and 12 comparisons. All 52 remain defined and are given below; all sixteen horizons and all eighteen values are supplied in the package.

Interpretation. A improves over P and local strength S but has no resolved primary advantage over M. The A/P and A/S changes favor A in all eighteen cohorts; A/M is adverse in sixteen. These counts are descriptive. Differential-error reduction coexists with greater false response at four steps and greater active/common error at sixteen steps against both mismatch controls. Against S, four-step false response and common error remain unresolved. Thus the correspondence and strength controls answer different questions.

The normalization interaction is $\log(M_{A16}/M_{P16}) - \log(M_A/M_P)$. Only its four-step active-error interval excludes zero. Both corresponding direct point estimates increase error, more strongly for

Table 3: Complete four-step secondary family entries (17 of 37). Relative changes and transformed corrected intervals use the same 37-comparison family as Table 4. The other three four-step active-effect comparisons belong to the primary family.

Ratio	Metric	Change (%)	Corrected interval (%)
A/P	F_{equiv}	-36.37	[-49.7641, -19.4107]
A/P	C	-39.49	[-53.0463, -22.0102]
A/P	Q	-94.62	[-95.9858, -92.7882]
A/P	G	-52.60	[-63.3187, -38.7553]
A/U	E_{active}	+4.86	[-1.3477, +11.4664]
A/U	F_{equiv}	+23.08	[+4.5006, +44.9571]
A/U	C	+12.36	[-1.4219, +28.0630]
A/U	Q	-13.61	[-20.7327, -5.8517]
A/U	G	+11.43	[-1.8846, +26.5434]
A/M	F_{equiv}	+22.32	[+4.4836, +43.2007]
A/M	C	+11.40	[-2.4794, +27.2615]
A/M	Q	-17.34	[-22.7787, -11.5146]
A/M	G	+10.34	[-3.0566, +25.5784]
A/S	F_{equiv}	-13.78	[-27.5117, +2.5527]
A/S	C	-17.92	[-33.9466, +1.9882]
A/S	Q	-89.55	[-91.2458, -87.5355]
A/S	G	-30.59	[-43.1723, -15.2213]

Table 4: Complete sixteen-step secondary family entries (20 of 37). Negative relative change favors A. The endpoint is beyond the loss-supervised horizon; the primary five arms also exclude steps 5–16 from their normalization statistics.

Ratio	Metric	Change (%)	Corrected interval (%)
A/P	E_{active}	-39.27	[-46.9342, -30.5025]
A/P	F_{equiv}	-56.36	[-69.3303, -37.9010]
A/P	C	-48.25	[-60.5698, -32.0688]
A/P	Q	-90.57	[-93.3677, -86.5788]
A/P	G	-66.24	[-74.0121, -56.1469]
A/U	E_{active}	+5.98	[+1.2656, +10.9216]
A/U	F_{equiv}	+12.44	[+0.6132, +25.6548]
A/U	C	+16.85	[+6.6920, +27.9649]
A/U	Q	-7.39	[-25.8846, +15.7123]
A/U	G	+13.13	[+3.1231, +24.1150]
A/M	E_{active}	+5.99	[+1.1358, +11.0815]
A/M	F_{equiv}	+12.16	[-0.0876, +25.9193]
A/M	C	+17.26	[+6.4597, +29.1563]
A/M	Q	-9.43	[-28.0708, +14.0527]
A/M	G	+13.16	[+2.8310, +24.5261]
A/S	E_{active}	-35.20	[-40.8454, -29.0173]
A/S	F_{equiv}	-58.94	[-72.3206, -39.0987]
A/S	C	-42.65	[-54.7814, -27.2756]
A/S	Q	-89.88	[-92.3819, -86.5596]
A/S	G	-63.07	[-70.5697, -53.6694]

P; the interaction does not mean that long-window normalization improves absolute accuracy. The short-window A/P and A/S gains therefore do not require longer normalization, without establishing normalization invariance. The study was fixed after earlier results but before these new data/models were fitted or scored; it is new-data evidence within the same dynamics, not unseen-environment transfer.

Table 5: Complete normalization family. All entries are mean paired log contrasts with twelve-comparison corrected intervals. P16/P and A16/A compare full-16 against short-4 normalization; the interaction is the difference of those log ratios. Interactions are not RMSE percentage reductions. Negative direct ratios favor the full-16 arm.

Contrast	Metric / step	Mean log contrast	Corrected interval
P16/P	E_{active} / 4	+0.10891	[+0.07803, +0.13979]
P16/P	F_{equiv} / 4	+0.20978	[+0.11096, +0.30859]
P16/P	E_{active} / 16	+0.10780	[+0.07094, +0.14467]
P16/P	F_{equiv} / 16	+0.09465	[-0.06799, +0.25729]
A16/A	E_{active} / 4	+0.03973	[-0.00575, +0.08520]
A16/A	F_{equiv} / 4	+0.05899	[-0.05032, +0.16829]
A16/A	E_{active} / 16	+0.09315	[+0.04426, +0.14204]
A16/A	F_{equiv} / 16	-0.00834	[-0.12761, +0.11093]
Interaction	E_{active} / 4	-0.06918	[-0.13166, -0.00671]
Interaction	F_{equiv} / 4	-0.15079	[-0.31538, +0.01380]
Interaction	E_{active} / 16	-0.01465	[-0.08305, +0.05374]
Interaction	F_{equiv} / 16	-0.10299	[-0.27952, +0.07354]

A2 PHYSICAL RELEVANCE OF CORRESPONDENCE: FIXED DESIGN AND ALL TESTS

The study was frozen after the first correspondence and native-forecast results, before generating its own data or reading endpoints. Root seed 260926983 generates 36 new cohorts: repetitions 0–17 pair $\eta = 0, 1$; 18–35 test only $\eta = .5$. Each world uses 2,048 training and 1,024 evaluation families, two starts, eight shared plans and sixteen steps. The four arms receive 2,000 updates each: 216 fits and 432,000 updates. Architecture, optimizer, FP32 execution and four-step loss are as in Appendix A1; all worlds use their cohort's $\eta = 1$ first-four-step training scales and initial sixteen-block calibration. No normalization uses later steps or evaluation labels.

Physical and pairing intervention. Acceleration is $b_\eta = (1 - \eta)(u_1 + u_2, 0) + \eta b_{\text{original}}$, with original damping .9 and integration step .1. Modes and gate are visible at all doses. Initial position and velocity are uniform on $[-1, 1]^2$ and $[-.2, .2]^2$; modes cycle 0–4, with mode-4 horizontal position redrawn on $[-1, .5]$. Thus mode/input geometry remains correlated even at $\eta = 0$. M makes one cycle within each mode in a 32-family library block. $M_\times$ has the identical five cycle lengths, but every edge crosses one of the second-state ($g = 1$) effect classes $\{0\}, \{1, 2\}, \{3\}, \{4\}$. Fixed random-priority backtracking assigns classes, then random within-class assignments select indices; this is not uniform sampling over all admissible permutations. Both are bijections without fixed points; predictions move with their correct labels, preserving action identity, branch marginals and full $\mathcal{L}_{\text{pred}}/G$ populations. Each uses its own current A auxiliary-gradient norm as reference before clipping/AdamW, not equal cross-arm final updates.

Reference access and primary population. Closed-form $\eta = 0, 1$ train/evaluation labels are generated also for intermediate cohorts, without reference-world models. Training $\eta = 1$ labels set scales and calibration; evaluation references only define the common eligibility mask. For action-pair/state/time index i, let I_i indicate a nonzero true response at both endpoints. The common-active metric is

$$E_{\cap,t}(\eta) = \left(\frac{\sum_i I_i \|\hat{\Delta}_i(\eta) - \Delta_i(\eta)\|^2_{\text{motion}}}{\sum_i I_i \|\Delta_i(\eta)\|^2_{\text{motion}}} \right)^{1/2}.$$

The index sums over all evaluation families, both starts and 28 action pairs at step t; scales are the shared training reference. Qualification uses exact physical equality, not a fitted threshold. This excludes pairs inactive at either endpoint and is not the full world-specific active population. The denominator cancels in each M/$M_\times$ log ratio. Reference access makes this a controlled experiment, not a deployable selector. Generating all 90 cohort/reference populations requires 70,778,880 closed-form branch steps; native environment calls are zero. No interim scientific outcome changes the fixed matrix.

All three primary intervals in Figure 3 use eighteen-cohort paired t_{17} inference with three-test Bonferroni correction. The endpoint interaction is a difference of within-cohort log ratios, not a direct percentage error change; the independent midpoint does not prove monotonicity. The secondary family below has its own 60-test correction. Its E_4 uses each world's original active set, F_4 uses equivalent paths, and Q_4, C_{16} use aligned evaluation pairs. It therefore complements rather than replaces the common-active primary tests. All four metrics use motion coordinates; full-state and sixteen-horizon curves are descriptive.

Table 6: Effect-error half of the 60-test secondary family (30 comparisons). Entries are mean cohort log metric ratios [simultaneous 95% interval], corrected with all 60 using t_{17}. Negative favors the numerator. $M_\times$ is `effect_crossing` in the package. E_4 uses world-specific active pairs; F_4 uses equivalent paths.

η	Contrast	E_4	F_4
0	A/P	−0.168 [-0.195,-0.141]	−0.184 [-0.225,-0.144]
0	A/M	0.167 [0.131,0.203]	0.202 [0.159,0.246]
0	$A/M_\times$	0.151 [0.117,0.185]	0.195 [0.151,0.238]
0	M/P	−0.335 [-0.371,-0.299]	−0.387 [-0.430,-0.343]
0	$M_\times/P$	−0.319 [-0.350,-0.289]	−0.379 [-0.416,-0.342]
1	A/P	−0.363 [-0.475,-0.252]	−0.570 [-0.838,-0.302]
1	A/M	0.055 [0.002,0.109]	0.244 [0.117,0.370]
1	$A/M_\times$	−0.209 [-0.317,-0.101]	−0.304 [-0.512,-0.096]
1	M/P	−0.419 [-0.541,-0.296]	−0.814 [-1.105,-0.522]
1	$M_\times/P$	−0.154 [-0.238,-0.071]	−0.266 [-0.534,0.002]
0.5	A/P	−0.424 [-0.555,-0.293]	−0.480 [-0.694,-0.265]
0.5	A/M	0.029 [-0.037,0.096]	0.138 [-0.058,0.334]
0.5	$A/M_\times$	−0.127 [-0.252,-0.002]	−0.409 [-0.676,-0.142]
0.5	M/P	−0.453 [-0.596,-0.311]	−0.618 [-0.892,-0.343]
0.5	$M_\times/P$	−0.297 [-0.431,-0.163]	−0.070 [-0.263,0.122]

Table 7: Residual-energy half of the same 60-test family (remaining 30 comparisons). Mean log ratios [simultaneous 95% intervals]; negative favors the numerator. C, Q are energies on aligned evaluation pairs; their ratios are not RMSE ratios. The correction covers both secondary tables together.

η	Contrast	Q_4	C_{16}
0	A/P	−1.575 [-1.728,-1.422]	−0.070 [-0.109,-0.031]
0	A/M	−0.849 [-0.945,-0.753]	0.632 [0.536,0.728]
0	$A/M_\times$	−0.944 [-1.041,-0.848]	0.624 [0.542,0.706]
0	M/P	−0.726 [-0.838,-0.614]	−0.702 [-0.789,-0.616]
0	$M_\times/P$	−0.631 [-0.725,-0.536]	−0.694 [-0.764,-0.624]
1	A/P	−2.978 [-3.292,-2.664]	−0.742 [-0.980,-0.504]
1	A/M	−0.155 [-0.223,-0.087]	0.163 [0.050,0.275]
1	$A/M_\times$	−2.591 [-2.921,-2.261]	−0.749 [-1.107,-0.390]
1	M/P	−2.823 [-3.135,-2.511]	−0.905 [-1.154,-0.655]
1	$M_\times/P$	−0.387 [-0.668,-0.105]	0.007 [-0.273,0.288]
0.5	A/P	−3.411 [-3.865,-2.956]	−0.387 [-0.560,-0.213]
0.5	A/M	−0.167 [-0.343,0.008]	0.153 [0.011,0.295]
0.5	$A/M_\times$	−2.651 [-3.015,-2.287]	−0.191 [-0.420,0.037]
0.5	M/P	−3.243 [-3.686,-2.800]	−0.540 [-0.747,-0.332]
0.5	$M_\times/P$	−0.760 [-1.136,-0.384]	−0.195 [-0.386,-0.005]

A3 Residual covariance under action-preserving re-pairing

Fix the model parameters, one shared-action library, one actuator mode, one candidate-action pair and one prediction time. Write $x_k, y_k \in \mathbb{R}^{d_e}$ for the two states' effect residuals in family $k \in$

$\{1, \ldots, n\}$. The new controlled study has six or seven families per such mode/library stratum. A permutation π only reassigns the second-state family. Define

$$D = \frac{1}{nd_e} \sum_k \|y_k - x_k\|^2, \qquad D_\pi = \frac{1}{nd_e} \sum_k \|y_{\pi(k)} - x_k\|^2.$$

Every permutation preserves the marginal squared terms. Consequently the exact finite-permutation difference is

$$D_\pi - D = \frac{2}{nd_e} \sum_k \langle x_k, y_k - y_{\pi(k)} \rangle.$$

This is an identity at the same parameters and branch predictions, not a comparison between separately trained models.

For a uniformly random single-cycle permutation, each $\pi(k)$ is uniformly distributed among the $n - 1$ indices other than k. Therefore

$$\mathbb{E}_\pi \frac{1}{n} \sum_k \langle x_k, y_{\pi(k)} \rangle = \frac{n\langle \bar{x}, \bar{y} \rangle - \frac{1}{n} \sum_k \langle x_k, y_k \rangle}{n - 1}.$$

Substituting into the finite difference proves Equation 6. The covariance uses the population divisor n; a sample-covariance divisor $n - 1$ would change the displayed factor. The fixed mode and action library are essential: this statement concerns residual covariance conditional on those variables. Averaging the per-stratum identities with their fixed population weights gives the full-batch expectation. The factor $n/(n - 1)$ stays inside that average because stratum sizes can differ. The same derivation commutes with fixed nonnegative time, coordinate and target-energy weights, as used by the training loss.

With $J_{x,k} = \partial x_k / \partial \theta$ and $J_{y,k} = \partial y_k / \partial \theta$, the corresponding gradient is

$$\nabla_\theta \Gamma = \frac{1}{nd_e} \sum_k \left[(J_{x,k} - \bar{J}_x)^\top (y_k - \bar{y}) + (J_{y,k} - \bar{J}_y)^\top (x_k - \bar{x}) \right].$$

Thus even fixed marginal losses can have different training gradients after re-pairing. Neither the value nor the sign of Γ alone determines its gradient or the effect on held-out error. For a common-parameter SGD comparison, an added training term A_B changes an evaluation quantity M_E relative to the same baseline update by

$$-\eta \langle \nabla M_E, \nabla A_B \rangle + O(\eta^2).$$

Training and evaluation populations and weights may differ, so a training decrease does not imply a held-out decrease. Gradient clipping and AdamW require the actual update displacement in place of $-\eta \nabla A_B$ (Appendix A13). No new gradient probe or long-training mediation result is claimed here.

The implemented study samples one cycle per mode/library stratum before training and keeps it fixed. Equation 6 describes that randomization law, not exact equality for each sampled cycle. The experiment, including its gradient-matched mismatch arm, tests the eventual effect of that finite intervention. It does not infer a universal application threshold from residual covariance or declare covariance a unique causal explanation of any observed gain.

Common variation and differential variation. The covariance has a more explicit interpretation. In a fixed stratum, write $x = c - q$, $y = c + q$, and use one symmetric positive semidefinite coordinate weight W. Define

$$B_c = \|\bar{c}\|_W^2, \qquad V_c = \frac{1}{n} \sum_k \|c_k - \bar{c}\|_W^2,$$

with analogous B_q, V_q. Taking $W = I/d_e$ recovers the convention above; coordinate-summed diagnostics instead use $W = I$ on the stated standardized coordinates. Symmetry cancels the scalar cross terms, giving $\Gamma = V_c - V_q$. This is a weighted-trace identity: the full cross-covariance matrix also contains $\mathrm{Cov}(c, q) - \mathrm{Cov}(q, c)$, which need not vanish. Therefore

$$D = 4(B_q + V_q), \qquad \mathbb{E}_\pi D_\pi = 4B_q + \frac{2n}{n-1} V_c + \frac{2(n-2)}{n-1} V_q. \tag{8}$$

At equal raw coefficients, re-pairing adds a penalty on common residual variation and reduces the coefficient on differential variation. Neither raw term directly measures the common stratum mean B_c. For $n = 6$, the two variance coefficients are $12/5$ and $8/5$; for $n = 7$, they are $7/3$ and $5/3$. Strata are weighted by their family populations, not equally. This is a reparameterization of the covariance identity, not an additional learning guarantee.

The actual fixed cycle. For $(\Pi v)_k = v_{\pi(k)}$, let the stacked norm include the family average and the same W. Then

$$D_\pi = \underbrace{\|(\Pi - I)c\|^2}_{H_c} + \underbrace{\|(\Pi + I)q\|^2}_{H_q} + \underbrace{2\langle (\Pi - I)c, (\Pi + I)q\rangle}_{H_{cq}}. \tag{9}$$

The first term measures differences of common residuals along the cycle; the second measures sums of adjacent differential residuals. The cross term is signed and equals $2\langle c, (\Pi^\top - \Pi)q\rangle$. Defining the correspondingly weighted $\Gamma_\pi = n^{-1}\sum_k \langle x_k - \bar{x}, y_{\pi(k)} - \bar{y}\rangle_W$, the finite deviation from the random-cycle expectation is

$$D_\pi - \mathbb{E}_\pi D_\pi = -2\left(\Gamma_\pi + \frac{\Gamma}{n-1}\right).$$

It need not be small. A fixed cycle's H_c is not proportional to V_c: a change from high- to low-frequency cycle modes can lower H_c while increasing V_c. Thus a reduction in common variance is one specific interpretation to examine, not a necessary implication of the fixed-cycle identity.

When $q = 0$, the mismatch term vanishes exactly when $W^{1/2}c_k$ is constant along the cycle. If W is positive definite on the measured space, this means constant c_k there. For an even cycle, $\Pi q = -q$ also gives a pure differential zero direction when c is constant, although the aligned energy is positive whenever q has positive weighted energy. An odd cycle has no such measured pure differential direction. These facts do not remove the joint c, q cancellation or the null space of a semidefinite W.

Why this is not the matched arm's effective coefficient. The actual mismatch arm uses the detached multiplier

$$\gamma_B = \mathrm{stopgrad}\left(\lambda \frac{\|\nabla_\theta D_B\|_2}{\|\nabla_\theta D_{\pi,B}\|_2}\right)$$

for the complete training batch, all six coordinates, and all four supervised horizons with their original energy weights. If the aligned auxiliary gradient is zero, the implementation sets $\gamma_B = 0$; a positive target norm with zero mismatch norm is an error. In particular, zero differential residual throughout that weighted batch disables both matched auxiliary updates, so matched mismatch does not eliminate the complete common-error blind direction. Zero q in one held-out stratum is insufficient for this condition, and nonzero q need not give a nonzero parameter gradient. Likewise, a raw mismatch zero with a nonzero aligned gradient is not a normally trainable free direction under the matching rule.

The raw expectation fixes the residual field. It is not an expectation over parameters learned using the same sampled training cycle; the parameter-dependent γ_B also cannot be pulled outside that expectation. Held-out operator decompositions can describe the saved residuals and their specified weights. They do not recover the training multiplier, the original minibatch gradients, or a unique mediator of the observed training effect.

A4 A FIXED-MIDPOINT INTERVENTION ON THE TRAINING STATE

This follow-up was designed after the completed eighteen-cohort correspondence study showed lower four-step differential error but higher sixteen-step common error for aligned supervision than for gradient-matched state mismatch. It uses the same cohorts and evaluation populations. Its interventions were fixed before their execution, but it is not a fresh-cohort confirmation and its intervals do not control selection across the preceding research process. The checkpoint is the midpoint of the original training clock, not a checkpoint selected from intermediate performance.

Eight continuations per cohort. Let A denote the original aligned objective and M the original action-preserving, gradient-matched mismatch. At update 1,000, we independently choose the parameter source $p \in \{\mathrm{A}, \mathrm{M}\}$, Adam-history source $o \in \{\mathrm{A}, \mathrm{M}\}$, and continuation objective $l \in \{\mathrm{A}, \mathrm{M}\}$. Their Cartesian product gives eight cells per cohort; a three-letter cell label always follows the order (p, o, l). Each cell performs the original remaining 1,000 updates. The 144 continuations therefore execute 144,000 additional updates, including the 36 unchanged AAA/MMM paths. The latter are actual continuation controls, not reused final predictions.

Parameters and Adam's first/second moments and step counters are loaded separately in the original parameter order. Learning rate, weight decay, gradient clipping, precision, batch size, and the global update clock remain unchanged. The same saved NumPy sampler continues the original library-block sequence. The sources have identical CUDA random state, input/output scales, evaluation reference, and fixed initial coefficient. Their CPU random states differ across the original processes; each continuation restores its parameter source's CPU state. The predictor computations use no CPU random draws. For M, the detached gradient-norm multiplier is recomputed at that continuation's own parameters and current batch, using the original zero-handling rules. Neither objective restarts the learning-rate schedule or recalibrates its coefficient.

Source data and checkpoints are read-only. All cells retain complete checkpoints at steps 1,500 and 2,000. Evaluation covers the two original parameter anchors at step 1,000 and every continuation at steps 1,500 and 2,000: 36+144+144=324 complete-population evaluations. Each uses all 1,024 original families and all sixteen prediction steps. The intermediate positions provide descriptive trajectories; all formal comparisons below use step 2,000. Training logs retain every prescribed fifty-update observation. New native-environment calls and newly generated scientific labels are both zero.

What the interventions identify. Changing only l compares two future objectives from the same complete training state. Changing p or o tests the corresponding specified training-history intervention, averaged over the other assignments when stated. An optimizer state learned with one parameter vector can be incompatible with the other vector. Shared initialization and parameter coordinates make the swap well-defined but do not remove this possible shock. These contrasts are not natural-mediator fractions, do not identify a unique gate mechanism, and do not explain the entire aligned-versus-paired gain or the controlled-versus-native gap. An unresolved contrast does not establish absence or equivalence.

Fixed primary family. For a positive error metric Z, write $z_{pol,r} = \log Z_{pol,r}$ for cohort r. The primary endpoint is the original motion-coordinate common residual energy C at prediction step sixteen. Four comparisons change only the future objective:

$$d_{po,r} = z_{po\mathrm{A},r} - z_{po\mathrm{M},r}, \qquad (p, o) \in \{\mathrm{A}, \mathrm{M}\}^2.$$

Three additional comparisons average each factor's A-minus-M change over the other four assignments:

$$T_{p,r} = \tfrac{1}{4} \sum_{o,l} (z_{\mathrm{A}ol,r} - z_{\mathrm{M}ol,r}), \quad T_{o,r} = \tfrac{1}{4} \sum_{p,l} (z_{p\mathrm{A}l,r} - z_{p\mathrm{M}l,r}), \quad T_{l,r} = \tfrac{1}{4} \sum_{p,o} d_{po,r}.$$

The seven comparisons are deliberately not independent: T_l averages the four simple objective effects. We retain this dependence and use a single seven-comparison Bonferroni family. Each contrast is first calculated within each of the eighteen cohorts, then summarized with its mean and corrected t_{17} interval. Positive values mean larger error for the A level in the stated intervention. Single-factor log contrasts can be exponentiated to geometric ratios, not ratios of arithmetic means.

Fixed secondary family. All four interactions are retained: parameter–optimizer, parameter–objective, optimizer–objective, and the three-factor interaction. A two-factor interaction is the difference of its simple effects, averaged over the remaining factor; for example,

$$I_{po,r} = \tfrac{1}{2} \sum_{l} (z_{\mathrm{A}\mathrm{A}l,r} - z_{\mathrm{A}\mathrm{M}l,r} - z_{\mathrm{M}\mathrm{A}l,r} + z_{\mathrm{M}\mathrm{M}l,r}).$$

For $s_{\mathrm{A}} = 1, s_{\mathrm{M}} = -1$, the three-factor interaction is

$$I_{pol,r} = \sum_{p,o,l} s_p s_o s_l z_{pol,r}.$$

The secondary family comprises these four interactions for C_{16} and all eleven contrasts (four simple effects, three averages, four interactions) for each of Q_4, F_4, and E_{16}: $4+3\times 11 = 37$ comparisons. They share a separate thirty-seven-comparison Bonferroni correction. Interactions are reported on the log scale; exponentiating them produces nested ratios, not percentage changes in an individual method's error. Any undefined or nonpositive metric leaves the affected contrast undefined, with all cohort values retained and no pseudocount or deletion.

The complete eight-cell tables, all eighteen within-cohort values, and all horizons accompany the fixed comparisons. Family/action pairs, checkpoints and cells do not count as additional independent training repetitions. The scientific interpretation is restricted to this midpoint, remaining budget, objective pair and prescribed system.

Table 8: All 44 midpoint comparisons on their original log scale: estimate above corrected interval. The first seven C_{16} cells form the seven-test primary family; all other 37 cells share the secondary correction. d_{po} changes only the future objective at fixed parameter/Adam sources. Direct percentages in the main text use $100(e^d - 1)$; interactions are not method-error percentages.

Contrast	C_{16}	Q_4	F_4	E_{16}
d_{AA}	+0.0007 [-0.0539, +0.0553]	-0.1277 [-0.1655, -0.0898]	+0.1429 [+0.0880, +0.1977]	-0.0141 [-0.0483, +0.0201]
d_{AM}	-0.0077 [-0.0632, +0.0479]	-0.1311 [-0.1702, -0.0920]	+0.1441 [+0.0891, +0.1992]	-0.0182 [-0.0529, +0.0166]
d_{MA}	-0.0028 [-0.0494, +0.0439]	-0.1434 [-0.1712, -0.1156]	+0.0753 [+0.0281, +0.1224]	-0.0132 [-0.0406, +0.0142]
d_{MM}	-0.0071 [-0.0540, +0.0398]	-0.1423 [-0.1695, -0.1150]	+0.0740 [+0.0296, +0.1185]	-0.0152 [-0.0427, +0.0123]
T_p	+0.1753 [+0.0909, +0.2598]	-0.0446 [-0.1071, +0.0178]	+0.0974 [-0.0497, +0.2444]	+0.0793 [+0.0267, +0.1319]
T_o	-0.0129 [-0.0175, -0.0083]	-0.0108 [-0.0174, -0.0042]	-0.0043 [-0.0105, +0.0019]	-0.0065 [-0.0093, -0.0036]
T_l	-0.0042 [-0.0538, +0.0454]	-0.1361 [-0.1657, -0.1065]	+0.1091 [+0.0632, +0.1549]	-0.0152 [-0.0453, +0.0150]
I_{po}	+0.0027 [-0.0079, +0.0132]	+0.0033 [-0.0020, +0.0085]	+0.0082 [-0.0006, +0.0170]	+0.0011 [-0.0028, +0.0050]
I_{pl}	+0.0014 [-0.0291, +0.0319]	+0.0134 [-0.0175, +0.0443]	+0.0688 [+0.0263, +0.1114]	-0.0019 [-0.0175, +0.0137]
I_{ol}	+0.0063 [+0.0022, +0.0104]	+0.0011 [-0.0016, +0.0038]	-4.59×10^{-6} [-0.0029, +0.0029]	+0.0031 [+0.0011, +0.0050]
I_{pol}	+0.0040 [+0.0004, +0.0076]	+0.0046 [+0.0006, +0.0085]	-0.0025 [-0.0073, +0.0022]	+0.0021 [+0.0001, +0.0041]

All formal intervals are retained, including every unresolved simple C_{16} and E_{16} objective effect. Future alignment lowers Q_4 and raises F_4 at each prescribed starting assignment. Resolved interactions prevent an additive mechanism interpretation. The complete eight-cell arithmetic means and all intermediate horizons remain in the saved-array outputs; they are not substitutes for these paired log estimands.

A4.1 What the saved residuals explain

All 324 unique prediction sources were analyzed descriptively without new fits, gradients, environment calls or tests. Each of 32 evaluation libraries has five mode strata: 96 strata contain six families and 64 contain seven, weighted by $n/1024$. All 28 candidate pairs are retained. Motion components sum four standardized coordinates; the separate weighted operator uses the original six-coordinate training weights and supervised steps 1–4. A saved evaluation permutation, never used for training, is applied to each fixed residual field.

The arithmetic parameter-source difference in C_{16} is $.00308573 = .00126506 + .00182068$: common stratum mean squared plus within-stratum variance. Both contribute; neither raw auxiliary operator directly penalizes the first. This does not decompose the 19.16% geometric contrast, estimate causal shares or exclude indirect effects through shared parameters. Continuing with M gives a lower mean H_c and higher H_q at supervised horizons from all four starting states. Signed H_{cq},

Table 9: All final cells: motion common residual at prediction step sixteen. Arithmetic means across eighteen cohorts, scaled by 10^3. These descriptive sums do not decompose the primary geometric percentage contrast.

Cell	$10^3 B_c$	$10^3 V_c$	$10^3 C$
AAA	13.25668	5.31232	18.56899
AAM	13.49794	5.19585	18.69379
AMA	13.28069	5.42074	18.70143
AMM	13.58383	5.39376	18.97759
MAA	11.93104	3.54559	15.47663
MAM	12.21792	3.37867	15.59659
MMA	12.04823	3.62241	15.67064
MMM	12.36172	3.49330	15.85502

finite-cycle deviations and all eight cells are retained in the complete operator tables. These held-out quantities cannot recover the parameter-dependent multiplier or original training gradients.

The package contains all 6,912 motion rows, 1,728 full-coordinate horizon rows, 432 weighted rows and 324 stratum arrays, spanning updates 1,000/1,500/2,000. Its CPU entry reaggregates those sufficient statistics, not their original prediction residuals. At update 1,000 eight labels map to only two parameter anchors per cohort; these shared values are not additional independent evaluations.

A5 Prospective forecasts of finite auxiliary-loss interventions

Population and data separation. The fixed population is DM-Control Cartpole-swingup, Reacher-hard, and Meta-World Push-v3, with 12 independently seeded, complete three-system suites. Each system/cohort has independent fitting (F), forecast-validation (V), and evaluation (E) streams: 32/16/32 episodes of 500 control calls, yielding 1024/256/512 families with horizons 4/4/16. A family contains two initial states and two shared candidate action sequences. Before branching, 64 F or 32 V/E initial-state indices per episode are sampled from $0, 5, \ldots, 480$, without outcome filtering. Their complete simulator, mutable-model, controller, and random-number snapshots and all behavior transitions are retained. E collection can precede forecasting, but model evaluation, analysis, and preview access to E require all 144 fixed terminal models and all 36 forecast records, including explicit failures.

Shared fitting specification. A recursive residual predictor uses two width-128 hidden layers and padded 64-state/4-action inputs. Mechanical/controller dimensions are 4/4/38, with 0/6/3 copied context coordinates; padding is masked. Restored solver caches are not inputs; strict Markov sufficiency is not asserted. F alone sets state-coordinate means/scales, counting each initial state and each future branch state once; standard deviations at most 10^{-3} use unit scale. Actions retain native coordinates. Arms share initialization, complete 32-family action-library batches, and 2000-update AdamW schedules (10^{-3} learning rate, $(0.9, 0.999)$ betas, 10^{-8} epsilon, 10^{-3} weight decay, 20-step warmup, gradient clipping at 1). One control call lasts 0.01/0.02/0.0125 seconds, respectively: shared dimensions and horizons do not equate information or physical duration.

Objectives and matched controls. Let $\mathcal{L}_{\mathrm{pred}}$ be standardized absolute-state mean squared error and let G_h, C_h, Q_h, D_h denote the response-residual energies at control step h, averaged over effective dynamic coordinates and families. For $e_1 = c - q, e_2 = c + q$, these satisfy $G_h = C_h + Q_h$ and $D_h = 4Q_h$. Training averages steps 1–4. Its response terms, denoted with tildes, share the F-only true paired-response-energy denominator at each step, floored at 10^{-4}. The paired baseline is $B = \mathcal{L}_{\mathrm{pred}} + \widetilde{G}$ (the protocol denotes $\mathcal{L}_{\mathrm{pred}}$ by P and writes $B = P + G$); $\mathcal{L}_{\mathrm{pred}}$ alone is not the paired baseline. The three auxiliary gradient directions are

$$\begin{aligned} &a_A = \lambda \nabla \widetilde{D}, \qquad a_M = \gamma_M \nabla \widetilde{D}_\pi, \qquad a_S = \gamma_S \nabla \widetilde{G}, \\ &\gamma_j = \frac{\lambda \|\nabla \widetilde{D}\|}{\|\nabla \widetilde{L}_j\|}, \quad (\widetilde{L}_M, \widetilde{L}_S) = (\widetilde{D}_\pi, \widetilde{G}). \end{aligned} \tag{10}$$

Here π is fixed within the same F episode and action library; each prediction moves with its own true response, preserving candidate identity and labels. The shared λ is the median of $0.25\|\nabla B\|/\|\nabla \widetilde{D}\|$ over 16 fixed F blocks at initialization. Matching is recomputed at each arm's own state before clipping, with detached γ_j in the parameter gradient. A zero target norm sets $\gamma_j = 0$; a positive target with zero denominator is a retained failure. This matches auxiliary gradient norms, not clipped updates or Adam trajectories.

Full-path local sensitivity. Following differentiation through optimization (Maclaurin et al., 2015; Franceschi et al., 2017), introduce the dose ρ through the explicit gradient field $\nabla B + \rho a_j$. With $z_t = (\theta_t, m_t, v_t)$ denoting parameters and both Adam moments, let $U_{j,t}$ include the fixed batch, schedule, clipping, and weight decay. Along the complete $\rho = 0$ baseline trajectory, forward Jacobian–vector products propagate three sensitivities:

$$\begin{aligned} &s_{j,0} = 0, \qquad s_{j,t+1} = D_z U_{j,t}(z_{B,t}, 0) s_{j,t} + \partial_\rho U_{j,t}(z_{B,t}, 0), \\ &p_j^X = \nabla_\theta \log X_V(\theta_{B,2000})^\top s^\theta_{j,2000}. \end{aligned} \tag{11}$$

Dependence of γ_j on parameters vanishes from this state Jacobian only because it is multiplied by $\rho = 0$; detaching the parameter gradient would not justify that omission at nonzero dose. For $X \in \{R_G, C, Q, R_P\}$, forecasts of the finite $\rho = 1$ A/M and A/B log ratios are $p_A^X - p_M^X$ and p_A^X. This extrapolation can fail; no coefficient or horizon is fitted to E, and no finite-dose accuracy or convergence guarantee is implied.

Endpoints and forecast comparisons. The primary endpoint is exactly control step 4, with $R_G = \sqrt{G_4}$ and $R_P = \sqrt{\mathcal{L}_{\text{pred}}(4)}$; reported energies use coordinate means without the training response-energy denominator. The 1–16 curves are pointwise, not cumulative, and actions change after step 5, so their longer horizon is not a pure time extension. Define actual E log ratios $y^X_{AM} = \log(X_A/X_M)$ and $y^X_{AB} = \log(X_A/X_B)$ and their forecasts p^X_{AM}, p^X_{AB} above. The fixed headroom forecast is $H = \frac{1}{2}\log[C_{B,V}/(C_{B,V} + Q_{B,V})]$, which assumes unchanged common error and removal of differential error. The initialization baseline uses gradient cosine similarity (Du et al., 2018): its direction is $k = -(\cos_A - \cos_M)$, comparing the V gradient of $\log R_G$ with each mean auxiliary vector over the same 16 F blocks. Zero vectors give zero direction. With $\ell(p, y) = (p - y)^2$ and direction error $d(p, y)$ equal to one for opposite signs, zero for matching nonzero signs, and $1/2$ if either argument is zero, the complete five-comparison family is

$$\begin{aligned} &\ell(p^{R_G}_{AM}, y^{R_G}_{AM}) - \ell(0, y^{R_G}_{AM}), \qquad \ell(p^{R_G}_{AB}, y^{R_G}_{AB}) - \ell(0, y^{R_G}_{AB}), \\ &\ell(p^{R_G}_{AB}, y^{R_G}_{AB}) - \ell(H, y^{R_G}_{AB}), \qquad \frac{1}{2}\sum_{X\in\{C,Q\}} [\ell(p^X_{AM}, y^X_{AM}) - \ell(0, y^X_{AM})], \\ &d(p^{R_G}_{AM}, y^{R_G}_{AM}) - d(k, y^{R_G}_{AM}). \end{aligned} \tag{12}$$

Each contrast first averages the three systems with fixed weights $1/3$. The 12 suite cohorts are the independent units. Two-sided 95% Bonferroni simultaneous intervals use $\bar{d} \pm t_{11,0.995} s_d/\sqrt{12}$ for all five contrasts; an upper endpoint below zero supports lower forecast loss, not necessarily a beneficial training intervention. No systems or failed cohorts are dropped. A nonfinite sensitivity with finite baseline parameters/moments is recorded while baseline training continues unchanged; any required forecast failure makes the complete five-test conclusion unavailable.

Numerical scope and stopping. Parameters, Adam states, gradients, and sensitivities use FP32 without AMP/TF32; normalization, native labels, and statistical reductions use FP64. Separate engineering checks used synthetic inputs with the actual 33,600-parameter model and covered 32/64 updates, recording three-direction FP32–FP64 derivative discrepancies at most 4.264×10^{-8}; they do not certify 2000-update sensitivity accuracy. The fixed 144-fit/36-sensitivity-path design has a 32 GPU-process-hour cap including engineering, failures, and evaluation. Budget exhaustion remains incomplete evidence; outcomes do not trigger replacement cohorts or extra comparisons.

A6 Complete native forecast results

The fixed matrix completed all 144 fits, 288,000 updates, 36 forecasts and 144 final-weight inference replays. All five main forecast comparisons are reported in Table 10; none establishes superi-

Table 10: All step-4 forecast comparisons. Negative favors the path forecast. Magnitude loss is squared log-ratio error; the last row uses direction error. Simultaneous 95% Bonferroni t_{11} intervals use twelve complete cohorts with equal system weights.

Forecast comparison	Difference	Simultaneous interval
A/M R_G, versus zero	1.050	$[-0.788, 2.888]$
A/B R_G, versus zero	6.590	$[-1.029, 14.209]$
A/B R_G, versus Q headroom	6.593	$[-1.019, 14.205]$
A/M mean C, Q, versus zero	4.405	$[-2.743, 11.552]$
A/M direction, versus initial cosine	-0.167	$[-0.635, 0.302]$

ority. Every required forecast and terminal metric was finite. Saved prediction/truth reconstruction retained all 36 cells and all sixteen horizons. Independent CPU calculations also reproduced all 18,432 metric values and the five comparisons; this is numerical verification, not an independent replication of training.

A6.1 Prospective forecasts do not establish an advantage

None of the five prespecified tests establishes superiority (Table 10). All twelve complete cohorts have greater magnitude-forecast error in each of four comparisons, although the corrected intervals do not establish inferiority. The tested prediction extrapolates a zero-dose derivative to unit dose and from V to E.

Descriptive outcomes retain system differences: in Cartpole/Reacher/Push order, A/B changes response RMSE by $-5.49\%/-39.70\%/+1.82\%$, but absolute RMSE by $+30.82\%/+13.04\%/+13.25\%$. A/M response changes are $+0.69\%/-1.00\%/-0.59\%$, without a formal efficacy family. A/M direction errors are 0/9/7 out of twelve per system, versus 8/8/6 for initial cosine. Accurate Cartpole signs coexist with poor magnitudes. Appendix A6 retains all arms and horizons.

Where does the forecast fail? A fixed post-result 396-fit diagnostic reuses all 36 cells and 2,000 updates (Appendix A7). FP64 A/B and A/M response-forecast mean squared errors are 6.681 and 1.046 on V, versus 6.689 and 1.051 on E. Squared V/E finite-effect gaps are 0.0200 and 0.000755; precision-change energies are 6.0×10^{-9} and 9.5×10^{-10}. Large unit-dose error thus already exists on V in double precision. These squared differences describe error magnitudes rather than causal fractions; finite-effect gaps also leave derivative gaps unconstrained. Post-hoc E derivatives worsen A/B forecast error but improve A/M; they do not uniformly repair prediction. Signed full-path probes at $10^{-3}, 10^{-4}$ retain exceptions: at 10^{-4}, median scaled discrepancy is 4.95×10^{-5} but maximum 0.801. The extrapolation fails, while the remaining numerical exceptions leave some local derivatives unresolved.

Full-population outcomes. Every one of the twelve complete-suite cohorts has a positive difference in each of the four magnitude comparisons: greater observed error than its comparator. All five corrected intervals cross zero. These are not 48 independent repetitions, a new inferiority test or evidence of equivalence. Heavy tails, twelve cohorts and the fixed system population limit the t inference. The full cell table retains exceptions: Push-r11's A/M response and mean-C/Q differences are $-.000518$ and $-.001612$; all twelve Push A/B-headroom cells favor the path forecast. None replaces the fixed equally weighted three-system test.

For A/M response direction the forecast makes 0/9/7 errors in Cartpole/Reacher/Push, against 8/8/6 for initialization cosine, each out of twelve. Cartpole's correct signs coexist with inaccurate magnitudes: cohort 4 forecasts a log change 2.71517 and observes .05432. A/B sign errors are 2/6/9; Reacher's A/B response improves in all twelve cohorts, while six predictions have the wrong sign. These are descriptive counts, not extra tests.

Interpretation of failed extrapolation. The local derivative includes parameter and Adam sensitivities along 2,000 clipped updates. Short 32/64-update FP64/finite-difference engineering checks do not establish full-path numerical accuracy. Consequently, the failed forecast cannot be assigned uniquely to finite-dose curvature, V-to-E sampling transfer or accumulated numerical error. No dose,

Table 11: Standardized response and absolute-state RMSE changes (%). Negative denotes smaller error; brackets are descriptive, unadjusted 95% intervals.

System	Contrast	R_G	R_P
Cartpole	A/B	-5.49 [-14.65, +4.65]	+30.82 [+2.97, +66.21]
Cartpole	A/M	+0.69 [-0.44, +1.83]	-0.95 [-12.00, +11.47]
Cartpole	A/S	+5.40 [-2.61, +14.06]	+0.43 [-19.26, +24.93]
Reacher	A/B	-39.70 [-44.23, -34.80]	+13.04 [+3.74, +23.17]
Reacher	A/M	-1.00 [-3.27, +1.32]	+0.15 [-4.62, +5.17]
Reacher	A/S	-38.10 [-42.47, -33.39]	+0.01 [-7.34, +7.94]
Push	A/B	+1.82 [+0.52, +3.14]	+13.25 [+11.05, +15.49]
Push	A/M	-0.59 [-1.14, -0.04]	-1.91 [-2.92, -0.88]
Push	A/S	+3.03 [+1.95, +4.12]	+2.10 [+0.72, +3.49]

Table 12: Common and differential residual-energy changes (%). Negative denotes smaller error; brackets are descriptive, unadjusted 95% intervals.

System	Contrast	C	Q
Cartpole	A/B	-9.63 [-28.81, +14.72]	-11.89 [-25.24, +3.84]
Cartpole	A/M	+2.20 [-0.47, +4.93]	+0.38 [-2.00, +2.83]
Cartpole	A/S	+14.29 [-5.09, +37.62]	+7.52 [-5.24, +22.00]
Reacher	A/B	-58.95 [-66.44, -49.78]	-74.65 [-78.54, -70.06]
Reacher	A/M	-2.34 [-7.87, +3.51]	-0.97 [-3.01, +1.12]
Reacher	A/S	-58.17 [-64.89, -50.18]	-71.06 [-76.37, -64.56]
Push	A/B	+5.55 [+2.22, +8.99]	+1.21 [-0.85, +3.32]
Push	A/M	-1.50 [-2.83, -0.15]	-0.77 [-1.85, +0.33]
Push	A/S	+8.03 [+5.36, +10.78]	+3.73 [+1.92, +5.57]

scale, horizon, intercept or threshold was fitted on E afterward. Appendix A7 now separates finite-effect population and precision discrepancies; full-path numerical exceptions remain. The standardized and native-coordinate measurements also differ: mixed mechanical/controller units cannot be called meters. Complete native-unit contrasts and every cohort/horizon remain supplied.

This study differs from the old ten-task bridge in architecture, model-visible coordinates, normalization, shared response-energy denominator, action-preserving mismatch and collection distribution. Its large Reacher A/B response change and the old .133% suite gain are different estimands, not an identified physical explanation of effect-size differences. Cartpole adds a structurally different fixed system; Reacher and Push add new data within previously studied task categories. No forecast-superiority test is a registered test of training-arm efficacy or control utility.

All 108 F/V/E arrays, 36 forecast/prediction cells, 144 final models and 36 initial models are supplied. Author-side execution of the standalone public GPU replay reproduced every final prediction exactly, with no new training or simulator calls. This is inference verification, not independent retraining. The fixed preview sample uses family 0/state 0/candidate 0 for every system/cohort/arm; 36 videos were fully decoded and fixed frames 1/8/16 inspected. Five Cartpole videos have the pole outside the fixed camera boundary in some inspected frames. Preview encoder/framebuffer repairs changed no trajectories or models; original failed attempts remain in the complete local archive.

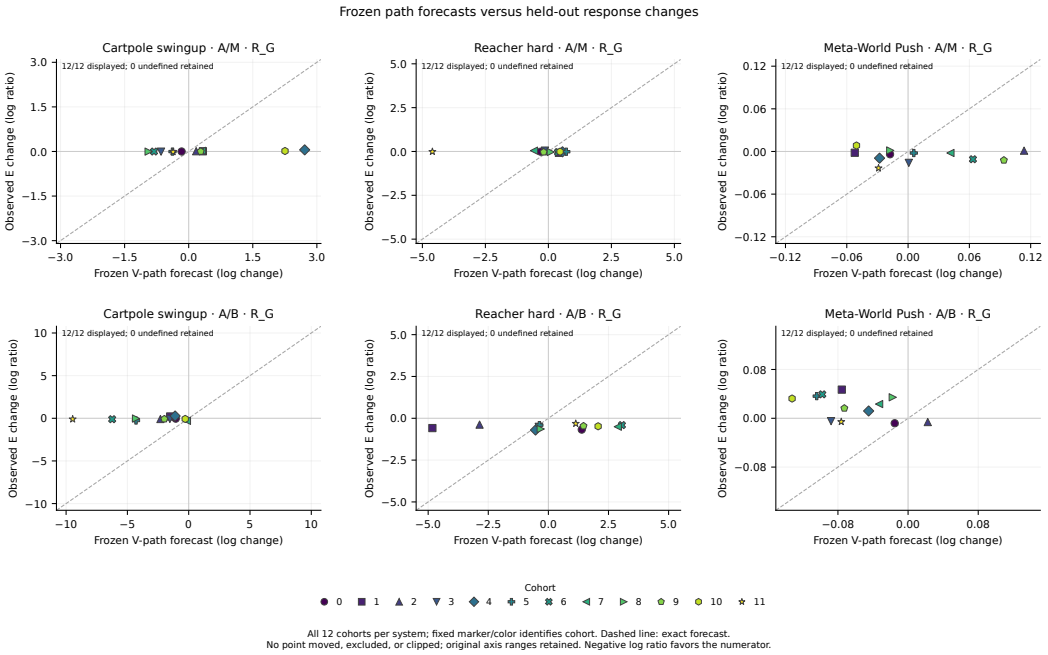

Figure 5: All 36 system/cohort response forecasts for both A/M and A/B, with equal x/y scales and identity lines. Extreme predictions and favorable cells remain. No scale was fitted on E. Complete separate C/Q plots and values accompany the data.

A7 Separating finite-dose, population and precision discrepancies

Designed after the failed forecasts, this diagnostic reuses all twelve three-system cohorts without new data or environment calls. Its 36 FP64 baseline/sensitivity and 360 signed/unit-dose fits complete 792,000 updates, all finite. Another 36 jobs re-evaluate original FP32 models. The original five prospective tests remain unchanged.

Fixed paths and precision intervention. Promote the original FP32-rounded inputs, normalization, initialization and coefficient to FP64; preserve all 2000 batches, scheduling, clipping and AdamW settings in Appendix A5. Propagate A/M full-state sensitivities at zero dose; independently fit both arms at $-10^{-3}, -10^{-4}, 10^{-4}, 10^{-3}, 1$. Matching is recomputed at each model's own state with the original stop-gradient convention. Signed doses are numerical probes. Every V/E family is evaluated at step 4; no cell, step or endpoint is replaced. Precision can change the optimizer path itself.

Exact signed decomposition. For A/B or A/M, let p_0 be the original forecast, $p_{32,V}, p_{64,V}$ the corresponding V projections, and $y_{d,S}$ the finite unit-dose log contrast in precision d on population S. The original reported E contrast is y_0. Then

$$\begin{aligned} p_0 - y_0 = (p_0 - p_{32,V}) + (p_{32,V} - p_{64,V}) + (p_{64,V} - y_{64,V}) \\ + (y_{64,V} - y_{64,E}) + (y_{64,E} - y_{32,E}) + (y_{32,E} - y_0). \end{aligned} \tag{13}$$

These are replay, derivative precision, implemented unit-dose, V/E, endpoint precision and reporting discrepancies. Squaring retains six squares and all fifteen doubled cross terms, not causal shares. Cohorts keep equal system weights; pointwise t_{11} summaries are descriptive, not new efficacy tests.

Across all 36 cells with equal system weights, large mean squared unit-dose discrepancy exists on V in both precisions. For response RMSE, A/B V squared error is 6.68165 in FP32 and 6.68148 in FP64; A/M is 1.04629 and 1.04634. Pointwise FP64 intervals are [1.22762,12.1353] and [−0.255083,2.34777]. Across all contrast/metric rows, derivative precision-change energy is at

Table 13: All contrast/metric means over twelve complete cohorts. "Dose on V" is $(p_{64,V} - y_{64,V})^2$; "V/E gap" is $(y_{64,V} - y_{64,E})^2$; "Precision" sums both precision squares; "Cross sum" includes all fifteen signed cross terms. Replay is zero; reporting energy is at most 5.2×10^{-14}. Display values are rounded.

Contrast	Metric	Original error	Dose on V	V/E gap	Precision	Cross sum
AB	R_G	6.68932	6.68148	0.0199902	5.96396e-09	-0.0121529
AB	C	35.1738	35.2717	0.120141	3.24926e-08	-0.218056
AB	Q	21.6901	21.5749	0.11124	2.40598e-08	0.0039162
AB	R_P	32.062	31.6393	0.0926356	1.57422e-07	0.330003
AM	R_G	1.05064	1.04634	0.000754874	9.49502e-10	0.00354004
AM	C	7.63671	7.58188	0.00573509	9.98524e-09	0.0490971
AM	Q	1.17701	1.21175	0.00203421	1.47438e-09	-0.0367685
AM	R_P	6.74694	6.67257	0.0384112	2.53505e-07	0.0359574

most 2.54×10^{-7} and finite-effect precision-change energy at most 7.33×10^{-11}. The largest absolute baseline log-metric change across cells/populations is 2.29×10^{-5}. The finite-effect V/E gap and this precision perturbation do not account for the large aggregate forecast-error scale. Individual cells can differ: the V/E term exceeds the V discrepancy in 30/288 rows. Shared FP32/FP64 code does not exclude a common implementation error; the V term need not be pure Taylor truncation. Small finite-effect gaps do not imply stable local derivatives across populations. Their response-projection difference has mean squared magnitude 6.30356 for A/B and 0.686944 for A/M. Using the post-result E projection gives respective E errors 14.5808 and 0.678707, without a uniform repair. These descriptive quantities do not replace the original V-based forecasts or their tests.

Full-path finite differences retain nonuniformity. For both fixed $h \in \{10^{-3}, 10^{-4}\}$, compute $f_h = [\log X(\theta_j(h)) - \log X(\theta_j(-h))]/(2h)$ and compare it with the FP64 analytic projection p_j. Table 14 reports $|f_h - p_j|/\max(1, |f_h|, |p_j|)$ for all 144 arm/population/cell combinations per metric and step size. The package also retains unscaled and relative discrepancies, all 1152 values and second differences.

Table 14: Both fixed full-path finite-difference steps, all cells. Smaller medians do not imply uniform accuracy. No best-step selection or outcome-based replacement occurs.

Metric	10^{-3} median	10^{-3} max	10^{-4} median	10^{-4} max
R_G	0.0009134	0.631299	3.74792e-05	0.157474
C	0.00137293	0.79008	4.95175e-05	0.191314
Q	0.00125932	0.43447	5.50612e-05	0.115526
R_P	0.00137835	1.27188	5.82824e-05	0.801012

At $h = 10^{-4}$, the overall median scaled discrepancy is 4.95×10^{-5}, but the maximum is 0.801: Cartpole cohort 4, M, V absolute RMSE gives $f_h = 16.6431$ versus $p_M = 3.31178$. Cartpole cohorts 4,10,11 and Reacher cohort 11 contain discrepancies above 0.01; that threshold is a descriptive locator, not a frozen pass criterion. The response-RMSE maximum is 0.15747. Both step sizes and every exception remain reported; 57/576 paired discrepancies increase at the smaller step. These calculations do not certify all local derivatives, provide a finite-dose guarantee, or select a new prospective rule.

Reproduction. Complete states, tangents, endpoint arrays and records are supplied. A separate NumPy reconstruction checks 4320 metrics, 288 decompositions, 1152 finite differences and 40 cohort summaries. CPU inference/terminal projection does not repeat training. Original fixed previews supply context, not new behavior.

A8 THE SEPARATE NATIVE COEFFICIENT BRIDGE

Five new complete repetitions (10–14) cover the fixed Meta-World 3.0.0 MT10 tasks: reach, push, pick-place, door-open, drawer-open, drawer-close, button-press-topdown, peg-insert-side, window-open and window-close, all v3. Each task has a separate model; these are not shared-policy MT10

scores. Each repetition has 32/16 training/diagnostic episodes of 500 actions, 256/64 quartets, two saved starts and two shared four-block plans, each block five native steps. Half hold actions, half switch them. Sampling uses no contact or outcome filter. Every arm receives the same ordinary trajectories and branch labels.

The direct split predictor uses three ensemble members, two width-143 hidden layers per branch and 189,306 total parameters. Each output is conditioned on the initial state, masked action prefix and horizon, never an intermediate model prediction. Training-only scales use four local increments with a 10^{-3} SD floor. The common coefficient κ is calibrated once from 16 training batches using the original D/S rule; no evaluation search or arm-specific recalibration occurs. All arms use 2,000 FP32 Adam updates, learning rate 10^{-3}, clipping at 10, with 64 ordinary examples and 16 whole quartets, with common initialization, bootstrap and batches. The 300 ensemble fits execute 600,000 updates and 2,480,000 native actions.

Table 15: Frozen auxiliary losses for the native bridge. Every row is added to the same $\mathcal{L}_{\mathrm{pred}}$ with the same multiplier κ. The first four form a 2×2 coefficient design.

Code	Auxiliary objective	Question
SG	$G = C + Q$	Paired reference
SWeakC	$\frac{1}{2}C + Q$	Weaken common weight only
SAdd	$C + \frac{3}{2}Q = G + D/8$	Add differential weight only
SX	$\frac{1}{2}C + \frac{3}{2}Q$	Original joint reweighting
SStrength	$a_t G$	Local auxiliary-gradient norm control
SAddperm	$G + D_\pi/8$	Fixed residual-mismatch control

SStrength sets $a_t = \mathrm{stopgrad}(\|\nabla(G + D/8)\|/\|\nabla G\|)$ at its own state. This matches the full auxiliary gradient norm, not direction, total gradient, Adam displacement or trajectory. All six arms compute both auxiliary gradients. Logged norm/clipping checks cover 41 positions per fit, not every update. SAddperm uses a fixed four-family cycle on second-state residuals; because different families have different action pairs, it disrupts action and state correspondence. It is not the action-preserving control of Appendix A1.

For each complete repetition, log response-RMSE ratios are averaged equally over ten tasks. The primary response pools all 64 families, both starts, four horizons and seven current-motion coordinates. Five t_4 contrasts—four direct comparisons and $(SX - SWeakC) - (SAdd - SG)$—share individual 99% intervals. The interaction is .000276 $[-.000512, .001063]$ in log units. The predeclared 1% reference is descriptive precision, not useful-control equivalence. No tasks, repetitions or large-error cells are removed.

SAdd/SG improves by .133% [.087,.179]; SX/SAdd instead has reduction $-.354\%$ $[-.604, -.106]$. SAdd has no resolved advantage over local strength or mismatch. Descriptive common/differential RMSE reductions for SAdd/SG are .100/.197%; SX/SAdd raises common RMSE .500%, with differential change unresolved. This supports a small coefficient-specific finding, not a large or correspondence-specific native gain.

All six direct models in Reach-r13 have first-block absolute hand errors 5.89–6.76 m, versus roughly 6–7 mm in the other four repeats. This failure occurs without recursive input feedback. It remains in every response aggregate. The scale intervention below measures its configuration sensitivity, without establishing a complete repair or explaining the controlled/native effect-size gap. Complete component, member/ensemble, physical-unit and horizon arrays are supplied; ensemble disagreement is not a train/test gap.

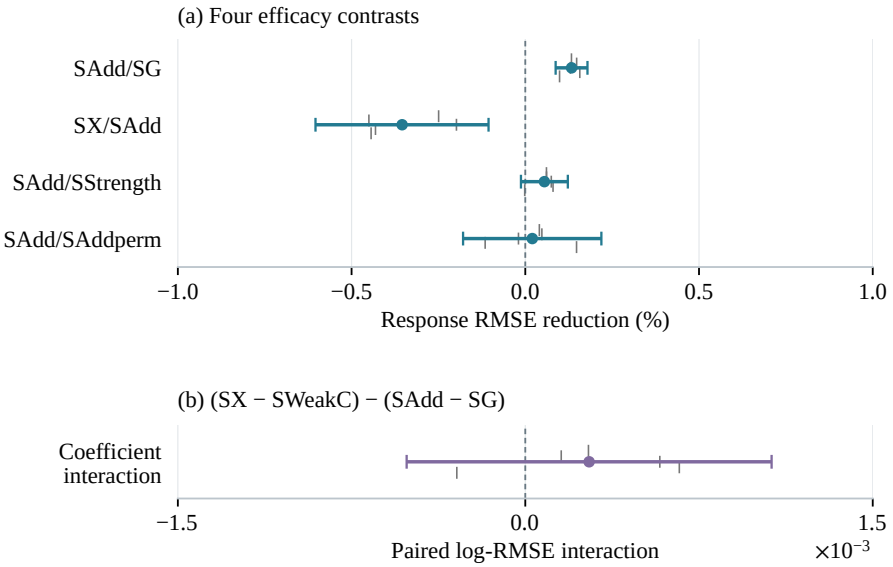

Figure 6: All five native coefficient comparisons and all five complete-suite repetition values. Direct comparisons report RMSE reduction; the interaction retains its log scale. Original family correction and the descriptive 1% reference remain.

A9 Output grouping as separate supporting evidence

This study holds the objective at prediction loss and changes output sharing. The 39-coordinate Meta-World observation contains 18 current coordinates, their preceding-native-step values and a known 3-coordinate goal. Each grouped network sees the complete normalized observation/action. Current-only C places indices 0–3 (hand XYZ/gripper) in one head; history-aligned H groups 0–3 with 18–21; mixed M uses 0–7. The second head predicts all remaining coordinates; only the known goal is copied. A five-step endpoint's history describes native step 4, not the initial state, so copying initial history would be incorrect.

Grouped predictors have two width-148 SiLU hidden layers per branch, three bootstrap members and 188,817 parameters. H/M have identical head shapes; all C/H/M hidden initializations match within a member. The dense width-212 reference has 188,373 parameters. Local training uses the true start of each five-step segment; recursive training propagates predictions through four segments and differentiates through all four. Evaluation always rolls out all four segments from the same initial observation. Equal parameter/update counts do not imply equal runtime.

After Push development, five new repetitions 5–9 regenerate all data, action plans, normalizers, bootstrap and initializations in the other nine tasks. Collection and training follow the 32/16-episode, 256/64-quartet, three-member, 2,000-update recipe above. The eight arms are dense/C/H/M under both local and recursive training: 360 ensemble fits, 720,000 updates and 2,232,000 native actions. This is fresh data within previously studied task families; Push is excluded by design, not removed after these results.

Six efficacy contrasts in main Table 21 plus two interactions share individual 99.375% t_4 intervals. Each independent value is the equal nine-task mean paired log response-RMSE ratio. HMP/CMP and HMP/MMP improve by 1.364% and 2.142%; HMP/RMP remains unresolved. Local H1P/R1P is adverse. The H/C and H/M recursive-minus-local log interactions are .1211 and .0253, transformed ratio changes 12.87% $[-3.60, 32.16]$ and 2.56% $[-4.48, 10.13]$. Neither establishes training-mode dependence.

The supplied task table retains all nine tasks, both modes, components and horizons. For example, HMP's object-response reductions against CMP/MMP are 3.00/4.46% with pointwise intervals crossing zero, despite the primary aggregate improvement; the descriptive HMP/RMP object gain

3.12% [.50,5.67] does not replace its unresolved aggregate comparison. Ordering decomposition finds that unchanged candidate order contributes 98.06/104.33% of HMP's hand-MSE gains over CMP/MMP; over 100% reflects an adverse changed-order contribution. This includes pairs where both orders are wrong. It is an arithmetic partition of an observed gain, not identification of a training mechanism or a successful decision rule. All 360 model populations and complete candidate outcomes remain in the package.

A10 What response accuracy changes in a specified selector

This post-hoc saved-data analysis uses all 180 recursive grouping ensembles and 300 bridge ensembles: 95 task/repetition cells, 64 families, two starts/candidates and horizons 5/10/15/20. It generates no new fits, neural queries or simulator labels. The studies retain separate five-repetition inference; 20 steps is the display, with every horizon and equal pooling supplied.

Let $m = (y_0 + y_1)/2$, $\delta = y_1 - y_0$ and define predicted counterparts. The four endpoints for candidate $a \in \{0, 1\}$ are

$$\hat{y}_a, \quad m + (a - \tfrac{1}{2})\hat{\delta}, \quad \hat{m} + (a - \tfrac{1}{2})\delta, \quad y_a.$$

They retain the original prediction, replace midpoint, replace response, or replace both. FP64 reconstruction promotes the saved FP32 ensemble mean; it does not re-average members. Reach chooses minimum hand-to-goal distance; other tasks choose object-to-goal distance. Exact predicted ties choose candidate zero. Geometric regret is selected true endpoint distance minus its minimum, in millimeters. Native regret uses the separately recorded prefix-mean rewards: maximum minus selected reward. Endpoint geometry and prefix reward differ in target and time aggregation.

For squared-distance selection the true margin is $2\delta^\top(m - g)$. With midpoint error e_m and response error e_δ, its prediction error is exactly

$$2\{e_\delta^\top(m - g) + \delta^\top e_m + e_\delta^\top e_m\}.$$

Thus exact response can coexist with an incorrect distance ordering through midpoint error. Full truth guarantees zero geometric regret, but need not maximize native reward. Its native regret is not a lower bound: an inaccurate geometric selector can choose the higher-reward candidate.

Population and inference. True nonzero margins define preference agreement or disagreement; a zero margin in either objective enters a separate tie stratum. These retrospective strata are not deployable conditions. Counts/numerators pool over fixed tasks before conditional division, so conditional task weights depend on eligible counts. The five complete repetitions, not members or pairs, give paired pointwise 95% t_4 intervals. Empty conditions remain undefined; no missing repetition receives a full-five interval. Descriptive intervals do not acquire primary-family correction.

True-response replacement lowers HMP geometric regret .704 to .109 mm. Its native-regret reduction is .12903 [.09050,.16756] in agreeing pairs and $-.11442$ [$-.13585, -.09299$] in disagreeing pairs; all-pair reduction .01754 [$-.00024, .03533$] remains unresolved. For SAdd, geometry improves .884 to .148 mm while the corresponding reward reductions are .13077 [.11176,.14979], $-.11807$ [$-.13064, -.10550$] and .01980 [.01263,.02698]. Full-truth all-pair reward regret remains .13204/.13196. The tables retain all ten methods and all four constructions.

Every focal between-method reward comparison still crosses zero even within preference-agreeing pairs. Objective disagreement therefore explains one readout failure but not all missing method gains. Positive pointwise HMP/CMP and HMP/RMP contrasts in truth-tie pairs remain in the full outputs; these do not establish general superiority. The older uniform object-goal proxy remains a historical analysis, not this corrected selector.

A10.1 Ties and candidate identity

Geometric-only ties have nonzero mean reward gaps .273231/.231662 in grouping/bridge. The original strict-zero convention retains all near-zero cases. Swapping candidate axes preserves every 841,267 non-tied selector choice/regret; 141,773 tied cases change physical choice because the first-index tie rule changes. These are evaluation cases, not independent repetitions. At 20 steps the

Table 16: Corrected task-geometry regret in millimeters, at twenty native steps. Entries are means over five complete repetitions, retaining every candidate pair; all ten models and all four endpoint constructions are shown. Full-truth geometry regret is zero by construction.

Arm	Predicted	True midpoint	True response	True endpoints
Nine-task grouping				
RMP	0.849	0.737	0.096	0.000
CMP	0.736	0.658	0.105	0.000
HMP	0.704	0.598	0.109	0.000
MMP	0.814	0.705	0.107	0.000
Ten-task loss bridge				
SG	0.864	0.759	0.147	0.000
SWeakC	0.878	0.761	0.150	0.000
SAdd	0.884	0.774	0.148	0.000
SX	0.876	0.747	0.150	0.000
SStrength	0.862	0.762	0.148	0.000
SAddperm	0.876	0.779	0.149	0.000

Table 17: Recorded native mean-reward regret for the same selections. Lower is better, but improved geometric fidelity need not lower native regret. Complete pointwise level intervals and paired differences are supplied as repetition-level CSVs; interval overlap must not be used as a paired test.

Arm	Predicted	True midpoint	True response	True endpoints
Nine-task grouping				
RMP	0.15201	0.15034	0.13124	0.13204
CMP	0.15293	0.15079	0.13135	0.13204
HMP	0.14882	0.14837	0.13128	0.13204
MMP	0.14528	0.14483	0.13066	0.13204
Ten-task loss bridge				
SG	0.16000	0.15596	0.13861	0.13196
SWeakC	0.16088	0.15649	0.13884	0.13196
SAdd	0.15865	0.15510	0.13885	0.13196
SX	0.15970	0.15651	0.13886	0.13196
SStrength	0.16014	0.15595	0.13861	0.13196
SAddperm	0.16017	0.15703	0.13830	0.13196

HMP/SAdd true-response all-pair reward regret rises .000286222/.003206305 after reversal, from geometric-only ties; agreement/disagreement-stratum results are unchanged. At earlier horizons nine predicted-distance ties occur even with nonzero true margins: six Peg-r10 arms at step 10 and SG/SWeakC/SAddperm in Peg-r13 at step 15. Floating-point-scale true margins and exchange effects remain in the data; truth non-ties do not guarantee selector non-ties.

Complete five-value and leave-one-repetition results retain every original native comparison. SAdd/SG leave-one reductions range .1272–.1418%; SAdd/SAddperm ranges $-.0117$–.0540%. HMP/CMP and HMP/MMP remain positive, while two of five HMP/RMP values are adverse. These are descriptive stability views with no replacement intervals, extra significance tests or new task filtering.

Table 18: The five focal method contrasts restricted to aligned non-tied pairs, at twenty steps, using original predictions. Gain is reference minus method native regret; positive favors the first method. These pointwise descriptive intervals all span zero.

Comparison	Aligned reward-regret gain	[pointwise 95%]
HMP/CMP	-0.00927	[-0.03487, 0.01632]
HMP/MMP	-0.01293	[-0.04665, 0.02079]
HMP/RMP	-0.00434	[-0.03497, 0.02630]
SAdd/SG	0.00292	[-0.00334, 0.00917]
SX/SAdd	-0.00282	[-0.01089, 0.00524]

Table 19: True preference populations at 20 steps, as percentages of all family/start-state pairs. The historical grouping study contains 5,760 pairs and the coefficient study 6,400 pairs across five repetitions. Each truth population is counted once, without multiplying by the number of models. "Geometry only" and "reward only" denote which true margin is zero; the other margin is strictly positive.

Study	Aligned	Misaligned	Geometry only	Reward only	Both ties
History grouping	37.378	23.160	21.424	15.434	2.604
Loss coefficients	36.750	22.016	24.672	13.734	2.828

A11 BOUNDED SENSITIVITY TO QUATERNION INPUT SCALING

Saved-data localization. The original history-grouping Reach r6 tail concentrates in diagnostic episode four; the independent loss-bridge Reach r13 tail concentrates in episode eleven (families 44–47). The native observation includes the object's unit quaternion at coordinates 7–10 and its preceding-native-step value at 25–28, in scalar-last order. Coordinate nine is an object-quaternion component, not a hand angle. Both datasets have nearly constant training rotation and the original 10^{-3} minimum standard deviation. In bridge r13, diagnostic values reach a standardized input magnitude of 1150.867, whereas the other four new Reach repetitions reach approximately 3–4. Episode eleven's quaternion norms are one to floating-point precision; saved initial observation, integration state and history match the corresponding original trajectory exactly. Physical object rotation exists before the sign changes in its quaternion representation. These checks do not support indexing, state-restoration or non-unit-quaternion explanations, and a quaternion sign convention is not the intervention below.

The original r6 mixed-input queries established sensitivity of one fixed predictor. The r13 localization initially established only an association between the input support and errors. Neither alone identified a shared training cause. We therefore perform one bounded real-input retraining comparison rather than treating the original severe tails as representation-independent phenomena.

Fixed twelve-fit design. The matrix is Reach r6/r7 $\times$ RMP/CMP/HMP/MMP and r12/r13 $\times$ SG/SAdd: twelve three-member ensemble fits, each with 2,000 updates (24,000 total). The known r6/r13 cases and adjacent-number references r7/r12 are fixed before this study's results. References are previously examined data, not unseen or guaranteed-normal controls. Original completed fits provide the untreated comparisons; no original model or result is overwritten. No simulator interaction, data collection, coefficient calibration, additional seed or scale grid is used.

The sole treatment sets the eight quaternion *input* standard deviations to the natural unit scale one. Training means, other input scales, all output/label scales, original scientific seeds, initial parameters, bootstrap and batch streams, Adam settings, gradient clipping, loss coefficients, 2,000 updates and FP32 arithmetic remain unchanged. Raw observations and true labels are neither replaced nor projected. Two fixed engineering representatives (r6 HMP and r13 SAdd) verify the initial parameters, buffers, ten batches and output scales exactly before the formal matrix; their twenty updates are recorded separately. There is no recalibration to compensate for the input treatment.

All 64 families, two starts, two candidates, four horizons and three members are evaluated. New recursive models receive the same four local feedback queries, anchored to each member's own new

free trajectory; direct S models have no artificial feedback condition. Results are descriptive per-cell configuration sensitivity, with no new significance tests or two-repeat confirmation. Original nine/ten-task primary inference retains its original configuration and correction.

Absolute tails shrink substantially but remain unequal. Table 22 retains all twelve cells. History-aligned r6 first-block hand absolute RMSE falls from 1109.417 to 8.847 mm, while response RMSE changes from 3.077 to 2.710 mm. Its twenty-step absolute error falls from 3541.995 to 18.232 mm. Direct SG r13 first-block absolute error falls from 6504.359 to 53.264 mm, while response error changes from 6.590 to 6.351 mm. The original severe magnitude is therefore strongly sensitive to input scaling in both specific cases. This does not make every prediction reliable: within the known r13 episode, SG still has 211.579 mm first-block absolute error, compared with 6.464 mm in the other sixty families. For HMP r6, the corresponding values are 25.562 and 6.318 mm. These strata locate errors and are never excluded from the full results.

Local comparisons depend on configuration. For seven-coordinate response RMSE pooled over four horizons, HMP's reductions against CMP/MMP remain positive in both selected repetitions: 1.56/0.98% becomes 1.69/1.22% in r6, and 2.73/3.91% becomes 2.24/3.29% in r7. Its comparison with dense RMP reverses in both cells: -0.47% becomes 0.01% in r6, while 0.87% becomes -0.93% in r7. SAdd/SG changes from -0.0028% to 0.0295% in r13, while r12 remains positive (0.2553% to 0.2680%). These small per-cell differences are not a new confirmation of any ordering and do not replace the original full-suite estimates. They show why a large absolute repair must not be renamed a universal method advantage.

A11.1 Definition of local feedback queries

At a fixed model's original rollout inputs, candidate errors are e^a, e^b, common error $b = (e^a + e^b)/2$ and differential errors $d^j = e^j - b$. With identical actions, query the next-block map at true input, true plus b, true plus d^j, and original predicted input. Their contrasts are $R_{00}, R_{10}, R_{01}, R_{11}$. For true contrast Δ,

$$L = R_{00} - \Delta, \quad B = R_{10} - R_{00}, \quad D = R_{01} - R_{00}, \quad I = R_{11} - R_{10} - R_{01} + R_{00}.$$

Then $R_{11} - \Delta = L + B + D + I$. Squaring retains four energies and six signed cross terms. These are four finite queries on one original trajectory, not four new recursive rollouts, additive causal shares or the across-start C/Q decomposition. Native members use their own 39-coordinate trajectories at times 0/5/10/15, with history referring to the preceding native step. Original FP32 inference/member averaging is retained; vector accounting is FP64. Direct-prefix models have no intermediate input feedback and are excluded.

The full diagnostic includes 30 original controlled models and 180 native grouping ensembles, with every family/horizon. At the first step all feedback terms are zero. Common-input inclusion changes direction by time/environment: the original continuous paired model's signed MSE change is -9.37×10^{-5} at step 4 and 3.19×10^{-3} at step 16, with pointwise intervals excluding zero in opposite directions. Hybrid inputs can be physically unattainable; each model has different upstream errors and a different receiving map. These queries do not isolate a training mediator.

Cancellation persists with smaller magnitude. Table 20 retains all eight recursive models. In HMP r6, true/original-input twenty-step hand-response RMSE changes from 148.994/12.448 mm to 16.561/11.046 mm. Its local–differential cross term changes from -44198.248 to $-487.868\,\text{mm}^2$. The cancellation example persists in this bounded alternative configuration, but its magnitude is strongly configuration dependent. Within the known episode-four stratum itself, true/original-input RMSE changes from 592.347/31.636 mm to 15.629/19.082 mm: the original net cancellation reverses there. The remaining sixty families still show 16.622/10.290 mm under unit scale, and reference r7 HMP shows 15.733/9.728 mm. Thus full-population persistence does not mean the original anomalous-episode behavior persists. Signs of other finite input-removal differences also change, so no monotone harmful-feedback rule follows. Every first-block feedback component remains zero. The queries measure fixed-function response at original trajectory inputs and do not identify a training mediator, deployable repair or cross-time causal share.

The companion archive supplies four Reach cells, 12 original and 12 treated final ensembles, scales, calibrations and saved outputs. Extracted inference replay completed 48 split predictions; all 40 his-

Table 20: All eight recursive predictors' twenty-step queries before and after the input-scale change. Complete-family hand-response RMSE and the signed local–differential cross term use the same original target units. All four component energies and six cross terms, all horizons and fixed episode strata are in the supplied tables.

Repeat / model	Scale	True	Common only	Differential only	Original	$2\langle L, D\rangle$
		Hand-response RMSE (mm)				(mm^2)
r6 RMP	original	144.410	107.749	13.986	13.036	-41314.774
r6 RMP	unit1	14.518	14.406	10.984	10.931	-340.414
r6 CMP	original	110.729	176.521	14.712	15.292	-23916.692
r6 CMP	unit1	15.686	15.460	12.360	12.518	-388.126
r6 HMP	original	148.994	152.765	12.649	12.448	-44198.248
r6 HMP	unit1	16.561	16.204	11.493	11.046	-487.868
r6 MMP	original	129.548	171.151	14.349	13.501	-33125.135
r6 MMP	unit1	15.701	15.387	12.099	12.221	-392.244
r7 RMP	original	14.386	14.498	10.048	10.077	-329.464
r7 RMP	unit1	14.199	14.289	10.095	10.077	-311.878
r7 CMP	original	14.463	14.464	10.382	10.514	-303.984
r7 CMP	unit1	14.358	14.249	10.463	10.502	-301.646
r7 HMP	original	15.611	15.808	9.564	9.488	-397.289
r7 HMP	unit1	15.733	16.000	9.731	9.728	-406.026
r7 MMP	original	14.769	14.671	10.646	10.710	-320.246
r7 MMP	unit1	14.551	14.392	10.595	10.508	-295.865

torically saved arrays matched absolute/relative tolerance 3×10^{-5}, maximum absolute discrepancy 3.815×10^{-6}. The eight other original-history training outputs had no saved reference. All 16 recursive variants' four-query member outputs matched exactly. This verifies inference, not independent training; original final weights lack Adam state, which inference does not require.

A12 Supporting grouping and input-scale summaries

A12.1 Response improvements have separate readout limits

A supporting experiment fixes the objective at prediction loss and changes the output partition across nine fixed Meta-World tasks and five fresh repetitions. Table 21 retains every primary efficacy contrast. History-aligned grouping improves over two recursive partitions, but not demonstrably over dense prediction; local training instead favors the dense reference. Both training-mode interactions remain unresolved. These results concern output partitioning under a fixed prediction objective.

Table 21: All grouping efficacy comparisons: history-aligned versus each reference. Response-RMSE reduction (%), positive for improvement; individual 99.375% intervals from the original eight-test family. Complete data and both interactions are described in Appendix A9.

Training	Current-only	Mixed	Dense
Local	12.61 [−1.85,25.02]	4.59 [−3.35,11.92]	−2.54 [−4.00,−1.10]
Recursive	1.36 [0.23,2.48]	2.14 [0.57,3.69]	0.58 [−2.09,3.18]

The recursive predictors RMP/CMP/HMP/MMP use dense/current/history/mixed output partitions, respectively. On all 480 saved loss/grouping models, a selector minimizes endpoint hand-to-goal distance for Reach and object-to-goal distance otherwise; native regret uses mean prefix reward. Replacing HMP's response by truth lowers twenty-step geometric regret from 0.704 to 0.109 mm, yet reward-regret reduction changes sign with true-preference agreement: 0.1290 [0.0905,0.1676] when agreeing, versus -0.1144 $[-0.1359, -0.0930]$ when disagreeing, in pointwise descriptive intervals. The coefficient-study cohort has the same directional split. The strata use future truth and therefore describe retrospective associations.

Ties alter some truth-replacement outcomes. None of the five focal method contrasts establishes reward benefit, overall or within the agreeing stratum (Appendices A10 and A10.1).

A12.2 SMALL RESPONSE ERROR CAN COEXIST WITH A LARGE ABSOLUTE FAILURE

Action subtraction removes shared endpoint error. In the known Reach r6 cell, HMP has 3.077 mm first-block response RMSE alongside 1109.417 mm absolute RMSE; all four recursive predictors fail in that cell. A separate direct-prefix SG model has 6504.359 mm first-block absolute error in r13, before any recursive feedback. These failures cannot be attributed solely to feedback.

A fixed twelve-fit sensitivity test changes only current/history object-quaternion input scales to one. Table 22 reports all matched comparisons. The known tails shrink, with substantial residual error: SG r13 retains 211.58 mm error inside its exceptional episode. Under the changed scale, HMP r6's true/original-input response RMSE is 16.56/11.05 mm at twenty steps: population cancellation remains while the exceptional episode reverses direction (Appendices A11 and A11.1).

Table 22: All twelve input-scale comparisons. Hand-XYZ RMSE in millimeters; each cell keeps all 64 families. Absolute errors use the stated endpoint; response pools all four horizons. "unit1" changes only quaternion input scales. RMP/CMP/HMP/MMP are dense/current/history/mixed recursive predictors; SG/SAdd are direct coefficient-study models.

Study / repeat	Predictor	5-step absolute Original → unit1	20-step absolute Original → unit1	Pooled response Original → unit1
Grouping r6	RMP	735.208 → 8.575	1952.173 → 10.214	8.967 → 7.416
Grouping r6	CMP	857.105 → 7.541	1316.043 → 13.269	10.051 → 8.270
Grouping r6	HMP	1109.417 → 8.847	3541.995 → 18.232	8.628 → 7.446
Grouping r6	MMP	918.188 → 7.546	2214.520 → 15.857	9.295 → 8.074
Grouping r7	RMP	5.570 → 5.535	8.696 → 8.477	6.755 → 6.702
Grouping r7	CMP	5.918 → 5.902	9.135 → 9.046	7.093 → 7.152
Grouping r7	HMP	5.680 → 5.756	8.436 → 8.586	6.433 → 6.585
Grouping r7	MMP	5.993 → 5.997	9.235 → 8.970	7.189 → 7.174
Loss r12	SG	6.161 → 6.171	10.025 → 10.000	8.438 → 8.447
Loss r12	SAdd	6.152 → 6.162	10.017 → 9.992	8.426 → 8.434
Loss r13	SG	6504.359 → 53.264	6505.849 → 61.831	10.946 → 10.220
Loss r13	SAdd	6758.631 → 53.618	6760.263 → 62.185	10.891 → 10.196

A13 DEFINITIONS, ALGEBRA AND CONTROLLED DYNAMICS

Residual identities and their limits. Expanding $e_1 = c - q, e_2 = c + q$ proves $G = C + Q, D = 4Q$ with identical pair/coordinate weights. Unequal state weights introduce $(a_2 - a_1)\langle c, q\rangle$. With positive supervised state-loss weights, a feasible exact fit makes every term zero; adding the cross-state term does not change those zero-loss solutions. The contrast alone vanishes for any common residual $e_1 = e_2 = v$, whose paired loss can be arbitrarily large. The anchor inequality $\|e_2\| \le \|e_1\| + \|e_2 - e_1\|$ and its L^2 version require an anchor on the same pair population, not unobserved-state generalization.

For uniformly weighted $K \ge 2$ candidate vectors,

$$\frac{1}{\binom{K}{2}} \sum_{i<j} \|v_i - v_j\|^2 = \frac{2K}{K-1} \frac{1}{K} \sum_i \|v_i - \bar{v}\|^2.$$

Expand the pair sum as $K \sum_i \|v_i\|^2 - \|\sum_i v_i\|^2$. Applied to residuals $r_{mi} = \hat{y}_{mi} - y_{mi}$, it gives

$$\mathcal{L}_{\text{pred}}^{\text{branch}} = B_{\text{abs}} + \frac{K-1}{2K} G, \qquad B_{\text{abs}} = \frac{\|\bar{r}_1\|^2 + \|\bar{r}_2\|^2}{2d_e}.$$

The coordinates/population must match; native $\mathcal{L}_{\text{pred}}$ also uses ordinary trajectories. B_{abs} is error shared across actions within a state, whereas C is effect error shared across states. Neither is the selector's midpoint error. Selective pair weights generally require explicit pairs.

For smooth F, the finite state/action contrast is

$$F(s+h,u+k) - F(s+h,u) - F(s,u+k) + F(s,u) = \int_0^1 \int_0^1 D_s D_u F(s+ah, u+bk)[h,k]\, db\, da.$$

Apply the fundamental theorem of calculus along u, then s. A uniform mixed-derivative bound M gives contrast norm at most $M\|h\|\|k\|$. Small finite residuals do not bound pointwise derivative error because the integral may cancel; discrete/nonsmooth dynamics retain the finite-contrast definition.

Orthogonal output components share the predictor Jacobian, so their parameter gradients need not be orthogonal. For an L-smooth early loss and actual displacement d, the absolute Taylor remainder is at most $L\|d\|^2/2$. An SGD update $d = -\eta(g+h)$, with early/late gradients g, h, has first-order early change $-\eta(\|g\|^2 + \langle g, h\rangle)$; sufficiently adverse alignment can increase early loss while preserving its coefficient. AdamW requires its actual clipped, moment-dependent displacement. This local statement is neither a full-training mediator nor a convergence guarantee.

Qualified effect metrics. For active pairs $\mathcal{A}_t$ and equivalent pairs $\mathcal{E}_t$,

$$E_t = \sqrt{\frac{\sum_{\mathcal{A}_t} \|(\hat{y}_i - \hat{y}_j) - (y_i - y_j)\|^2}{\sum_{\mathcal{A}_t} \|y_i - y_j\|^2}}, \qquad F_t = \sqrt{\frac{\sum_{\mathcal{E}_t} \|\hat{y}_i - \hat{y}_j\|^2}{d\,|\mathcal{E}_t|}}.$$

Equivalent pairs have distinct command prefixes but identical full native paths. Qualification uses truth, never predictions; endpoint convergence, object-only equality and visual near-equivalence are separate categories. A zero denominator stays undefined. An action-independent predictor has $F = 0$ and $E = 1$ on nonempty active populations, so neither metric substitutes for the other. The original studies use full standardized coordinates; fresh/midpoint motion metrics explicitly remove gate/mode outputs. Sufficient statistics sum within each original family/seed before the ratio and square root. Original percentages compare arithmetic seed means; new log-ratio studies use geometric effects. These aggregations are not interchangeable.

Prescribed numerical dynamics. The continuous six-state system has two positions, two velocities, gate and mode. Five modes implement independent two-axis drive, redundant summed horizontal drive, gate-multiplied drive, saturation to $[-.5, .5]$, and weak $.1\tanh$ drive. Velocity retention is .9 with .1 drive increment; position advances .1 times the new velocity. At fixed mode/gate the action contrast is independent of base positions/velocities. Gate-flipped pairs mix nonzero gated interactions with invariant modes. These are custom units, not calibrated robot coordinates or benchmark returns.

The older discrete five-state system has two grid positions clipped to $[-16, 15]$, gate, timer and mode. Eight commands include wait, moves, gate toggle, a redundant horizontal command and timer activation; wait can move under an active timer. Original short/long losses both used normalization from all 16 training futures: later evaluation was beyond supervision but not beyond normalization. The fresh study changed that design explicitly. Scientific configurations and eligibility masks retain these distinctions.

A14 Coordinate conventions and recorded computation

The prediction loss $\mathcal{L}_{\text{pred}}$ is distinct from the paired controlled arm P and the native paired arm B. B_{abs} denotes candidate-mean absolute-residual energy in the algebra. The native sensitivity dose is ρ, leaving α for the within-state auxiliary coefficient. Original protocol and data files retain their recorded names, including P for prediction loss and alpha for dose.

Recorded resource use for the physical intervention. Table 24 summarizes all 216 completed fits from their original training receipts. Training wall time covers the update loop, logging and checkpoint writes, excluding separately recorded initialization. Runs used a shared GPU with concurrent work, so these times describe the recorded execution load rather than a controlled method-speed comparison. Allocator peaks cover process initialization and training and exclude memory outside the PyTorch allocator. The companion CSV retains each cohort, world and arm, initialization/training times, and allocated/reserved peaks.

Loss components and pre-clipping gradient norms were logged at the first update and every fifty updates; these sparse records do not recover an all-update clipping frequency. Actual per-update parameter-displacement norms and validation curves were not logged. The endpoint study evaluates fixed terminal models, so the reported ordering is a fixed-budget fitting result rather than a convergence comparison. This resource summary introduces no new fits, model queries or environment interactions.

Table 23: Coordinate and residual-energy conventions. Motion endpoints and training objectives need not use the same coordinates or normalization.

Study	Training / representation	Reported effect energies
Fresh continuous correspondence, physical intervention and midpoint	Six state coordinates; training effect terms sum coordinates and divide by true-energy scales	Four motion coordinates; C, Q, G sum coordinates, with separate active/false-response definitions
Earlier controlled studies	Six continuous or five discrete state coordinates	Original full-state endpoints; motion-only reanalyses are labeled separately
Earlier Meta-World coefficient bridge	Full observation prediction; seven current-motion effect coordinates	Seven-coordinate response endpoints, with original horizon and repetition aggregation
Native forecast study	Padded 64-state input; effective dynamic dimensions 4/4/38 for Cartpole/Reacher/Push	Effective dynamic coordinates; C, Q, G use coordinate means; copied context and padding are excluded
Output grouping and readout diagnostics	Prediction loss; current/history output partitions	Seven-coordinate response or hand/object XYZ and specified readouts, as labeled

Table 24: Recorded resources, with 2,000 updates per fit. The final column is the maximum PyTorch allocated-memory peak among the eighteen fits, not total device memory.

Physics	Arm	Fits	Training seconds: median [range]	Peak allocated MiB
0	P	18	30.6 [28.6, 34.7]	73.6
0	A	18	32.1 [30.8, 36.2]	73.5
0	M	18	65.2 [59.2, 68.8]	73.5
0	$M_\times$	18	65.5 [59.9, 72.6]	73.5
.5	P	18	30.8 [28.5, 32.6]	73.6
.5	A	18	33.4 [30.3, 35.5]	73.5
.5	M	18	65.8 [58.0, 68.6]	73.5
.5	$M_\times$	18	66.7 [58.2, 71.8]	73.5
1	P	18	31.4 [28.3, 34.4]	73.6
1	A	18	32.7 [30.3, 36.2]	73.5
1	M	18	64.5 [59.8, 68.6]	73.5
1	$M_\times$	18	66.5 [59.1, 69.0]	73.5

A15 Earlier controls and unsuccessful extensions

These studies motivated the central tests but are not extra repetitions of them. Their complete original comparisons and numerical entries remain supplied. RMP/DMP and the other project variants are internal references, not independently published external baselines.

Action-mismatch diagnostic. Before the fresh action-preserving test, 20 fits reused five original initializations, 2,048 training families and 1,024 evaluation families. Four arms used 2,000 updates and original 16-step normalization. The mismatch cycled second-state candidate indices, moving predictions with their own labels; it changed action identity. The local-strength control matched the added gradient before clipping at its own state. Motion results used 30 tests, 112,000 crossed seed/family bootstrap draws and linear-quantile Bonferroni tails (about 93 draws per tail). This is not fresh-cohort confirmation or campaign-wide error control.

Alignment reduces step 4/16 active error 24.16/53.32% against action mismatch and 15.42/41.10% against local strength. Four-step false response and common energy against strength remain unresolved: corrected absolute-difference intervals $[-.011199, .000629]$ and $[-.003035, .000205]$. This supports a gain against the tested mixed action/state control, not necessity of precise state correspondence.

Temporal and architecture failures. The original four-step controlled families have 28 comparisons and crossed seed/family bootstrap inference. Short-preserved training weights steps 1–4 by

Control	Step / metric	Aligned	Control	Change (%)	Corrected Δ interval
Paired	4 / E_{active}	0.063016	0.103700	-39.23	$[-0.060715, -0.027616]$
Paired	4 / F_{equiv}	0.021729	0.045875	-52.63	$[-0.049886, -0.008953]$
Paired	4 / C	0.005405	0.011548	-53.20	$[-0.010839, -0.002715]$
Paired	4 / Q	0.000147	0.004266	-96.56	$[-0.007130, -0.002747]$
Paired	4 / G	0.005552	0.015814	-64.89	$[-0.017841, -0.005609]$
Paired	16 / E_{active}	0.085728	0.163530	-47.58	$[-0.092503, -0.055011]$
Paired	16 / F_{equiv}	0.053227	0.140502	-62.12	$[-0.128047, -0.047346]$
Paired	16 / C	0.021857	0.052471	-58.34	$[-0.043013, -0.013908]$
Paired	16 / Q	0.002887	0.043534	-93.37	$[-0.057375, -0.024887]$
Paired	16 / G	0.024744	0.096005	-74.23	$[-0.090081, -0.041062]$
Re-paired	4 / E_{active}	0.063016	0.083095	-24.16	$[-0.032271, -0.008848]$
Re-paired	4 / F_{equiv}	0.021729	0.031481	-30.98	$[-0.017947, -0.004776]$
Re-paired	4 / C	0.005405	0.008043	-32.80	$[-0.005081, -0.000384]$
Re-paired	4 / Q	0.000147	0.001802	-91.85	$[-0.002507, -0.001220]$
Re-paired	4 / G	0.005552	0.009845	-43.61	$[-0.007509, -0.001722]$
Re-paired	16 / E_{active}	0.085728	0.183634	-53.32	$[-0.129532, -0.065441]$
Re-paired	16 / F_{equiv}	0.053227	0.269609	-80.26	$[-0.353920, -0.079193]$
Re-paired	16 / C	0.021857	0.077887	-71.94	$[-0.088542, -0.021645]$
Re-paired	16 / Q	0.002887	0.073729	-96.08	$[-0.101249, -0.033651]$
Re-paired	16 / G	0.024744	0.151616	-83.68	$[-0.186128, -0.057815]$
Strength	4 / E_{active}	0.063016	0.074506	-15.42	$[-0.021752, -0.003881]$
Strength	4 / F_{equiv}	0.021729	0.025552	-14.96	$[-0.011199, +0.000629]$
Strength	4 / C	0.005405	0.006587	-17.94	$[-0.003035, +0.000205]$
Strength	4 / Q	0.000147	0.001270	-88.43	$[-0.001699, -0.000794]$
Strength	4 / G	0.005552	0.007856	-29.33	$[-0.004692, -0.000682]$
Strength	16 / E_{active}	0.085728	0.145560	-41.10	$[-0.090848, -0.036756]$
Strength	16 / F_{equiv}	0.053227	0.173153	-69.26	$[-0.184989, -0.034804]$
Strength	16 / C	0.021857	0.047312	-53.80	$[-0.045179, -0.007129]$
Strength	16 / Q	0.002887	0.038484	-92.50	$[-0.055276, -0.017420]$
Strength	16 / G	0.024744	0.085796	-71.16	$[-0.098072, -0.024627]$

Table 25: All 30 original diagnostic contrasts. Aligned minus control; negative favors alignment. Percentages compare arithmetic seed means, with the 30-test bootstrap family.

1/4 and 5–16 by 1/48, total 1.25; its uniform-long control uses 1.25/16 throughout. Preserving early coefficients did not preserve short-model accuracy. Normalized protection increases step 4 active error 5.75/3.56% in continuous/discrete confirmation while lowering false response 18.39/7.90%. At equal updates, discrete paired 16-step error improves 8.61%, its cross-state gain is unresolved, and continuous cross-state error worsens. Update count alone is insufficient. Real intermediate-start injection did not establish a16-step active-error rescue; truth injections can make unattainable hybrid states and are not additive causal shares.

The observed-gain Pendulum variant shows a modest gain, but numeric PushT and both recurrent conditions increase active error. Their horizon/budget also differ, so recurrence is not isolated as the cause. Numeric PushT lacks eligible full-state strict equivalents, making its specified protection inactive. In visual PushT, all 320 challenge families have identical paired full physical states and futures from deterministic collection replay. This is not aliasing between different physical states: same-state response is measurable, but cross-state transfer was not tested.

Earlier native and grouping development. The ten-task study found SX/RMP and DP/RMP error increases 12.07/13.26%, with cross-state and partition increments unresolved. Its original five-test corrected results remain in the package. Fixed Push grouping yielded local H1P/C1P and H1P/M1P reductions 42.13/16.26%, but the separate descriptive H1P/R1P comparison was adverse 5.20%. Recursive HMP/CMP and HMP/MMP gains were unresolved; HMP/RMP and CMP/RMP had nearly equal descriptive reductions .914/.910%. Push therefore motivated the separate nine-task confirmation, not a universal grouping claim.

Earlier native Push had 43.28% differential residual share without a resolved cross-state advantage. Available differential headroom is not sufficient for realized benefit. The CQM/WorldSync discussion distinguishes objectives; it is not a head-to-head defeat of their full systems. Output sharing and differentiation through optimization have established precedents. Scientific value here comes from the controlled empirical distinctions, not algebraic novelty.

A16 CHRONOLOGY AND REPRODUCIBLE SCOPE

The fresh 18-cohort control followed earlier results; the midpoint study reused those cohorts. The physical study separately fixes 36 new cohorts, all 216 fits and its 3/60 contrast families before results. Native forecasts use new fitting/validation/evaluation streams and a global evaluation barrier after all 144 terminal models and 36 forecasts. No outcome triggered cohort replacement, a new test or forecast calibration. Fixed task sets are not random samples of unseen tasks. Per-study correction does not remove selection across the research history. The forecast-failure decomposition then reuses these cohorts in 396 FP64 fits with fixed signed/unit doses; it is post-result, not new confirmation. All post-hoc diagnostics remain labeled.

Historical bootstrap families preserve their paired family draws: 56 comparisons use two 28-test families, 112,000 draws, batch 256 and linear quantiles; other factorial families use 200,000 draws. The 30-test diagnostic and newer paired-t families remain separate, with their original populations.

From the anonymous package root, run

```
python3 reproducibility/recompute_primary.py \
  --output build/primary-reproduction
```

After extracting the code and data archives into the same directory, the CPU entries reconstruct the frozen contrast families and descriptive readouts from complete supplied statistics. They need no model, simulator, private path or tracking account. `reproducibility/PRIMARY_REPRODUCTION.md` maps commands to studies; extracted-package execution and direct comparison verify the reported outputs.

Table 26: Reproduction levels. Each level requires the assets and computation listed here.

Level	Included population/assets	Boundary
Statistics	Complete arrays, comparisons, frozen configs and CPU code	Reconstructs numbers, not training
Fresh controlled	126 final models, exact data-generation streams	Separate inference/retraining commands
Physical correspondence	216 final models, 36 calibration streams, complete predictions	All-model CPU replay; retraining separate
Midpoint	36 full parameter/Adam/RNG sources, 144 final models, 324 family arrays	Explicit continuation; reused cohorts
Native forecast	108 data arrays, 36 forecasts/prediction cells, 144 final and 36 initial models	Inference self-contained; full training/recollection needs original runtime/snapshots
Forecast diagnostic	396 final states, all V/E endpoints, original data reused	Statistics/inference separate from full-path training
Reach scale/feedback	Four cells, 24 ensembles, calibration and references	Companion archive; original final models lack Adam
Historical native/visual	Statistical inputs and versioned code	Full raw data/weights remain separate

The 36 midpoint sources retain parameter/Adam tensors, counters and RNG states; learning rates use value-identical Python scalars for restricted loading. Full plots and fixed-index lossless samples accompany every study; samples do not replace complete-population inference.