# MORPHADAPT: TWO-TIMESCALE CONTEXT FOR SHARED CONTROL ON UNSEEN ROBOTS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Shared robot control faces context variation at two timescales: embodiment changes across robots, while interaction dynamics can change within deployment. MorphAdapt assigns these information roles to a static morphology pathway and a temporal adaptation pathway. A morphology-conditioned shared controller consumes context inferred by a temporal Student from recent observations and previous actions; privileged dynamics supervise training but are unavailable at deployment. We evaluate this system jointly across unseen embodiments and altered dynamics on 97 held-out UNIMAL robots spanning 87 exact-XML morphology clusters. MorphAdapt achieves 9.3%–16.4% higher mean episodic return than MetaMorph-DR in four dynamics-mutation conditions, together with lower point estimates of maximum normalized velocity drop and recovered-only recovery time. MetaMorph-DR provides a system baseline without online temporal adaptation; a matched State-only Student separately tests the value of explicit action history. Previous-action conditioning yields substantial post-perturbation point gains in an individual development realization, while its incremental effect remains heterogeneous in the matched input ablation. The results support the complete two-timescale system on this benchmark and identify reliable use of action history across training realizations as an open learning problem.

## 1 INTRODUCTION

Robot control faces context variation at two timescales: embodiment changes across robots, while dynamics change within deployment. Limb counts, joint arrangements, and nominal physical properties describe which robot is being controlled. Actuator strength, ground friction, and carried mass can change how that same robot responds during interaction. A shared controller therefore needs both an embodiment interface and a way to infer its current interaction context.

Shared and morphology-conditioned policies provide interfaces for sharing control across embodiments (Huang et al., 2020; Gupta et al., 2022; Xiong et al., 2023). History-based adaptation supplies deployment-time context from recent interaction (Yu et al., 2017; Kumar et al., 2021). Their conjunction raises our primary question: *can temporal adaptation learned across heterogeneous robots remain effective on unseen embodiments when interaction dynamics change during deployment?* A secondary question concerns the temporal input: *what additional value does explicit previous-action conditioning provide beyond state history?*

MorphAdapt (Fig. 1) uses *two-timescale context factorization*: static embodiment information enters a morphology pathway, while recent interaction enters a temporal adaptation pathway. This is a factorization of information roles and variation timescales. Nominal attributes condition a shared per-limb controller; a temporal Student infers context from observations paired with previous policy actions and fuses it with the controller's limb representations. Actions supply an explicit input–response relation. A privileged Teacher supervises training, while deployment uses causal history and fixed network weights. During Student training, randomized dynamics remain fixed within each episode; evaluation introduces within-episode dynamics changes on held-out bodies without deployment-time parameter updates.

At the system level, MorphAdapt obtains 9.3%–16.4% higher mean return than MetaMorph-DR on 97 held-out robot IDs across four reset conditions under persistent dynamics mutation. This comparison



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

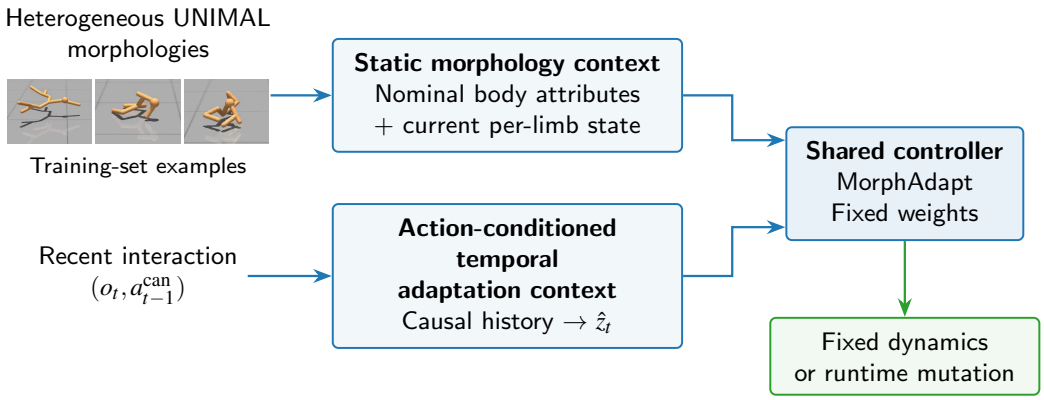


Figure 1: **Two context pathways, one shared controller.** Static morphology context and action-conditioned temporal adaptation context support shared control under fixed dynamics or runtime mutation. The MuJoCo renders are UNIMAL training-set examples; evaluation uses held-out robot instances with no exact-XML overlap with training.

measures the complete training and deployment pipeline. Controlled input and context-update diagnostics then examine the information sources and scope of these system-level gains.

Our contributions are:

1. We formulate heterogeneous shared control through static embodiment context and time-varying interaction context, with separate information roles across variation timescales.
2. We integrate a morphology-conditioned shared controller with a causal temporal Student and evaluate the joint condition of unseen morphology and altered dynamics.
3. We obtain 9.3%–16.4% higher mean return under dynamics mutation than the no-online-temporal-adaptation system baseline, together with lower maximum normalized post-mutation velocity drop.
4. We use action-input, history-horizon, and context-update diagnostics to distinguish complete-system gains from the incremental contribution of explicit action history.

## 2 Related Work

**Shared control across morphologies.** NerveNet uses graph structure, while shared modular policies reuse local control modules (Wang et al., 2018; Huang et al., 2020). Amorpheus and AnyMorph investigate attention-based interfaces for incompatible or variable morphologies (Kurin et al., 2021; Trabucco et al., 2022). MetaMorph represents limbs with morphology-conditioned Transformer tokens (Gupta et al., 2022); Body Transformer studies structure-informed attention (Sferrazza et al., 2025). ModuMorph introduces contextual modulation, and HyperDistill distills morphology-conditioned controllers for efficient deployment (Xiong et al., 2023; 2024). URMA studies end-to-end multi-embodiment locomotion (Bohlinger et al., 2025). MorphAdapt builds on shared morphology interfaces and supplies an additional context inferred from recent interaction.

**Dynamics adaptation and privileged supervision.** UP-OSI combines a universal policy with history-based system identification (Yu et al., 2017). RMA learns deployment-time history inference using privileged training context, and A-RMA extends adaptation to bipedal control (Kumar et al., 2021; 2022). Temporal Student imitation with latent and action supervision on Student-visited states is also established in locomotion (Lee et al., 2020). Concurrent estimator–policy learning and DreamWaQ infer useful state or implicit context; DreamFLEX conditions fault-aware control on history including previous actions (Ji et al., 2022; Nahrendra et al., 2023; Lee et al., 2025). Our focus is their integration and controlled evaluation within a heterogeneous morphology interface.

**Cross-platform temporal control.** GenLoco uses observation–action histories for quadruped generalization (Feng et al., 2023). MetaLoco uses recurrent context and tests abrupt within-episode mass changes on unseen quadrupeds (Zargarbashi et al., 2024). MorAL and PAL study history-based adaptation across quadrupedal platforms (Luo et al., 2024; Rytz et al., 2025). Shared Modular Recurrence supplies per-limb observations, previous actions, and recurrent state under contextual conditions sampled at episode start (Engwegen et al., 2025). LocoFormer uses long state–action context and cross-episode memory (Liu et al., 2025); rapid embodiment adaptation also evaluates abrupt changes on nominal quadrupedal platforms (Li et al., 2026). MorphAdapt uses a reset-local finite history on heterogeneous UNIMAL bodies. We instead focus on integrating these ideas with heterogeneous shared control and evaluating them under matched and intervention-based protocols.

# 3 Two-Timescale Context Formulation

Let $m$ describe nominal topology and physical attributes, $o_t$ the current observation, and $d_t$ the unobserved runtime dynamics in $s_{t+1} \sim p_m(\cdot \mid s_t, a_t, d_t)$. Embodiment varies across robots and is fixed within an episode; $d_t$ can change during that episode. We assign these information roles to

$$\phi_t = f_m(m), \qquad \hat{z}_t = f_h(h_t), \qquad a_t = \pi(\phi_t, \hat{z}_t, c_t), \tag{1}$$

where $\phi_t$ denotes static nominal morphology context, $\hat{z}_t$ is inferred temporal context, and $c_t$ collects current control information, including $o_t$. The subscript on $\phi_t$ indicates its availability at each decision, not a changing embodiment. There is no independent task or velocity command in the reported experiments.

Equation 1 distinguishes the roles of nominal embodiment information and runtime interaction information without requiring statistical independence. In the implementation, nominal attributes and current limb state jointly form the shared Transformer tokens, giving state-dependent representations $u_t = F(m, o_t)$. The history encoder produces temporal context $\hat{z}_t$, which is fused with limb representations after attention for the shared decoder (Sec. 4). The temporal pathway can infer current response properties while the controller retains explicit access to the robot's nominal structure.

# 4 Method

## 4.1 Morphology-conditioned shared controller

The padded interface contains at most 12 limbs with 52 features each. Valid tokens include current proprioception and nominal morphology/hardware attributes. Each token is embedded into 128 dimensions with learned positional embeddings and processed by five Transformer layers with two attention heads and feed-forward width 1024 (Vaswani et al., 2017). Padding masks distinguish absent limbs. The 16-dimensional adaptation latent is broadcast to each limb and concatenated *after* attention, producing a 144-dimensional decoder input. Each limb produces two action slots, yielding a 24-dimensional padded action vector. Policy and value branches have separate parameters. This shared architecture permits different bodies to use a common control interface without identifying morphology from action history.

## 4.2 Action-conditioned temporal context

The Student re-encodes a fixed sliding window of length $H = 10$ at each decision. With $L = H - 1$, its causal history is

$$h_t = \{(o_{t-L}, a^{\text{can}}_{t-L-1}), \ldots, (o_t, a^{\text{can}}_{t-1})\}, \qquad \hat{z}_t = f_h(h_t). \tag{2}$$

The latest pair is $(o_t, a^{\text{can}}_{t-1})$, never $(o_t, a_t)$. Reset fills the window with repeated $(o_0, 0)$ pairs. History does not cross episodes, and no persistent RNN hidden state is passed between decisions. Canonical actions are clipped in policy-side padded ordering, with invalid slots zeroed. They are not multiplied by motor strength or nominal gear and are not torques or generalized forces.

Except in input ablations, MorphAdapt denotes the complete system using state and previous-action history; in input ablations, SA denotes this configuration and SO removes explicit action history. The per-step input has width 648 for State+Action (SA) and 624 for State-only (SO). A linear projection

to 32 dimensions and ReLU precede two 32-channel temporal convolutions (kernel three, padding one), each followed by ReLU. Temporal average pooling and a $32 \to 16$ linear layer with output ReLU produce the latent. SO omits only the action block. Both variants use the same morphology pathway, temporal backbone, Teacher, latent width, history length, optimizer, budgets, and split. The wider SA input adds 768 trainable policy-encoder weights; including the frozen value encoder, it adds 1536 stored weights. Thus, the comparison is closely architecture-matched but not exactly parameter-matched.

### 4.3 Privileged Teacher and interactive distillation

The Teacher encodes privileged dynamics through

$$z_t = f_T(d_t), \qquad d_t = [\lambda_{\text{motor}}, \lambda_{\text{friction}}, \lambda_{\text{mass}}, 0, \ldots, 0] \in \mathbb{R}^{16}. \tag{3}$$

A $16 \to 64 \to 64 \to 16$ MLP applies ReLU after every layer. Its latent is a learned privileged reference, not a ground-truth physical state or a mathematical performance upper bound. The Teacher is trained with PPO (Schulman et al., 2017); Student construction copies same-name, same-shape non-adaptation parameters, including all latent-channel decoder columns, and independently initializes its history encoders (Fig. 5).

Following privileged locomotion imitation (Lee et al., 2020), the current Student collects fresh interactive rollouts and the frozen Teacher labels the same visited states. For minibatch size $b$, the distillation objective is

$$\mathcal{L} = \frac{1}{16b} \sum_{i=1}^{b} \|\hat{z}_i - z_i\|_2^2 + \frac{0.1}{24b} \sum_{i=1}^{b} \|\mu_{S,i} - \mu_{T,i}\|_2^2. \tag{4}$$

Only the Student policy history encoder updates. The shared controller and value history encoder remain frozen; action-loss gradients still pass through the frozen decoder. Action supervision averages over all padded policy-mean coordinates without an active-action mask. Canonical actions stored in history and unmasked means used for supervision are distinct quantities. Stage-II uses no PPO, value, entropy, or explicit physical-parameter regression objective.

### 4.4 Deployment

The Student produces actions from nominal attributes, current observations, masks, and causal history. It requires neither privileged dynamics nor Teacher targets, and performs no gradient updates. Adaptation changes the inferred context, not the network parameters. Controller observations use running-statistics normalization, whereas temporal history uses raw padded features; this distinction is retained in both SO and SA.

## 5 Experimental Setup

**Morphologies and training.** We use UNIMAL morphology assets (Gupta et al., 2021) in MuJoCo (Todorov et al., 2012). The training set contains 100 robot instances. Strict-OOD97 contains 97 held-out walker IDs grouped into 87 exact-XML clusters, with no exact train/test XML overlap. Exact-XML identity does not imply that all test topologies are absent from training. Morphology names, clusters, training seeds, evaluation seeds, and rollouts are separate statistical units.

Student training uses reset dynamics randomization (Peng et al., 2018): independent motor and mass/inertia multipliers in $[0.8, 1.2]$ and friction in $[0.7, 1.3]$, fixed within each training episode. SO and SA each use training seeds 1409, 1410, and 1411. Their nominal 100M budget comprises 99,942,400 recorded environment transitions and 78,080 optimizer updates. Adam uses learning rate $3 \times 10^{-4}$ and eight minibatch epochs per rollout. Appendix B gives the remaining settings.

**Baselines and estimand.** MetaMorph-DR is the no-online-temporal-adaptation system baseline: a dynamics-randomized morphology-conditioned shared controller without deployment-time temporal context inference, evaluated at training seeds 1412, 1413, and 1415. This comparison measures the complete training and deployment pipelines. Within the MorphAdapt pipeline, SO retains state history, and SO versus SA tests the incremental contribution of explicit action input. SA–SO is

training-seed matched; SA–DR is not. A released-code ModuMorph reproduction uses the reported 100M configuration and one training seed (1409), so its values are descriptive. The Teacher supplies privileged supervision; a protocol-matched eight-condition Strict-OOD97 Teacher comparison is not available, and no numerical upper bound is inferred from older populations. HyperDistill uses an adapted MetaMorph-DR Teacher configuration, documented in Appendix H.

**Fixed and changing dynamics.** The benchmark crosses Fixed and Mutation settings with four *reset* conditions: nominal, ID, OOD-mild, and OOD-strong. Fixed retains reset parameters. Mutation independently replaces them before the physics advance of one-based transition 250, using the same post-change distributions in every reset condition:

$$\lambda^{\text{post}}_{\text{motor}}, \lambda^{\text{post}}_{\text{mass}} \sim U[0.6, 1.4], \qquad \lambda^{\text{post}}_{\text{friction}} \sim U[0.4, 1.6]. \tag{5}$$

The mass multiplier also scales inertia. Changes persist to termination; asymmetric actuator faults are disabled. Thus, Mutation OOD-strong describes the initial condition, not a distinct stronger post-change distribution. Control is 50 Hz with a 1000-step horizon. Appendix C specifies reset intervals and indexing.

**Metrics and statistics.** Each trained model contributes one deterministic episode per walker and condition using evaluation seed 1409. Episodic return is the primary metric. Within each training seed, eligible walker values are averaged; the reported mean and sample SD summarize three seed means. Point estimates retain walker weighting. Existing 95% intervals use 5000 shared block-bootstrap draws over the 87 XML clusters, preserving instance weighting within each seed. They quantify morphology-level uncertainty conditional on the observed models, not training-seed population uncertainty; seeds are not resampled and intervals are unadjusted for multiple comparisons.

Mutation summaries include maximum normalized drop (MND) in forward velocity. Recovery time, sustained recovery rate, and normalized post-mutation performance retain their own eligibility rules (Appendix D). Termination is a proxy outcome, not a classified fall. The repeated-change experiment uses a separate equal-cluster estimand and its specified continuation rule; its results are not pooled with the main table.

# 6 Results

## 6.1 Joint Unseen-Morphology × Altered-Dynamics Performance

On 97 held-out morphology IDs in 87 exact-XML clusters, MorphAdapt achieves higher mean episodic return than MetaMorph-DR in all eight fixed and mutation conditions (Fig. 2; full values in Appendix A, Table 2). Across the full test matrix, gains range from 6.8% to 16.4%, with gains of 9.3%–16.4% across the four mutation conditions and the largest relative gain in Mutation OOD-strong. Return-difference intervals exclude zero only for Fixed-ID, $[21.6, 352.6]$, and Mutation OOD-strong, $[26.8, 333.2]$; the other intervals include zero. These aggregate differences do not imply improvement on every morphology.

MorphAdapt also exceeds the evaluated ModuMorph reproduction in mean return in all eight conditions and the HyperDistill adaptation in all four mutation conditions (Fig. 2(b), Table 8). These comparisons evaluate the complete temporal-adaptation system, including its training procedure.

## 6.2 Temporal Adaptation Under Dynamics Changes

Figure 3 complements return with maximum post-mutation velocity drop, relative post-mutation locomotion, and horizon completion. MorphAdapt has lower maximum normalized post-mutation velocity drop (MND) than MetaMorph-DR, ModuMorph, and HyperDistill in all four mutation conditions (Fig. 3(a)); its MND-difference intervals relative to MetaMorph-DR exclude zero in nominal and OOD-mild. NPP measures mean post-mutation velocity relative to each method's own pre-mutation baseline; Survival is the fraction of episodes reaching the evaluation horizon. NPP and Survival point estimates each favor MorphAdapt over MetaMorph-DR in three conditions and are lower in ID (Fig. 3(b,c)). Survival differences are $+1.7/-1.0/+2.4/+3.1$ percentage points; all NPP and Survival difference intervals include zero.

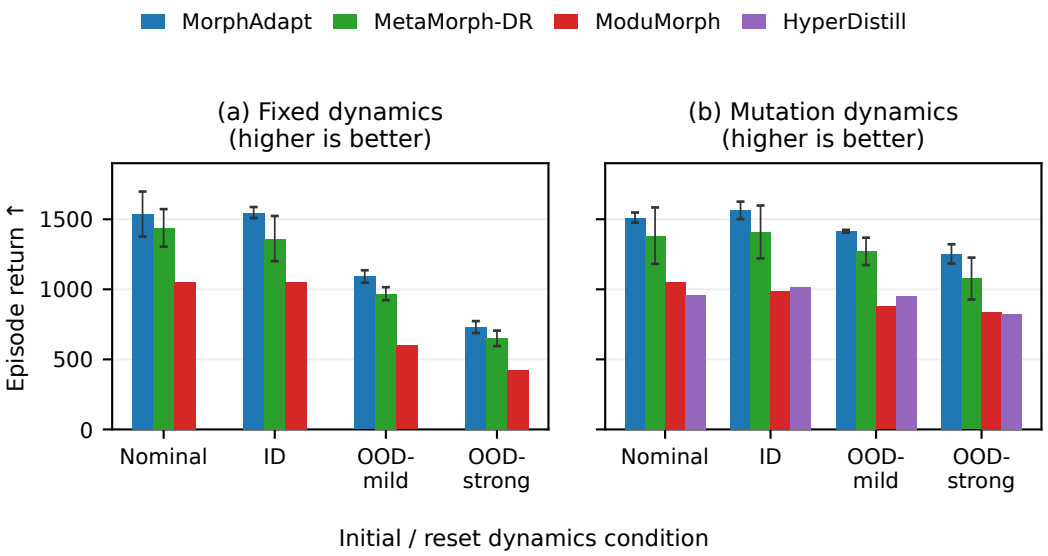


Figure 2: **System-level performance across fixed and changing dynamics.** (a) Fixed dynamics. (b) Runtime mutation. Higher episode return is better. MorphAdapt and MetaMorph-DR show means and sample SD over three training seeds; ModuMorph and HyperDistill are single-training-seed external baselines. HyperDistill is shown only for mutation cells under the common protocol. Full values are in Tables 2 and 8.

RT and SRR provide complementary information: recovered-only RT is lower for MorphAdapt than MetaMorph-DR in all four conditions, with all difference intervals including zero, while SRR has mixed directions. RT averages only recovered episodes, and SRR uses mutation-exposed episodes; lower RT does not establish faster recovery over all exposed episodes. Appendix D reports the full summaries. A separate fallen-state test provides a complementary, restricted sensitivity result: MorphAdapt shows higher relaxed recovery rates on a post-hoc moderate-state subset, while strict resumed locomotion remains zero for both methods (Appendix G).

### 6.3 What Does Previous-Action Conditioning Add?

**Matched action-input ablation.** The matched $H = 10$ ablation shows changes in the direction of SA–SO return-difference point estimates across conditions (Fig. 4(a); full values in Table 3, Appendix A). SA has higher mean return in Fixed-ID and Mutation OOD-mild; SO has higher means in the other six conditions. Every return-difference interval includes zero. Under mutation, SA lowers MND relative to SO in nominal and OOD-mild, with the opposite direction in ID and OOD-strong. This comparison did not show a consistent benefit from explicit action input.

The independent State+Action horizon-sensitivity sweep in Fig. 4(b) compares $H \in \{1, 2, 5, 10\}$ at 10M with GAP readout. Shorter histories remain competitive, with $H = 2$ showing the largest return increments under OOD-strong relative to $H = 10$ within the same sweep.

**Earlier development comparison.** In the earlier matched nominal-100M SO/SA development comparison, Gate7, SA has higher mutation Return point estimates in all four protocols. In OOD-mild and OOD-strong, its restricted mean recovery time (RMRT) is lower by 53.96 and 42.75 control steps, respectively (Table 1). Both RMRT-difference intervals cross zero, as do all four Return-difference intervals.

The earlier Gate7 comparison produced larger post-perturbation point estimates under its separate protocol, whereas the main matched ablation shows mixed directions (Appendix I).

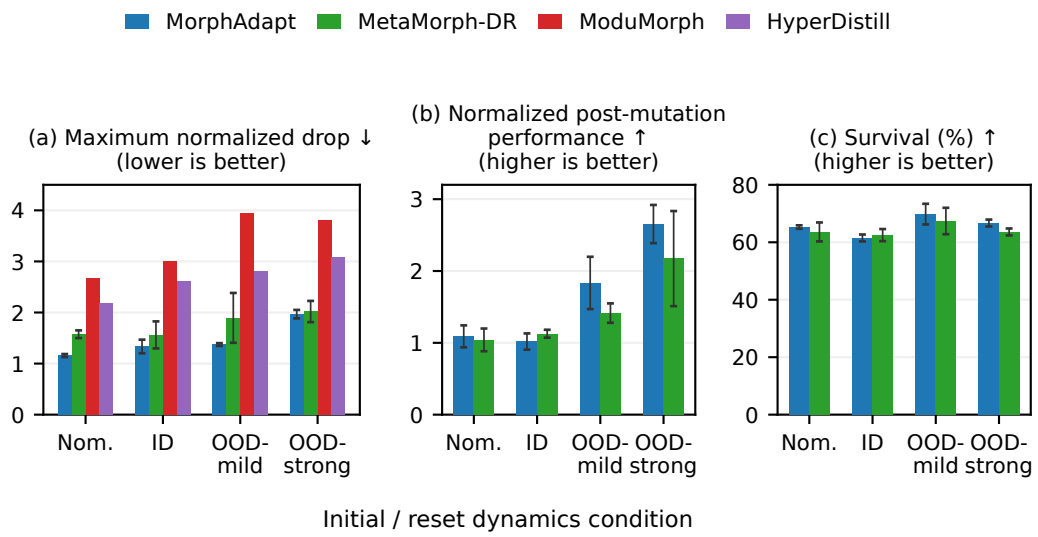


Figure 3: **Post-mutation robustness and post-shift performance quality.** (a) Maximum normalized post-mutation velocity drop (MND; lower is better) compares MorphAdapt with all three system baselines. (b) Normalized post-mutation performance (NPP) and (c) survival compare MorphAdapt with MetaMorph-DR; higher is better. NPP measures mean post-mutation velocity relative to each method's own pre-mutation baseline. Survival is the percentage of episodes reaching the evaluation horizon. MorphAdapt/MetaMorph-DR bars show means and sample SD; external baselines show point estimates (Tables 6, 8).

Table 1: Gate7 development case: paired RMRT differences (lower is better). One training realization, nominal-100M configuration, legacy OOD98. Both 95% intervals cross zero; these are point effects, not replicated or significant improvements. The horizon and population differ from the main benchmark.

| Reset condition | Paired $N$ | SA−SO RMRT (steps) | 95% paired CI |
|---|---|---|---|
| OOD-mild | 75 | $-53.96$ | $[-115.75, 8.67]$ |
| OOD-strong | 73 | $-42.75$ | $[-104.11, 16.51]$ |

### 6.4 Repeated Dynamics Transitions

A separate mechanism diagnostic changes motor strength from $A = 1.0$ to $B = 0.80$ at index 250 and back to $A$ at 550. UPDATE supplies the live temporal latent; FREEZE holds the controller-facing latent captured before the second change while history, encoder computation, and the environment continue. After excluding discovery XML clusters, confirmation contains 91 IDs/81 clusters, with training and evaluation seed 1409. The equal-cluster UPDATE-minus-FREEZE reward contrast $D_{\mathrm{RETURN}}$ over $[550, 800)$ has 23 positive, 11 negative, and 47 exact-zero values, mean $+0.215574$, and median zero.

Among 35 exposure-qualified clusters, 23 (65.7%) have positive effects, 11 negative, and one zero; each morphology family has a positive sign majority. This positive majority is conditional on exposure and does not characterize the full OOD population; the exposed floor-family mean remains negative. A matched HOLD branch stays at $B$ at 550. Its interaction $I_{\mathrm{RETURN}} = D_{\mathrm{RETURN}} - D_{\mathrm{HOLD}}$ has mean $-2.828963$ and 18 positive, 17 negative, and 46 zero values. The full-population interaction does not establish a positive return-transition-specific advantage (Appendix E, Fig. 6).

### 6.5 Structural Distance and Context-Update Effects

The main benchmark tests exact-XML-held-out embodiments. As a secondary diagnostic, we measure each confirmation morphology's distance to the nearest training morphology using nine structural

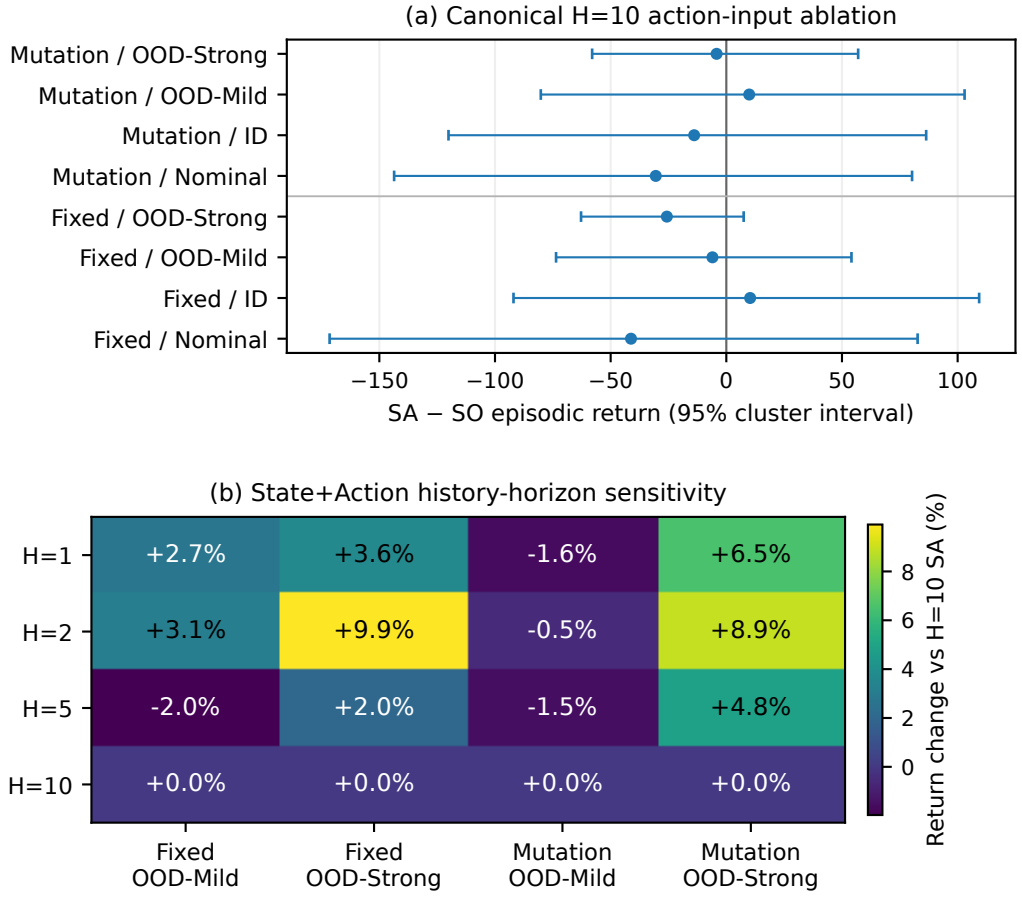


Figure 4: **Action-input ablation and State+Action horizon sensitivity.** (a) Main $H = 10$ matched SA−SO return differences and 95% XML-cluster bootstrap intervals. (b) Return changes relative to $H = 10$ within the same 10M State+Action sweep on Strict-OOD97 with GAP readout. At $H = 2$, Mutation OOD-strong sustained recovery also increases by 7.9 percentage points. Protocol details are in Appendix A.

descriptors, training-only standardization, and group-balanced RMS. Across all 81 confirmation clusters, Spearman correlation between this distance and the context-update effect $D_{\mathrm{RETURN}}$ is $+0.248079$: no systematic attenuation of update benefit is observed with increasing descriptor distance. This is neither a farther-is-better result nor evidence of morphology invariance. The HOLD-relative interaction correlation is $-0.012667$; within-mvt correlation is undefined because distance is constant (Appendix E).

## 7 Discussion

### 7.1 Seed and Optimization Sensitivity of Action Conditioning

The mutation Return comparison favors SA in 2/4, 3/4, and 2/4 cells for training seeds 1409, 1410, and 1411, respectively. Every seed is retained. Together with the heterogeneous aggregate ablation, this shows that the incremental effect varies across trained realizations and conditions. Gate7 illustrates larger possible point gains, but its different implementation lineage, population, and horizon prevent attributing the cross-campaign difference solely to random seed or optimization.

Three hypotheses could explain why explicit actions are not exploited consistently. State history may already encode substantial actuator-response information. Different optimization trajectories

may make different use of the additional action channel. Privileged latent and action distillation may also permit solutions that do not consistently rely on that channel. The current experiments do not distinguish these hypotheses or establish optimization as the cause. They motivate reliable learning from action history as a specific open problem, rather than a conclusion that action history is universally beneficial or ineffective.

### 7.2 Behavioral Utility and Privileged-Latent Agreement

An observer-only diagnostic compares Student latents with the frozen Teacher's post-ReLU reference without changing rollout actions. The diagnostic reports consistently lower SA Teacher-L2 in mvt, but mixed behavior in floor and vt. This auxiliary finding concerns a learned reference, whose sparse output constrains the interpretation of Euclidean distance. It neither replaces behavioral evaluation nor makes perfect Teacher matching a prerequisite for control utility (Appendix F).

## 8 Limitations

Evaluation is simulation-only and confined to the UNIMAL morphology distribution and the tested motor, friction, and mass/inertia perturbations. No real-robot transfer or sim-to-real robustness is established. Three training seeds support the main benchmark, while the ABA confirmation uses only one training and one evaluation seed. Its exposure-qualified result is conditional, exact positive signs need not represent practically large effects, and structural distance covers selected descriptors rather than all morphology differences.

The flattened temporal interface retains canonical slot-order sensitivity, and full-padded action supervision can weight morphologies differently. Action-conditioning effects vary across the evaluated training realizations. Gate7 is a single development case with a different protocol and cannot establish cross-seed robustness. The timescale formulation is an information-role distinction, not a learned-independence guarantee; latent updating and explicit physical-parameter identification remain separate questions. A qualified main-population Teacher performance comparison and complete historical training provenance remain unavailable.

## 9 Conclusion

MorphAdapt combines nominal embodiment information and recent interaction context in a shared controller, adapting through context inference without privileged information or network-parameter updates at deployment. Under dynamics mutation on held-out UNIMAL morphologies, MorphAdapt achieves 9.3%–16.4% higher mean return than MetaMorph-DR and smaller maximum normalized post-mutation velocity drops than the compared systems. The input ablation shows that these system-level gains do not translate into a uniform benefit from explicit action history. Together, the results support combining morphology-conditioned shared control with online interaction context for changing-dynamics control on unseen embodiments.

## REPRODUCIBILITY STATEMENT

The information interface and objective are specified in Section 4. Section 5 and Appendices B–I describe the training settings, evaluation protocol, statistics, and auxiliary diagnostics.

## AI USE STATEMENT

Generative AI tools were used to assist with research planning and experimental design, method implementation and code debugging, analysis and interpretation of experimental results, literature search and organization, manuscript structuring, drafting and editing, translation between Chinese and English, and code for tables, figures, compilation, and consistency checks. All AI-assisted research decisions, code, analyses, and manuscript content were reviewed by the author, who takes responsibility for the final submission. Training runs, simulator rollouts, and numerical results were produced by the research code and verified from experimental outputs rather than synthesized by an AI system.

## ETHICS STATEMENT

The reported evaluation is in simulation. It does not establish the safety of deploying these controllers on physical robots. Real-world use would require hardware-specific validation and safety measures.

## REFERENCES

Nico Bohlinger, Grzegorz Czechmanowski, Maciej Piotr Krupka, Piotr Kicki, Krzysztof Walas, Jan Peters, and Davide Tateo. One Policy to Run Them All: an End-to-end Learning Approach to Multi-Embodiment Locomotion. In *Proceedings of The 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pp. 3356–3378. PMLR, 2025. URL https://proceedings.mlr.press/v270/bohlinger25a.html.

Laurens Engwegen, Max Weltevrede, Caroline Horsch, Daan Brinks, and Wendelin Böhmer. Shared modular recurrence in contextual MDPs for universal morphology control. *arXiv preprint arXiv:2506.08630*, 2025. URL https://arxiv.org/abs/2506.08630.

Gilbert Feng, Hongbo Zhang, Zhongyu Li, Xue Bin Peng, Bhuvan Basireddy, Linzhu Yue, Zhitao Song, Lizhi Yang, Yunhui Liu, Koushil Sreenath, and Sergey Levine. GenLoco: Generalized Locomotion Controllers for Quadrupedal Robots. In *Proceedings of The 6th Conference on Robot Learning*, volume 205 of *Proceedings of Machine Learning Research*, pp. 1893–1903. PMLR, 2023. URL https://proceedings.mlr.press/v205/feng23a.html.

Agrim Gupta, Silvio Savarese, Surya Ganguli, and Li Fei-Fei. Embodied intelligence via learning and evolution. *Nature Communications*, 12:5721, 2021. doi: 10.1038/s41467-021-25874-z. URL https://www.nature.com/articles/s41467-021-25874-z.

Agrim Gupta, Linxi Fan, Surya Ganguli, and Li Fei-Fei. MetaMorph: Learning universal controllers with transformers. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=Opmqtk_GvYL.

Wenlong Huang, Igor Mordatch, and Deepak Pathak. One policy to control them all: Shared modular policies for agent-agnostic control. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 4455–4464. PMLR, 2020. URL https://proceedings.mlr.press/v119/huang20d.html.

Gwanghyeon Ji, Juhyeok Mun, Hyeongjun Kim, and Jemin Hwangbo. Concurrent Training of a Control Policy and a State Estimator for Dynamic and Robust Legged Locomotion. *IEEE Robotics and Automation Letters*, 7(2):4630–4637, 2022. doi: 10.1109/LRA.2022.3151396. URL https://arxiv.org/abs/2202.05481.

Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. RMA: Rapid motor adaptation for legged robots. In *Proceedings of Robotics: Science and Systems*, 2021. doi: 10.15607/RSS.2021.XVII.011. URL https://www.roboticsproceedings.org/rss17/p011.html.

Ashish Kumar, Zhongyu Li, Jun Zeng, Deepak Pathak, Koushil Sreenath, and Jitendra Malik. Adapting Rapid Motor Adaptation for Bipedal Robots. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1161–1168. IEEE, 2022. doi: 10.1109/IROS47612.2022.9981091. URL https://ieeexplore.ieee.org/document/9981091/.

Vitaly Kurin, Maximilian Igl, Tim Rocktäschel, Wendelin Boehmer, and Shimon Whiteson. My Body is a Cage: the Role of Morphology in Graph-Based Incompatible Control. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.01856.

Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. Learning quadrupedal locomotion over challenging terrain. *Science Robotics*, 5(47):eabc5986, 2020. doi: 10.1126/scirobotics.abc5986. URL https://www.science.org/doi/10.1126/scirobotics.abc5986.

Seunghyun Lee, I Made Aswin Nahrendra, Dongkyu Lee, Byeongho Yu, Minho Oh, Hyeonwoo Lee, and Hyun Myung. DreamFLEX: Learning fault-aware quadrupedal locomotion controller for anomaly situation in rough terrains. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 16001–16007, 2025. doi: 10.1109/ICRA55743.2025.11127805.

Dichen Li, Bo Ai, Nico Bohlinger, Jan Peters, Hao Su, and Henrik I. Christensen. Rapid embodiment adaptation for quadrupedal locomotion, 2026. URL https://arxiv.org/abs/2608.01506. arXiv preprint arXiv:2608.01506.

Min Liu, Deepak Pathak, and Ananye Agarwal. LocoFormer: Generalist locomotion via long-context adaptation. In *Proceedings of The 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pp. 532–546. PMLR, 2025. URL https://proceedings.mlr.press/v305/liu25a.html.

Zeren Luo, Yinzhao Dong, Xinqi Li, Rui Huang, Zhengjie Shu, Erdong Xiao, and Peng Lu. MorAL: Learning Morphologically Adaptive Locomotion Controller for Quadrupedal Robots on Challenging Terrains. *IEEE Robotics and Automation Letters*, 9(5):4019–4026, 2024. doi: 10.1109/LRA.2024.3375086. URL https://ieeexplore.ieee.org/document/10463132/.

I Made Aswin Nahrendra, Byeongho Yu, and Hyun Myung. DreamWaQ: Learning Robust Quadrupedal Locomotion With Implicit Terrain Imagination via Deep Reinforcement Learning. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5078–5084. IEEE, 2023. doi: 10.1109/ICRA48891.2023.10161144. URL https://ieeexplore.ieee.org/document/10161144/.

Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3803–3810. IEEE, 2018. doi: 10.1109/ICRA.2018.8460528. URL https://xbpeng.github.io/projects/SimToReal/index.html.

David Rytz, Suyoung Choi, Wanming Yu, Wolfgang Merkt, Jemin Hwangbo, and Ioannis Havoutis. Reference Free Platform Adaptive Locomotion for Quadrupedal Robots using a Dynamics Conditioned Policy. In *2025 European Conference on Mobile Robots (ECMR)*. IEEE, 2025. doi: 10.1109/ECMR65884.2025.11162971. URL https://ieeexplore.ieee.org/document/11162971/.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. arXiv preprint arXiv:1707.06347, 2017. URL https://arxiv.org/abs/1707.06347.

Carmelo Sferrazza, Dun-Ming Huang, Fangchen Liu, Jongmin Lee, and Pieter Abbeel. Body Transformer: Leveraging Robot Embodiment for Policy Learning. In *Proceedings of The 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pp. 3407–3424. PMLR, 2025. URL https://proceedings.mlr.press/v270/sferrazza25a.html.

Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033. IEEE, 2012. doi: 10.1109/IROS.2012.6386109. URL https://ieeexplore.ieee.org/document/6386109/.

Brandon Trabucco, Mariano Phielipp, and Glen Berseth. AnyMorph: Learning Transferable Polices By Inferring Agent Morphology. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 21677–21691. PMLR, 2022. URL https://proceedings.mlr.press/v162/trabucco22b.html.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Tingwu Wang, Renjie Liao, Jimmy Ba, and Sanja Fidler. NerveNet: Learning Structured Policy with Graph Neural Networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=S1sqHMZCb.

Zheng Xiong, Jacob Beck, and Shimon Whiteson. Universal morphology control via contextual modulation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 38286–38300. PMLR, 2023. URL https://proceedings.mlr.press/v202/xiong23a.html.

Zheng Xiong, Risto Vuorio, Jacob Beck, Matthieu Zimmer, Kun Shao, and Shimon Whiteson. Distilling morphology-conditioned hypernetworks for efficient universal morphology control. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 54777–54791. PMLR, 2024. URL https://proceedings.mlr.press/v235/xiong24c.html.

Wenhao Yu, Jie Tan, C. Karen Liu, and Greg Turk. Preparing for the Unknown: Learning a Universal Policy with Online System Identification. In *Proceedings of Robotics: Science and Systems*, 2017. doi: 10.15607/RSS.2017.XIII.048. URL https://www.roboticsproceedings.org/rss13/p48.html.

Fatemeh Zargarbashi, Fabrizio Di Giuro, Jin Cheng, Dongho Kang, Bhavya Sukhija, and Stelian Coros. MetaLoco: Universal Quadrupedal Locomotion with Meta-Reinforcement Learning and Motion Imitation. arXiv preprint arXiv:2407.17502, 2024. URL https://arxiv.org/abs/2407.17502.

Table 2: Strict-OOD97 episodic return. SA and DR: mean ± sample SD over three training seeds; SA–DR seeds are unpaired. †ModuMorph: one training seed, released-code reported-100M reproduction. Gain is the relative SA–DR mean difference. 97 IDs, 87 exact-XML clusters; one evaluation seed and one episode per ID/condition.

| Setting | Initial condition | SA Return | DR Return | ModuMorph† | Gain |
|---|---|---|---|---|---|
| Fixed | Nominal | 1537.0 ± 160.8 | 1438.6 ± 133.9 | 1049.7 | 6.8% |
| Fixed | ID | 1547.8 ± 39.4 | 1362.3 ± 161.2 | 1053.0 | 13.6% |
| Fixed | OOD-mild | 1091.7 ± 44.4 | 968.8 ± 46.4 | 599.3 | 12.7% |
| Fixed | OOD-strong | 730.6 ± 42.6 | 649.9 ± 55.5 | 424.4 | 12.4% |
| Mutation | Nominal | 1511.8 ± 36.2 | 1383.3 ± 201.0 | 1054.1 | 9.3% |
| Mutation | ID | 1563.6 ± 62.4 | 1409.5 ± 188.6 | 986.1 | 10.9% |
| Mutation | OOD-mild | 1414.6 ± 9.8 | 1271.0 ± 97.7 | 877.6 | 11.3% |
| Mutation | OOD-strong | 1253.1 ± 68.5 | 1076.9 ± 149.3 | 838.2 | 16.4% |

Table 3: Matched history-input ablation. SO returns use the same three training seeds as SA. Intervals are the 95% XML-block bootstrap intervals, conditional on observed trained models; all return-difference intervals include zero. MND values are mutation point estimates (lower is better).

| Setting | Initial condition | SO Return | SA−SO [95% CI] | MND SA / SO |
|---|---|---|---|---|
| Fixed | Nominal | 1578.3 ± 57.3 | -41.2 [-171.5, 82.7] | – |
| Fixed | ID | 1537.5 ± 86.9 | +10.3 [-92.0, 109.3] | – |
| Fixed | OOD-mild | 1097.6 ± 47.5 | -6.0 [-73.6, 54.1] | – |
| Fixed | OOD-strong | 756.3 ± 25.0 | -25.7 [-62.8, 7.5] | – |
| Mutation | Nominal | 1542.3 ± 27.2 | -30.5 [-143.6, 80.3] | 1.16 / 1.29 |
| Mutation | ID | 1577.5 ± 39.5 | -13.9 [-120.1, 86.4] | 1.34 / 1.28 |
| Mutation | OOD-mild | 1404.7 ± 11.4 | +9.9 [-80.2, 103.0] | 1.37 / 1.60 |
| Mutation | OOD-strong | 1257.3 ± 35.5 | -4.2 [-58.0, 57.0] | 1.97 / 1.74 |

# A Complete benchmark and action-input results

Tables 2 and 3 retain the full numerical summaries underlying Figs. 2 and 4(a), respectively. The main $H = 10$ ablation retains all three training seeds, and its bootstrap intervals are conditional on the observed trained models. The State+Action horizon sweep in Fig. 4(b) uses 10M, training seed 1409, evaluation seed 1409, and GAP readout on Strict-OOD97. It is a single-seed sensitivity diagnostic, not a matched SA–SO sweep across horizons. System comparisons include training-pipeline differences and do not isolate history, previous actions, or distillation as the sole causal source of the gain.

Table 4: Architecture and Student training settings specified by the experiment configurations and implementation.

| Component | Frozen setting |
|---|---|
| Limb interface | 12 limbs $\times$ 52 features; 24 padded action slots |
| Shared controller | 128-wide tokens; 5 layers; 2 heads; feed-forward 1024; dropout 0 |
| Temporal encoder | $H = 10$; SO/SA input 624/648; projection 32; two Conv1d layers; latent 16, post-ReLU |
| Teacher encoder | $16 \to 64 \to 64 \to 16$; ReLU after each linear layer |
| Fusion | 128-dimensional limb output plus 16-dimensional latent; decoder width 144 |
| Student optimizer | Adam; LR $3 \times 10^{-4}$; epsilon $10^{-5}$; gradient norm 0.5 |
| Student budget | 32 environments; 2560 steps; minibatch 10240; 8 epochs; 1220 iterations |
| Student completion | 99,942,400 transitions; 78,080 updates; only policy history encoder trainable |
| Policy distribution | Gaussian; fixed standard deviation 0.9 during interactive collection |

## B Implementation and training details

The shared control interface uses depth-first XML limb ordering and canonical mirror mapping. Masks are applied in controller attention and valid-limb value aggregation. The temporal encoder receives flattened padded features without a separate limb-mask input. Its history is reconstructed after every reset, preventing observations or actions from the previous episode from entering a new episode's context.

Student behavior samples a diagonal Gaussian policy during interactive collection; the configured standard deviation is fixed at 0.9. History stores the clipped, invalid-slot-zeroed policy-side command corresponding to execution, while the rollout buffer can retain the raw policy sample. Deterministic evaluation uses the policy mean through the same execution interface. Nominal gear and dynamics scaling are downstream of the history-action representation. The action loss in Eq. 4 compares unmasked means, so padded mean coordinates can contribute gradients even though invalid executed history slots are zero.

Each Student iteration collects $32 \times 2560 = 81920$ transitions. Eight epochs with minibatch size 10240 yield 64 optimizer updates per iteration. Across 1220 iterations, the completion records report 99,942,400 transitions and 78,080 updates. Adam uses epsilon $10^{-5}$, cosine learning-rate decay to zero, five warmup iterations with factor 0.1, and gradient-norm clipping at 0.5. Only the policy temporal encoder receives updates. These numbers describe Student Stage-II, not Teacher training.

The recovered Teacher configuration uses 100 training instances, the same reset dynamics ranges, PPO learning rate $3 \times 10^{-4}$, discount 0.99, GAE 0.95, clipping 0.2, minibatch size 5120, and eight epochs. Its configured budget is 100M state–action pairs. A complete Teacher completion record and historical Teacher-to-Student initialization byte receipts are not available; configured values must not be interpreted as measured Teacher completion counts. The policy/value Transformers and adaptation encoders are separate modules. Copying the non-adaptation controller includes all 144-dimensional decoder inputs, preserving the source-defined latent interface.

## C Evaluation identity and dynamics protocol

Exact-XML clusters group identical XML SHA256 values. Strict-OOD97 retains 97 IDs but only 87 distinct XML identities; no exact identity overlaps training. The split includes 64 floor, 11 mvt, and 22 vt IDs. A structural descriptor may be identical for distinct XML bodies, and topology overlap is compatible with exact-XML disjointness. The structural-distance analysis identifies eight test IDs with a topology seen in training and 89 without one.

For two-interval reset distributions, either interval is selected with equal probability, then a value is sampled uniformly inside it. The three dynamics multipliers are sampled independently. Parameters are restored to nominal before applying new multipliers to avoid cumulative scaling. Motor scaling

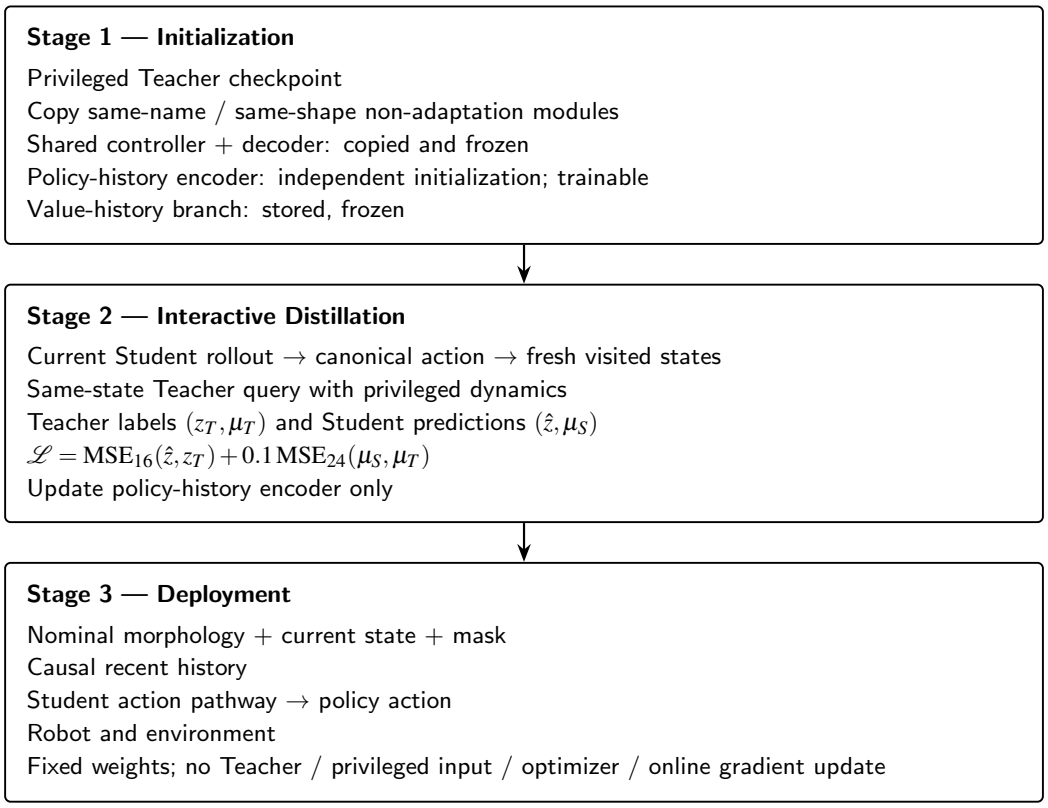


Figure 5: Initialization, interactive distillation, and deployment. The depicted initialization phase is distinct from privileged Teacher PPO pretraining. The value history encoder is stored but receives no Stage-II distillation supervision.

Table 5: Reset distributions and the shared persistent mutation target. All values multiply nominal parameters.

| Reset condition | Motor / mass–inertia | Friction |
|---|---|---|
| Nominal | $\{1\}$ | $\{1\}$ |
| ID | $[0.8, 1.2]$ | $[0.7, 1.3]$ |
| OOD-mild | $[0.7, 0.8] \cup [1.2, 1.3]$ | $[0.55, 0.7] \cup [1.3, 1.45]$ |
| OOD-strong | $[0.6, 0.7] \cup [1.3, 1.4]$ | $[0.4, 0.55] \cup [1.45, 1.6]$ |
| Mutation target (all rows) | $U[0.6, 1.4]$ | $U[0.4, 1.6]$ |

is global over actuators; friction affects ground/terrain geometry; body mass and inertia share one multiplier.

The simulator timestep is 0.005 s and control decimation is four, giving 0.02 s per policy step. In the main benchmark, the mutation occurs before the advance of one-based transition 250: zero-based decision 249 produces the action for the first affected transition, and decision 250 first observes its response. The transition spans 4.98–5.00 s. This convention is distinct from the serialized decision-index windows of the ABA experiment. The episode limit is 1000 steps. A torso-height proxy at or below half the initial standing height can also terminate the episode. Early termination is not a classification of fall cause.

Main-table training seeds are SO/SA 1409, 1410, 1411 and DR 1412, 1413, 1415. Evaluation seed 1409 is held common with one rollout per ID and condition. Additional evaluation seeds in other archives are not independent training runs and are not silently pooled into this table. ModuMorph

Table 6: Mutation metrics, mean ± training-seed sample SD. RT is in control steps and recovered-only; SRR is over exposed episodes; MND/NPP require eligible baselines and finite post samples. Survival is over all evaluated episodes. Metric-specific denominators are not interchangeable.

| Metric | Initial condition | SA | SO | DR |
|---|---|---|---|---|
| MND | Nominal | $1.157 \pm 0.030$ | $1.292 \pm 0.127$ | $1.575 \pm 0.075$ |
| MND | ID | $1.335 \pm 0.133$ | $1.276 \pm 0.074$ | $1.561 \pm 0.265$ |
| MND | OOD-mild | $1.373 \pm 0.027$ | $1.596 \pm 0.126$ | $1.895 \pm 0.488$ |
| MND | OOD-strong | $1.968 \pm 0.083$ | $1.740 \pm 0.193$ | $2.019 \pm 0.208$ |
| RT | Nominal | $121.4 \pm 28.4$ | $127.5 \pm 7.1$ | $126.6 \pm 17.4$ |
| RT | ID | $107.9 \pm 17.5$ | $113.7 \pm 27.3$ | $122.5 \pm 39.0$ |
| RT | OOD-mild | $71.8 \pm 14.9$ | $58.5 \pm 3.3$ | $88.1 \pm 27.1$ |
| RT | OOD-strong | $49.7 \pm 5.3$ | $45.5 \pm 3.0$ | $61.9 \pm 12.3$ |
| SRR | Nominal | $0.683 \pm 0.022$ | $0.708 \pm 0.016$ | $0.686 \pm 0.015$ |
| SRR | ID | $0.664 \pm 0.013$ | $0.672 \pm 0.018$ | $0.627 \pm 0.060$ |
| SRR | OOD-mild | $0.788 \pm 0.028$ | $0.824 \pm 0.018$ | $0.780 \pm 0.075$ |
| SRR | OOD-strong | $0.799 \pm 0.046$ | $0.811 \pm 0.030$ | $0.788 \pm 0.029$ |
| NPP | Nominal | $1.090 \pm 0.153$ | $1.005 \pm 0.122$ | $1.041 \pm 0.158$ |
| NPP | ID | $1.018 \pm 0.113$ | $1.128 \pm 0.173$ | $1.127 \pm 0.055$ |
| NPP | OOD-mild | $1.834 \pm 0.364$ | $1.762 \pm 0.234$ | $1.414 \pm 0.135$ |
| NPP | OOD-strong | $2.654 \pm 0.266$ | $2.582 \pm 0.052$ | $2.173 \pm 0.662$ |
| Survival | Nominal | $0.653 \pm 0.006$ | $0.632 \pm 0.016$ | $0.636 \pm 0.033$ |
| Survival | ID | $0.615 \pm 0.012$ | $0.649 \pm 0.010$ | $0.625 \pm 0.021$ |
| Survival | OOD-mild | $0.698 \pm 0.036$ | $0.670 \pm 0.021$ | $0.674 \pm 0.046$ |
| Survival | OOD-strong | $0.667 \pm 0.012$ | $0.656 \pm 0.036$ | $0.636 \pm 0.012$ |

contributes one training seed. ABA and the latent diagnostic each use training/evaluation seed 1409; the auxiliary fallen-state study uses its own three matched training seeds, 1410, 1411, and 1415.

## D METRIC DEFINITIONS AND ADDITIONAL MUTATION RESULTS

Let $B$ be mean forward velocity over the 25 pre-mutation samples. It is eligible only if finite and at least 0.05. For finite observed post-mutation samples $x_i$, including the first affected transition, define

$$\mathrm{MND} = \frac{\max(0, B - \min_i x_i)}{|B|}, \qquad \mathrm{NPP} = \frac{\mathrm{mean}_i\, x_i}{|B|}. \tag{6}$$

Survival is the fraction reaching the 1000-step horizon among all evaluated episodes; it does not classify a physical fall. These MND/NPP quantities require a valid baseline and at least one finite post sample. MND can exceed one under negative velocity, and NPP is not bounded by one. No post-termination zero-velocity samples are appended in the main benchmark.

Recovery time (RT) is the one-based post-change confirmation index of 25 consecutive finite samples at or above $0.8B$. Immediate sustained recovery is confirmed at RT=25; the onset index would instead be one. Missing samples break the run. RT averages recovered episodes only, with missing times excluded rather than set to zero. Sustained recovery rate (SRR) divides confirmed recoveries by all mutation-exposed episodes; invalid baselines or absence of a recovery count as non-recovery. Later sustained relapse is recorded separately and does not revoke a prior confirmation. Methods use their own eligible sets, not the intersection of their successful episodes.

All SA–DR RT, SRR, and NPP difference intervals include zero. The MND-difference intervals exclude zero only for nominal and OOD-mild. These metrics answer different questions, and recovered-only RT cannot support faster recovery over the entire exposed population. RMRT is a separate restricted-time quantity in the evaluator; it is not substituted for recovered-only RT here. Because velocity-command conditioning is disabled, $x_\mathrm{vel}$ describes forward locomotion, not velocity tracking.

The cluster bootstrap samples 87 XML blocks with replacement, shares draws across methods and training seeds, and retains each sampled block's full valid-instance sum and count. Seed-level ratios

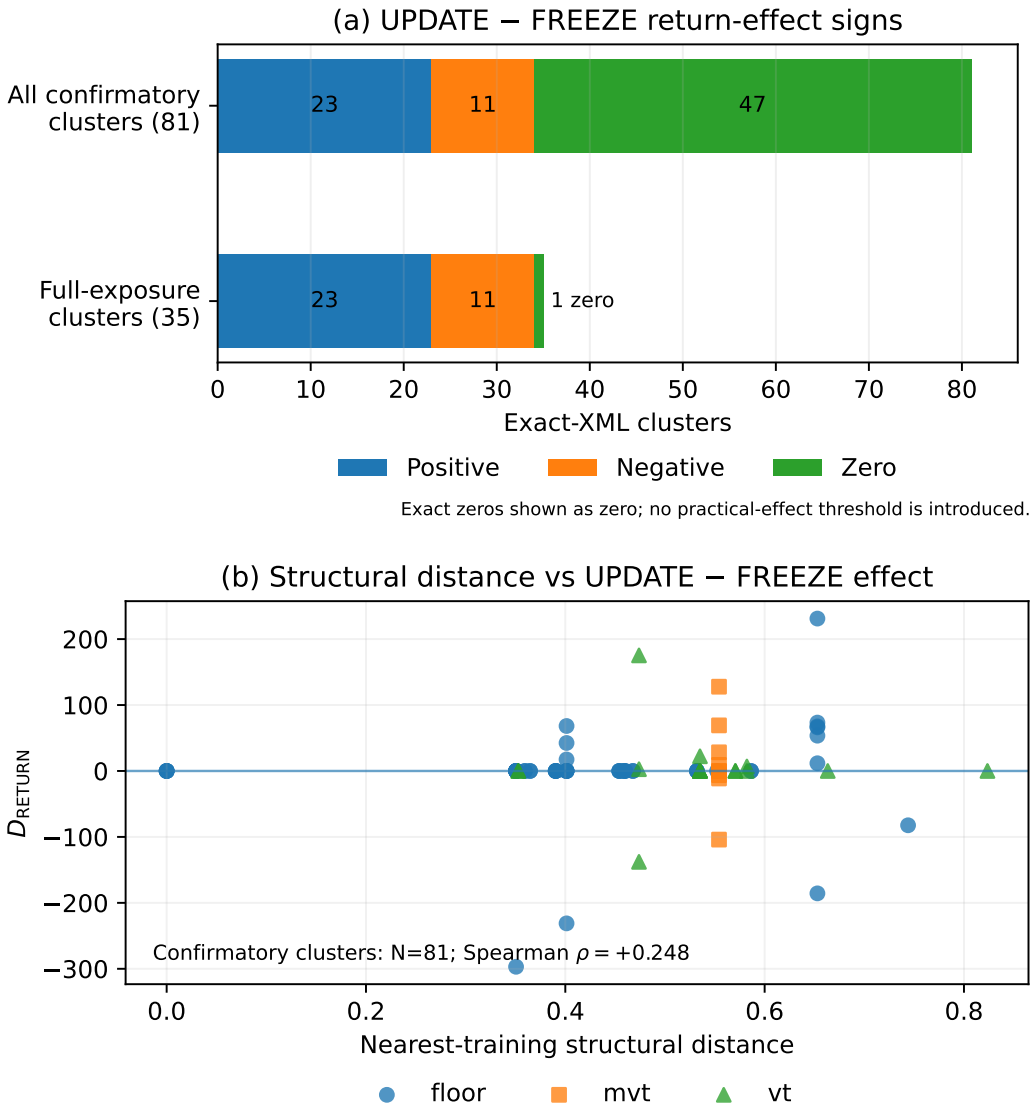


Figure 6: Frozen ABA confirmation. (a) Signs of $D_{\mathrm{RETURN}}$ for all 81 clusters and the 35 exposure-qualified clusters. (b) The distance–effect scatter for all 81 confirmation clusters, one point per cluster; overlaps can hide coincident points. Exact-zero signs use no new practical-effect threshold. One training seed and one evaluation seed.

are equally averaged. Percentile endpoints use 0.025 and 0.975 quantiles with linear interpolation. Training seeds are not resampled. This procedure preserves the original point estimator instead of reweighting each morphology cluster equally. In contrast, ABA explicitly uses equal-cluster effects.

## E Repeated-change exposure and structural distance

The moderate RETURN schedule sets motor strength to 0.8 at serialized index 250 and to 1.0 at 550. The matched HOLD schedule retains 0.8 at 550 and returns to 1.0 at 800, outside the primary $[550, 800)$ window. Friction and mass are not the manipulated factors in this experiment. FREEZE, also called HOLD_Z_PRE2, fixes only the controller-facing latent at its pre-second-change anchor. Encoder computation, observations, and previous-action history continue, so this is not an environment freeze.

Table 7: Frozen repeated-change cluster counts. Exposure is conditional; signs use exact zero. No new threshold or significance test is introduced.

| Population / quantity | $N$ | Positive | Negative | Zero |
|---|---|---|---|---|
| Confirmation $D_{\mathrm{RETURN}}$ | 81 | 23 | 11 | 47 |
| Full-exposure $D_{\mathrm{RETURN}}$ | 35 | 23 | 11 | 1 |
| Exposed floor | 19 | 13 | 6 | 0 |
| Exposed mvt | 8 | 5 | 3 | 0 |
| Exposed vt | 8 | 5 | 2 | 1 |
| Confirmation $I_{\mathrm{RETURN}}$ | 81 | 18 | 17 | 46 |
| ID12 $D_{\mathrm{RETURN}}$ | 12 | 0 | 0 | 12 |

For each branch, $D = Y_{\mathrm{UPDATE}} - Y_{\mathrm{FREEZE}}$ uses the sum of serialized step rewards at indices 550–799. Verified environmental termination permits zero reward continuation to the end of this window; no observations, latents, or simulated transitions are fabricated. This continuation rule belongs to this estimand and is not applied to the main-table velocity metrics. Walker effects are averaged inside exact-XML clusters; cluster effects then receive equal weight.

Exposure requires a changed live latent relative to the frozen anchor and a changed paired action, using the specified $10^{-8}$ latent and action tolerances over observed decisions from 550 onward. Paired actions use common decisions only. Every constituent walker must qualify for its cluster to enter the full-exposure cohort; the interaction cohort additionally requires both RETURN and HOLD branches. Full exposure is not synonymous with surviving the complete reward window. There are 58 second-event-eligible, 55 complete-window, and 35 RETURN-full-exposure confirmation clusters. The primary sign count uses exact zero, with no newly chosen near-zero threshold.

Pilot discovery robots are not pooled into confirmation. The separate ID12 diagnostic has zero measured latent/action exposure and $D = I = 0$ throughout, so it supplies no positive in-distribution updating result. The HOLD-relative interaction is not established as positive at population level. No bootstrap or multi-training-seed inference is added to this one-seed ABA diagnostic.

The structural descriptors are policy-body token count, valid action dimensions, maximum tree depth, mean nonleaf branching, maximum branching, leaf fraction, and edge-length mean, standard deviation, and maximum. Training-only mean and standard deviation define z-scores; zero-variance dimensions are dropped. Let each group contain only its retained features (none were dropped in this analysis). Complexity, tree, and geometry groups receive equal weight in the RMS distance:

$$d(m, \mathcal{M}_{\mathrm{train}}) = \min_{m' \in \mathcal{M}_{\mathrm{train}}} \left[ \frac{1}{3} \sum_{g \in \{C,T,G\}} \frac{1}{|g|} \sum_{j \in g} (\tilde{q}_j(m) - \tilde{q}_j(m'))^2 \right]^{1/2}. \tag{7}$$

This descriptor-based distance does not capture all morphology or dynamics differences. The $D$ and $I$ correlations use all 81 confirmation clusters, not the 35 exposed clusters. The within-mvt correlation is undefined due to constant distance; it is not assigned zero.

## F PRIVILEGED-LATENT TRACKING DIAGNOSTIC

The observer-only reference supplies the true motor, friction, and mass multipliers plus 13 zeros to the frozen Teacher and reads its 16-dimensional post-ReLU policy latent. Student and Teacher are compared in the same output space. The reference does not replace deployed actions, modify simulator state, or introduce future observations.

The diagnostic covers Strict-OOD97 (64 floor, 11 mvt, 22 vt IDs), training seed 1409, evaluation seed 1409, and mutation OOD-strong. The diagnostic reports family-dependent Teacher-L2 behavior: SA has lower error consistently in mvt, but does not uniformly reduce error in floor or vt. The post-ReLU representation also contains sparse and zero-valued conditions. The underlying numerical result records are unavailable; this auxiliary diagnostic is therefore reported qualitatively, without latent-error magnitudes.

Euclidean reconstruction measures agreement with a particular learned privileged encoder. It is not physical-parameter accuracy, a representation-invariant metric, or a complete measure of control

Table 8: Single-training-seed external comparisons. Hyper denotes HyperDistill with a MetaMorph-DR Teacher, not an exact official reproduction. Unreported fixed cells are omitted.

| Setting | Initial condition | Modu Return | Hyper Return | Modu MND | Hyper MND |
|---|---|---|---|---|---|
| Fixed | Nominal | 1049.7 | 1025.3 | – | – |
| Mutation | Nominal | 1054.1 | 958.7 | 2.67 | 2.18 |
| Mutation | ID | 986.1 | 1016.4 | 3.01 | 2.62 |
| Mutation | OOD-mild | 877.6 | 954.6 | 3.94 | 2.81 |
| Mutation | OOD-strong | 838.2 | 825.0 | 3.81 | 3.08 |

utility. Determining how the temporal representation encodes useful dynamics beyond this reference remains an open mechanism question. Earlier morphology-agnostic explicit motor/friction/mass regression is a separate auxiliary diagnostic and is not the deployed MorphAdapt method; its failure on disjoint morphologies would not establish failure of the complete morphology-conditioned controller.

## G FALLEN-STATE ROBUSTNESS

The paired stress test evaluates identical frozen states under fixed nominal physics with deterministic rollouts. There are 34 states per method and matched seed, or 204 episodes across two methods and seeds 1410, 1411, and 1415. These seeds are distinct from the main-table pairing. Twenty-six states are inverted; the post-hoc moderate subset contains eight states with positive initial signed uprightness.

Strict self-righting requires 0.5 s without torso-ground contact and normalized height/uprightness thresholds 0.8/0.9. The relaxed thresholds are 0.75/0.8. Resumed locomotion further requires one second upright and mean forward speed at least half the controller's upright-reference speed. Thus, a posture recovery is not itself resumed locomotion.

Strict resumed locomotion is 0/34 for both methods in every seed. In the relaxed moderate subset, SA self-righting is 6/8 in each seed and locomotion is 2/8 in each seed. DR self-righting is 1/8, 7/8, and 4/8; locomotion is 1/8, 1/8, and 0/8. These yield the descriptive seed averages of 75% versus 50% for self-righting and 25% versus 8.3% for locomotion. Relaxed inverted self-righting is 0/26 for both methods in each seed. A qualitative tendency in an individual shape, if observed, would not meet these successful-recovery criteria. No general self-righting, significance, or previous-action causal claim follows.

## H ADDITIONAL BASELINES AND PROVENANCE BOUNDARIES

ModuMorph uses released FT code, a reported 100M training configuration, and training seed 1409 under the common Strict-OOD97 evaluation. Its native evaluation differs from the paper's described rollout protocol; it is not claimed as exact reproduction of the published number. Mean mutation MND is reported by protocol alongside the main comparisons below.

HyperDistill uses a MetaMorph-DR Teacher rather than the original paper's universal Transformer/ModuMorph Teacher. Its HN-MLP is distilled from 100 training-parent morphologies and 1000 morphology-mutated robots, with 8000 expert transitions each (8M total), for 150 epochs. The evaluated checkpoint is epoch 150. Morphology-specific weights bind once per episode and do not regenerate after mutation. The single distillation seed supports only descriptive point estimates; its identifier and evaluation-seed identifier are not explicitly recorded in the available documentation. Missing fixed-dynamics cells are not inferred. These budgets differ from MorphAdapt's interactive Student training.

The main table is linked to the Student configuration and checkpoint manifests and the final scalar aggregation. Source checks establish numerical agreement and distinguish legacy runs. Remaining reproducibility gaps concern the complete Teacher training lineage, historical initialization load receipts, some DR completion/source records, and the original bootstrap generator bytes. The present manuscript preserves the reported estimates without claiming a fresh simulator replay or

reconstruction of missing provenance. The older privileged Teacher evaluation uses a different OOD98 population and cannot supply a matched Strict-OOD97 upper-bound column.

# I GATE7 DEVELOPMENT CASE AND ITS SCOPE

Gate7 is an earlier, nominal-100M matched SO/SA development comparison with training and evaluation seed 1409. It uses one deterministic episode per robot/protocol on legacy OOD98: 98 IDs and 88 exact-XML clusters, including one ID excluded from Strict-OOD97. Mutation occurs at step 250 and the horizon is 1200 steps. Its source, configurations, and recorded checkpoints differ from the main three-seed comparison. The scheduled training count is 99,778,560 transitions; full training logs and checkpoint bytes were not available for a fresh verification of completion or checkpoint loading. Thus the 100M label denotes the configured scale, not exactly 100M verified completed transitions.

Mutation Return SA-minus-SO point differences are $+6.20$, $+19.48$, $+37.95$, and $+55.34$ in nominal, ID, OOD-mild, and OOD-strong, respectively. Their archived 95% walker-paired bootstrap intervals are $[-155.90, 160.32]$, $[-123.24, 165.21]$, $[-72.46, 159.01]$, and $[-49.44, 162.55]$. Each comparison contains 98 paired walkers; all intervals include zero. The archive uses 20,000 walker-level paired resamples, distinct from the main XML-cluster bootstrap and not a measure of training-seed uncertainty.

Gate7 RMRT uses the sustained recovery confirmation time for recovered mutation-exposed episodes and assigns $1200 - 250 + 1 = 951$ control steps to exposed episodes without recovery. Recovery requires 25 consecutive samples at or above 0.8 times the 25-step pre-mutation forward-velocity baseline. Episodes terminating before mutation are excluded. The paired comparison uses walkers valid in both methods: 75 for OOD-mild and 73 for OOD-strong. These intersections differ from the individual-method counts (77/78 and 75/74, SO/SA), so the paired effects in Table 1 are not obtained by subtracting the separate aggregate means. RMRT is not recovered-only RT. The source-bound control interval is 0.02 s, but we retain control-step units to avoid conflating this restricted estimator with measured wall-clock recovery.

Gate7 is not the main seed1409 run, and it is not an independent replicate under the main protocol. It supports a development-case observation of substantial point effects with uncertainty crossing zero. The three-seed results characterize the current pipeline's average incremental action-history effect; no seed or main evaluation condition is removed or replaced by Gate7.

Previous-action conditioning can yield substantial post-perturbation point gains in an individual training realization, but the current three-seed comparison does not show reliable gains across training seeds.