When Predictions Become Inputs: Hidden Feedback Errors in Latent World Models

Anonymous authors
Paper under double-blind review

Abstract

In autoregressive latent world models, a prediction describes a task state and becomes input to the next prediction. We show that these roles can diverge: changing feedback improves future prediction while its current nonlinear task readout stays fixed. Our oracle diagnostic guides a predicted token toward its observed encoding, preserving the readout, original ReLU activation region, earlier history and future actions. Free rollout tests usefulness; norm-matched donor guidance tests source sensitivity. Across Transformer and GRU PushT models, stronger readouts and matched coordinate training retain terminal block-position MSE reductions of 9.3% and 8.8% for coordinate-teacher and physical-label objectives. On 256 fresh recipients, corrections preserving two pose readouts retain reductions of 6.4% and 6.7% under a reserved evaluator. Benefits also survive joint pose and qualified velocity constraints. Recursive Transformer training lowers free-rollout MSE by 23.1% and 30.9% under the preserved readout. Prospective tests do not confirm incremental prediction of training response or improved action selection; native DINO-WM PushT and PointMaze do not confirm correction over free rollout. These findings establish feedback insufficiency relative to specified task measurements and separate it from diagnostic value and decision quality.

1 Introduction

A robot imagining a push uses predicted representations both to describe the object and to predict its next interaction. Latent prediction underlies image and video representation learning (Assran et al., 2023; Bardes et al., 2024) and planning with predicted visual features (Zhou et al., 2024).

Error accumulation describes how prediction errors become input errors and propagate, leaving open which input properties must be correct. Tokens with the same pose readout can lead the predictor to different futures, so current pose-readout error alone cannot determine the better prediction input.

We introduce a controlled diagnostic of *feedback usefulness at a fixed task readout*. Oracle guidance from the corresponding observation changes one predicted token while preserving all six PushT nonlinear pose-readout outputs, the original ReLU activation region, earlier history and future actions. Free rollout tests usefulness; norm-matched unrelated-donor guidance tests source sensitivity.

1. **A controlled test of feedback usefulness.** Changing one token while preserving its nonlinear task readout improves future prediction across six Transformer/GRU groups. Norm-matched donor guidance separates the value of the source from displacement size.
2. **Persistence under stronger measurements.** Benefits survive simultaneous pose constraints, evaluation by a separately reserved readout, and joint pose and qualified velocity constraints. The phenomenon extends beyond a single fitted task measurement.
3. **Boundaries between prediction, diagnosis and decisions.** Recursive training improves free rollout; prospective tests leave incremental diagnosis and action-selection benefits unresolved. Native-model and candidate-cost analyses identify limits of transfer.

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.

2 FROM ERROR ACCUMULATION TO A CONTROLLED DIAGNOSTIC

E includes image preprocessing, the encoder and observation projector; action-conditioned predictor F_θ includes the action encoder and prediction projector. Task readout g maps tokens to physical outputs using fixed input and output normalization. For three-token history $H[z] = (h_1, h_2, z)$, we study

$$g(z) = g(z') \quad \text{while} \quad g(F_\theta(H[z], U)) \neq g(F_\theta(H[z'], U)). \tag{1}$$

With identical actions U and earlier history, Equation 1 violates readout sufficiency: equal current readouts yield unequal next readouts in the learned computation, without asserting equal physical states.

Useful correction must improve prediction of the same executed future. A donor comparison alone is insufficient: an unrelated donor may damage useful feedback.

Observed-history prediction supplies encodings at every transition; full reset replaces only the insertion token. Both may change its readout. Aggregate rollout gaps do not isolate these changes (Table 18; Appendix Q).

Preserving the pose readout and earlier decoded samples fixes their deterministic finite differences, without establishing equal instantaneous velocity, contact state or all image information. We call an improvable feedback difference invisible to these specified readouts a "hidden feedback error"; this does not identify a missing physical variable.

Relation to existing explanations. Observed/predicted-input distribution shift motivates scheduled sampling and imitation learning (Bengio et al., 2015; Ross et al., 2011); multi-step latent consistency and action-conditioned prediction are established tools (Hafner et al., 2019; Schwarzer et al., 2021; Ye et al., 2021; Zhao et al., 2023). Our intervention isolates *useful feedback changes at a fixed task description*. Subsequent dynamics can still amplify the changed component, so it complements error accumulation.

DeepMDP and value equivalence distinguish reconstruction from behaviorally relevant information (Gelada et al., 2019; Grimm et al., 2020; 2021). You et al. (2026) distinguish data-averaged prediction error from plan-cost error under planner-induced distribution shift; Vakalis (2026) audit whether imagined action effects match environment interventions. We change the *feedback representation*, holding the task readout, earlier history and actions fixed, to isolate a useful within-model difference invisible to the preserved outputs.

3 EXPERIMENTAL DESIGN

3.1 MODELS AND STRONGER TASK MEASUREMENTS

Three training pools and Transformer/GRU predictors define six groups, each with three continuation objectives. The latent objective matches predicted and observed encodings in raw-token space. Coordinate-teacher training matches their decoded outputs; physical-label training matches decoded predictions to simulator labels, both with fixed six-output standardization (Appendix I). Perception is shared and frozen within each group. Tokens have 192 dimensions; the readout has two ReLU hidden layers and six outputs, $y = (x_a, y_a, x_b, y_b, \sin\theta_b, \cos\theta_b)$. Training uses 256 expert episodes and 21,000 updates per pool; continuation uses 1,280 planner windows and 2,100 updates. The GRU restarts on each three-token window, without persistent hidden state.

We fit 72 pose readouts and apply fixed validation gates (Appendix A). All six stronger g_A fits pass; observed-encoding planner-validation block MSE remains 1364.0–2869.7 pixels squared. Five second-seed g_B fits pass; the failed fit is retained. Normalizers remain fixed.

Coordinate continuations are retrained with g_A under the original schedule, optimizer and complete-gradient normalization; validated latent continuations are reused. This tests stronger readouts *and matched training*, not measurement error alone (Appendix A; Table 50).

Evaluation-only measurement. Six GELU residual readouts, g_{eval}, use separately assigned fitting, validation and qualification data. They share frozen encoders but never enter predictor train-

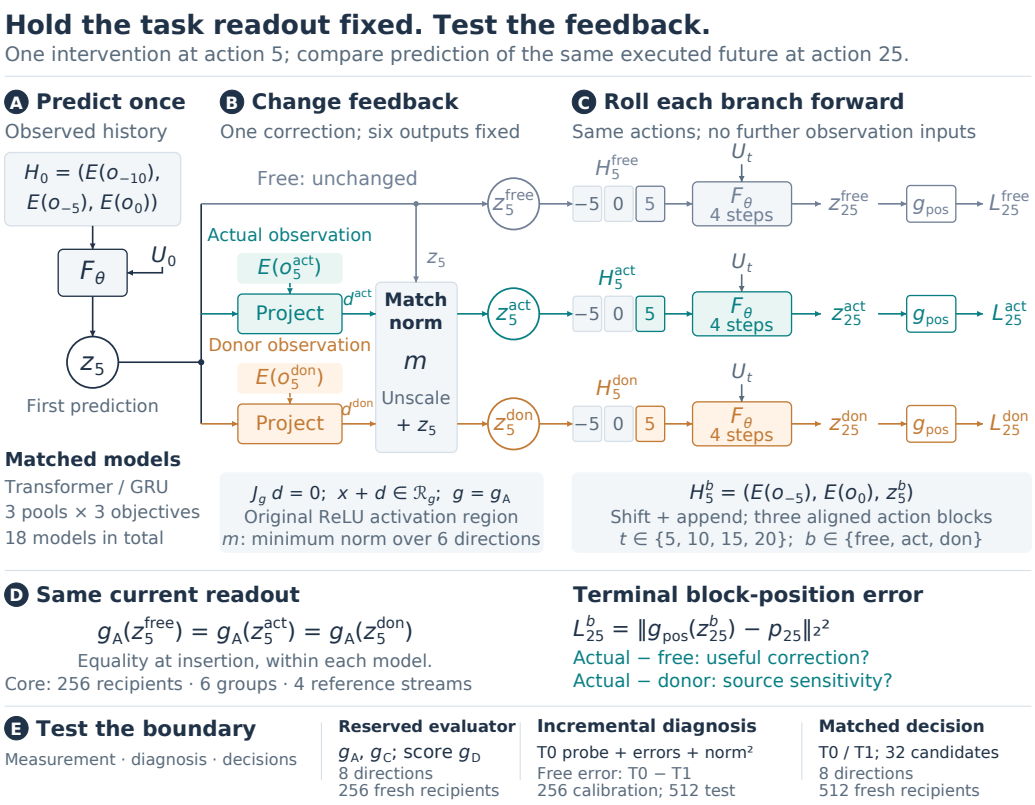

Figure 1: **Hold the task readout fixed; test the feedback.** (A–B) Actual/donor guidance changes z_5 while preserving all six outputs of g_A and its original ReLU activation region. Corrections share the minimum standardized norm over three objectives and two sources. (C) Separate branches take four five-action transitions with frozen F_θ and identical actions; history cells show token times. (D) g_{pos} selects block x, y; p_{25} is the true endpoint. (E) Prospective measurement, diagnosis and decision tests have separate populations and eight-, four- and eight-direction matching families, respectively (Appendix B).

ing, constraints or displacement selection. They rescore all saved pose-confirmation and training branches post hoc (Appendix R).

3.2 Intervention, Timing, and Displacement Controls

Time counts primitive actions in blocks of five. From $(E(o_{-10}), E(o_{-5}), E(o_0))$, each transition uses three tokens, three aligned action blocks and F_θ's final successor. We intervene at z_5 and evaluate at action 25, fixing earlier history, parameters and future actions.

For standardized input $x = (z - \mu_z) \oslash \sigma_z$ and guidance encoding x_o, we solve

$$\min_d \ \tfrac{1}{2}\|d - (x_o - x)\|_2^2 \quad \text{s.t.} \quad J_g d = 0, \quad x + d \in \mathcal{R}_g \quad \text{for every } g \in \mathcal{G}. \tag{2}$$

Here $\mathcal{R}_g$ is the original ReLU activation region and J_g the Jacobian with respect to x. The readout is affine within this region, so the constraints preserve its full output. Core uses $\mathcal{G} = \{g_A\}$ and dual uses $\{g_A, g_B\}$, with shared input standardization. Projection finds the closest displacement within this region, not the full nonlinear fiber. In an earlier 720-pair test, every naive local-null projection violated nonlinear readout preservation; every constrained projection passed (Appendix E).

Actual guidance uses the recipient's $E(o_5)$; donor guidance uses an independently assigned observation from a separate bank. Projected directions share their family's minimum norm. Core matches six directions: three objectives by two sources. Dual uses Pool 0 Transformer/GRU and twelve directions: single/dual constraints by three objectives by two sources. Its single branch is rescaled and rerolled at that common norm, with g_A scoring.

Table 1: **Main tests of feedback usefulness.** Counts refer to recipients; diagnosis has separate calibration/test banks. All coordinate continuations use g_A for training. † marks post-hoc scoring. The complete historical map, including dual, native and earlier decision tests, is in Appendix B.4.

Test	Fixed models / recipients	Preserved → scored	Main question
Core	6 Transformer/GRU groups / 256	$g_A \to g_A, g_{\text{eval}}^{\dagger}$	Is readout-preserving feedback useful?
Reserved evaluator	Pool 0 Transformer/GRU / 256 fresh	$g_A, g_C \to g_D$	Does benefit survive an extra pose constraint?
Motion	Pool 0 Transformer/GRU / 512	Pose/velocity variants $\to g_A$	What remains under richer measurements?
Training	3 Transformer pools / 256	$g_A \to g_A, g_{\text{eval}}^{\dagger}$	Does recursive training reduce error?
Diagnosis	6 T0/T1 pairs / 256; 512	$g_A \to g_{\text{eval}}$	Does the probe add predictive information?
Matched decision	6 T0/T1 pairs / 512 fresh	$g_A \to g_A$	Does feedback improve a subsequent choice?

Dynamics and insertions use FP32; independent FP64 checks recompute full nonlinear readouts and original activation-region constraints on inserted tokens, with 10^{-6} tolerances in normalized coordinates. Legal zero displacements remain; solver failures block analysis. All 14,592 original PushT pose, native and training matching families pass with nonzero norms, no numerical backoff and no missing families.

3.3 A TEST WITH RICHER MOTION CONSTRAINTS

Preserved pose history and actions fix any velocity estimate computed from them. We therefore fit a five-output readout of instantaneous agent and block velocities, including block angular velocity, using all three latent tokens plus a frozen pose/action baseline. Both Pool 0 architectures qualify on a separate 1,024-state test; complete fitting, qualification and earlier failures appear in Appendix C.

For conditional readout $h_c(x)$ and fixed orthogonal Q, compare preservation of g_A, (g_A, h_c) and $(g_A, h_{Q,c})$, where $h_{Q,c}(x) = h_c(x_0 + Q(x - x_0))$. The rotated control preserves the full conditional Jacobian singular spectrum, but its pose-tangent spectrum and activation geometry can differ. On 512 fresh recipients and 512 separate donors, three constraint sets, two objectives and two sources share a 12-direction matching family. All 4,096 families pass. Each architecture's three primary contrasts test source benefit, attenuation by velocity preservation and attenuation beyond the rotated control. The six primary contrasts and six separate secondary usefulness contrasts each use 99.1667% recipient-bootstrap intervals. Objectives and reference streams are averaged within recipients before resampling.

3.4 NATIVE-MODEL AND TRAINING TESTS

Native DINO-WM. Released DINO-WM PushT (epoch 2) and prospectively selected PointMaze (epoch 10) checkpoints retain their native representations, predictors and rollout timing (Zhou et al., 2024). Only the 196×384 visual entries change at insertion; predicted proprioception and known-action entries stay fixed. Proprioception subsequently evolves per branch; aligned actions follow the official replacement rule. Each task has its own readout, bank and two-direction matching family. PointMaze uses 256 fresh trajectories, a qualified two-output position readout, relative insertion control 5 and endpoint 25 (Appendix S). Absolute errors are not comparable across tasks.

Training and remaining correctability. Three Transformer pools and two coordinate objectives start from their pose-confirmation checkpoints. T0 uses observed history for three rolling transitions; T1 feeds predictions back with gradients across steps; T2 adds standardized latent anchoring, with $\lambda = 0.1$ selected on development validation. All share continuous six-frame examples, target times, 2,100 updates, batch schedule and frozen perception/readout. T0 takes each transition's final

successor, replacing the old one-forward three-successor baseline. Update times differ by less than 1%.

Define $G = \mathbb{E}[L^{\text{free}} - L^{\text{act}}]$ at action 25. Positive G measures benefit at the specified readout, guidance and norm. Within each pool/objective/recipient/reference-stream case, the three conditions and both sources share the six-direction minimum norm. We jointly assess free-rollout MSE, G, measurement, displacement and action-response diagnostics.

3.5 Prospective diagnosis and matched decisions

Two further tests use the six fixed T0/T1 pairs from the three Transformer pools and two coordinate objectives. For diagnosis, 256 calibration recipients and 512 separate test recipients each have a T0 probe stream and a separate T0/T1 response stream. The probe feature $s_{\text{probe}} = L^{\text{don}} - L^{\text{act}}$ measures source sensitivity under g_{eval}. A fixed ridge predictor estimates the response $Y = L^{T0,\text{free}} - L^{T1,\text{free}}$ from ordinary free-rollout errors, observed-history error and squared displacement norm, using their full quadratic basis. The augmented predictor adds only s_{probe}. Test predictions are fixed before the response is computed. This evaluates extra information on fresh contexts for known model pairs, rather than unseen-model selection (Appendix B.2).

For decisions, another 512 recipients share an executed five-action prefix and 32 fixed 20-action suffixes within each pool. Each T0/T1 model retains its own predicted insertion token; actual/donor guidance is shared across suffixes. Both conditions, objectives and sources share an eight-direction matching family. Predictions use the same g_A to score block position and wrapped angle. We then compare realized physical cost of the selected candidates, averaging exact predicted-cost ties. The two primary contrasts are T1 actual minus free and T1 actual minus donor, with 97.5% recipient-bootstrap intervals. This tests one subsequent choice with the newer models, keeping the earlier decision results separate (Appendix B.3).

3.6 Frozen populations and inference

Table 1 summarizes the main tests; Appendix B.4 maps the earlier confirmation populations. Pose/training, native PushT and PointMaze each have a separate bank of 256 donors. Frozen actions are shared across models and branches. Admission and donor assignment do not use future intervention effects; models, inputs, endpoints, budgets and analysis code are locked before confirmation.

The PushT endpoint is block-position squared Euclidean error, $L = (\hat{x}_b - x_b)^2 + (\hat{y}_b - y_b)^2$, in pixels squared. Groups/reference streams are averaged within recipient before 20,000 recipient-bootstrap draws. The nine original pose/native primary contrasts use $1 - 0.05/9 = 99.444\ldots\%$ percentile intervals; eight training contrasts separately use 99.375%. Relative changes recompute aggregate means per draw. Intervals condition on fixed models, actions and realized donors. Earlier 128-recipient prediction and 512-recipient decision confirmations remain separate.

4 Results: what remains hidden, and when it matters

4.1 Useful corrections survive stronger and simultaneous pose readouts

With g_A, actual-guided correction reduces terminal block-position MSE for the coordinate-teacher objective from 8448.1 to 7662.0: a change of -786.1 $[-998.8, -609.0]$, or 9.3%. For the physical-label objective, terminal block-position MSE falls from 8583.8 to 7829.3: -754.5 $[-926.0, -607.8]$, or 8.8%. All six group-level changes are negative for both objectives (Figure 2). The aggregate benefit therefore persists after both readout quality and coordinate training are strengthened.

Actual guidance also outperforms matched donor guidance for all three objectives: latent -737.9 $[-938.8, -546.9]$, coordinate teacher -643.8 $[-811.9, -488.3]$, and physical labels -551.7 $[-699.6, -413.5]$. Thus source matters at a fixed standardized displacement norm, without identifying the hidden physical information it carries.

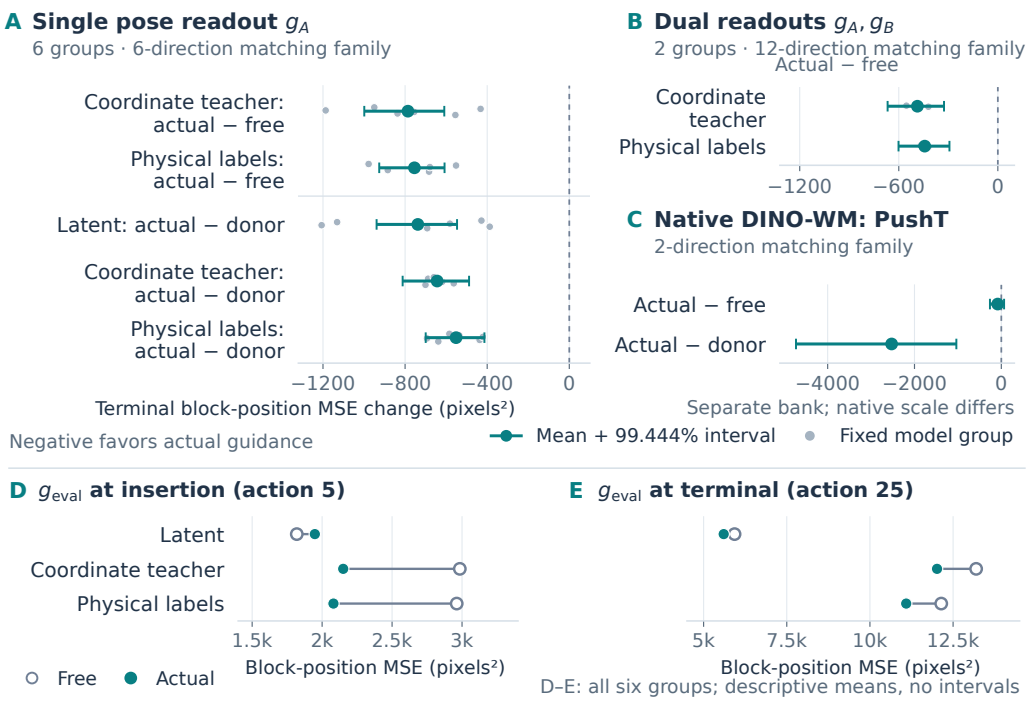

Figure 2: **Correction benefits and insertion-time measurement changes.** A–C: original nine PushT pose/native primary contrasts, with 99.444% recipient-bootstrap intervals and all fixed group estimates (gray). Negative favors actual guidance. Matching families contain six, twelve and two directions for single, dual and native tests, respectively. Single/dual score with g_A; native PushT uses separate recipients and a different scale. D–E: evaluation-only g_{eval} scores the same core branches at insertion and the terminal horizon. Hollow/filled points are free/actual means over all six groups, descriptive without intervals. All three objectives are retained. g_A stays fixed at insertion; other pose estimates can change.

The latent objective also improves over free rollout: -322.6 $[-419.4, -226.6]$, an exploratory 95% interval outside the primary family (Appendix Q).

Preserving both original pose fits retains coordinate-teacher and physical-label improvements of -487.2 $[-667.6, -325.9]$ and -442.8 $[-601.5, -293.1]$ (5.8% and 4.8%). The unused-in-training second fit fails its Pool 0 GRU angle-qualification gate. Appendices A and Q.5 retain that failure, matched single/dual comparisons and post-hoc second-readout scoring.

Future benefit persists under evaluation-only measurement. All six g_{eval} fits qualify. On the unchanged core branches, coordinate-teacher and physical-label terminal block-position MSE fall by 8.9% and 8.7%: -1175.7 $[-1458.5, -909.5]$ and -1053.5 $[-1278.2, -850.3]$. Actual guidance also outperforms donor guidance for all three objectives; both dual-readout actual–free comparisons remain favorable. These seven comparisons use a separate nine-slot 99.444% interval family.

The same evaluator detects insertion-time block-position RMS displacements of 30.2 and 30.1 pixels, with insertion MSE falling from 2984.1 to 2150.7 and from 2961.9 to 2081.0. Latent-objective MSE instead rises from 1819.0 to 1948.0 at insertion while falling from 5930.7 to 5603.9 at the endpoint (Figure 2D–E). Thus "hidden" is relative to the specified readout: future benefit survives independent measurement, but preserving g_A does not preserve every pose estimate. This does not identify non-pose information or attribute future benefit to insertion-error reduction (Appendix R).

A prospective test with a reserved evaluator. An additional pose constraint g_C and reserved evaluator g_D use disjoint measurement parents; all four Pool 0 Transformer/GRU fits qualify before confirmation. On 256 fresh recipients, corrections preserve g_A and g_C while g_D only evaluates

predictions. Eight directions—two constraint sets, two objectives and two sources—share a common standardized norm.

Under the reserved evaluator, actual guidance improves over free rollout by 676.6 [$398.2, 960.9$] and 670.7 [$421.6, 922.5$] pixels squared for coordinate teacher and physical labels, respectively. Its advantages over matched donor guidance are 733.0 [$459.8, 1033.4$] and 371.6 [$149.5, 601.4$]. All four primary 98.75% intervals exclude zero; the descriptive reductions relative to free rollout are 6.4% and 6.7%. Adding g_C attenuates correction benefit at the same norm, but useful, source-sensitive correction remains. Thus the effect survives an extra pose constraint and an evaluator reserved before confirmation. Appendix B.1 gives all eight contrasts, including the unresolved physical-label insertion-displacement contrast.

4.2 Useful corrections survive joint pose and velocity constraints

With both pose and velocity readouts fixed, prespecified secondary comparisons show Transformer error falling from 6908.6 to 6492.6, a benefit of 416.0 [$293.1, 541.1$], or 6.0%; GRU error falls from 9667.3 to 8754.3, a benefit of 913.0 [$684.1, 1160.6$], or 9.4%. All three constraint conditions improve over free rollout in both architectures (Figure 7B). Full five-output rank on the pose tangent rules out a redundant velocity readout as the reason for persistence.

The fresh 512-recipient test confirms a pose-preserving source benefit in both architectures: $\Delta_{\text{pose}} = 463.5$ [$311.9, 622.3$] for the Transformer and 534.0 [$311.9, 752.9$] for the GRU. Preserving the velocity readout attenuates that benefit by 81.0 [$4.0, 164.1$] in the Transformer, but its difference from the rotated control is only 27.9 [$-59.7, 117.7$]. The corresponding GRU contrasts are -21.3 [$-147.5, 103.9$] and 4.3 [$-129.1, 136.7$]. Neither architecture meets the planned criterion for velocity-specific attenuation (Figure 7A).

The velocity readout's mean block-component training-mean R^2 is 0.09–0.39 on predicted tokens versus 0.52–0.64 on observed tokens. This combines prediction and measurement error, rather than isolating readout error. Complete post-hoc branch changes are in Appendix Q.

4.3 Recursive training improves free rollout

Under g_A, T1 reduces free-rollout terminal block-position MSE relative to T0 by 23.1% for the coordinate-teacher objective and 30.9% for the physical-label objective. The paired changes are -1689.7 [$-2237.6, -1217.3$] and -2352.8 [$-2978.2, -1802.5$]. At the same time, G falls from 546.9 to 328.2 and from 661.8 to 364.2. Their changes, -218.7 [$-397.3, -43.6$] and -297.6 [$-467.9, -138.3$], also exclude zero under the separate eight-contrast correction. All three pools move toward lower free-rollout MSE and lower G (Figure 3).

A smaller G alone need not imply improvement. Here free-rollout MSE improves while the readout and observed-encoding MSE stay fixed. Standardized displacement norms are identical across conditions (means 3.193 and 3.187 for coordinate teacher and physical labels), with no zero cases. T1 retains positive G; Appendix A reports output variance and action-response diagnostics.

All four T2–T1 contrasts are unresolved with g_A. With g_{eval}, however, T2 lowers free-rollout MSE by 1344.7 [$860.4, 1801.1$] and 1221.4 [$824.0, 1621.7$]. T1 improves free rollout under both readouts, but its reduction of G under g_{eval} is resolved only for physical labels. Thus recursive training improves prediction, while remaining correctability and the advantage of extra latent anchoring depend on measurement. Appendix R retains all eight separately corrected 99.375% comparisons and branch means.

4.4 The probe does not confirm incremental prediction of training response

On the fresh 512-recipient test bank, adding s_{probe} changes response-prediction MSE from 464.392 to 463.923 million pixels to the fourth power. The paired reduction is 0.469 million [$-0.223, 1.650$ million; 95% interval], a descriptive 0.10%. This does not confirm extra predictive information for these known model pairs (Appendix B.2). A controlled correction can reveal feedback insufficiency without serving as a validated forecast of training response.

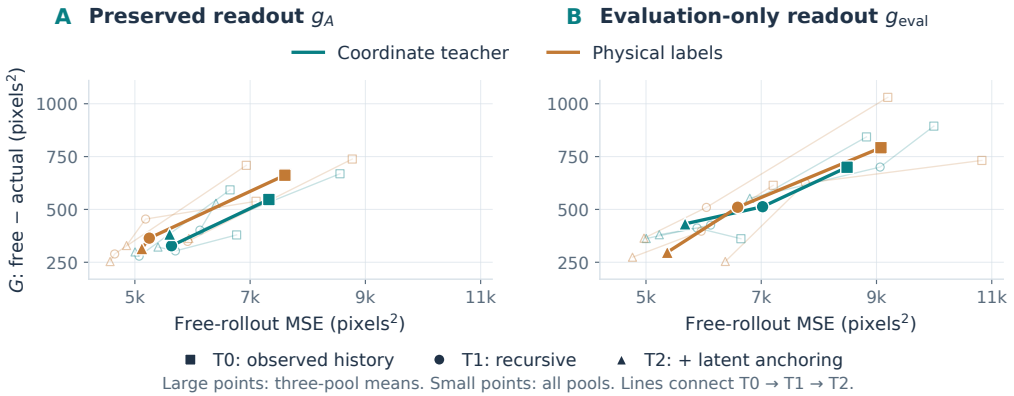

Figure 3: **Training comparisons under the preserved and evaluation-only readouts.** Both panels score the same saved training branches at the original common standardized displacement norm. Large points average all three Transformer pools; small points retain every pool. Each panel computes free-rollout terminal block-position MSE and G at action 25 with its named readout. T2 improves free rollout beyond T1 under g_{eval}; the g_A comparisons are unresolved. Points are descriptive, not joint confidence regions; paired intervals are in Appendices A and R.

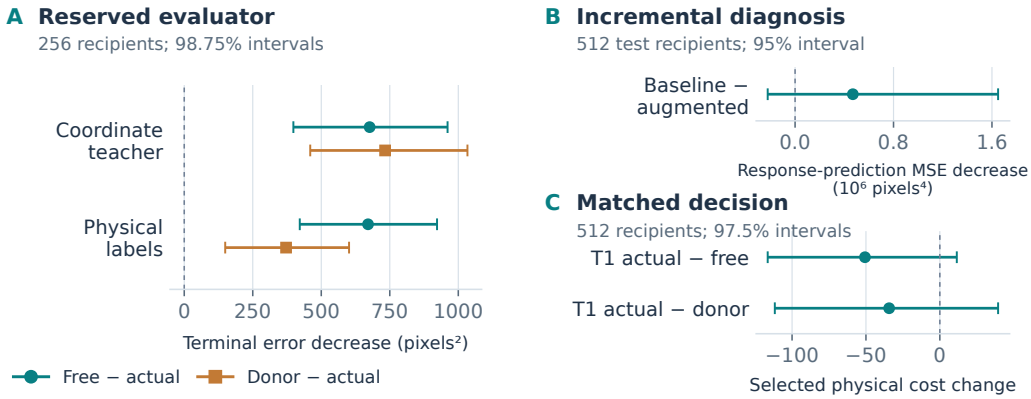

Figure 4: **Prospective tests separate measurement, diagnosis and action choice.** A: corrections preserve g_A and g_C, while reserved g_D scores the endpoint (98.75% intervals). B: change in response-prediction MSE after adding the feedback probe to the ordinary-error baseline (95%). C: realized physical cost of T1's selected candidates, with uniform exact-tie selection (97.5%). All seven primary contrasts are retained, with separate populations and units. Positive favors actual guidance or the augmented predictor in A/B; negative favors actual guidance in C.

4.5 Native transfer separates source sensitivity from useful correction

For the released PushT DINO-WM checkpoint, terminal block-position MSE is 1522.2 free, 1439.4 actual-guided, and 3965.8 donor-guided. Actual–donor is -2526.5 $[-4729.4, -1035.5]$, whereas actual–free is -82.9 $[-261.9, +63.8]$. The latter relative interval spans a 16.3% reduction to a 4.5% increase. It neither confirms a benefit nor establishes equivalence.

On PointMaze, terminal position MSE is 0.116572 free, 0.119259 actual-guided and 0.125025 donor-guided in native units squared. Actual–free is $+0.002687$ $[-0.010752, +0.018450]$; actual–donor is -0.005766 $[-0.023783, +0.012721]$. Both are unresolved in the nine-contrast family shared with seven evaluation-only comparisons. Neither task confirms correction over free rollout. Appendices A and S retain all horizons and reset references.

4.6 BETTER COST ESTIMATES NEED NOT IMPROVE DECISIONS

The earlier 512-recipient decision bank uses original models/readouts, not T1 or stronger-readout models. Each group/objective has 32 fixed suffix candidates sharing the intervened prefix. For physical pose-cost errors $\epsilon_k = \hat{C}_k - C_k$ and $\bar{\epsilon} = K^{-1}\sum_k \epsilon_k$,

$$\underbrace{K^{-1}\sum_k \epsilon_k^2}_{\text{cost MSE}} = \underbrace{\bar{\epsilon}^2}_{\text{shared bias squared}} + \underbrace{K^{-1}\sum_k (\epsilon_k - \bar{\epsilon})^2}_{V\text{: centered cost-error variance}} . \tag{3}$$

A shared additive error cancels from pairwise differences. Pairwise margin-error MSE equals $2K/(K-1)\,V$; candidate pairs are not additional independent recipients.

Coordinate-teacher and physical-label cost MSE decrease by 27.38 and 24.15 million squared cost units; shared bias squared decreases by 27.30 and 24.59 million. Their centered cost-error variance changes are -0.078 million $[-0.359, +0.200]$ and $+0.445$ million $[+0.140, +0.753]$, respectively (Appendix A, Figure 6). Most gains concern a component that cannot affect ranking; centered cost-error variance has an unresolved change for the coordinate-teacher objective and worsens for the physical-label objective. Both full-pool and true-top-three pair-agreement changes have intervals spanning zero.

All four original pose-score selected-cost comparisons remain unresolved under 98.75% intervals (Appendix K). Positive secondary results for latent-objective models using native latent-distance scoring remain in that appendix.

A matched test with the recursive-training models. On 512 fresh recipients, T1 actual guidance changes selected physical cost relative to free rollout by -50.7 $[-116.5, 11.5]$ and relative to donor guidance by -34.3 $[-111.7, 39.4]$. Neither 97.5% interval excludes zero (Figure 4C). Thus the matched test does not confirm useful, source-sensitive action selection in these newer models. Appendix B.3 retains all 12 model rows and full-reset references.

5 IMPLICATIONS AND SCOPE

For robotics, these tests separate task measurement, feedback usefulness and decision quality. Useful corrections provide counterexamples to readout sufficiency in the learned predictor. Stronger constraints and reserved evaluation test measurement dependence; training-response prediction and physical action selection test broader uses. The cost decomposition identifies gains in a component shared by candidates that cannot change their ranking.

The oracle requires the insertion observation; constraints cover finite readouts within the original ReLU activation region. Native tests cover one checkpoint per task, and recipient-bootstrap intervals do not cover unseen training seeds or model families. Historical encoder training has a small exact-image overlap with validation; stronger pose-readout fits exclude those images, and confirmation inputs pass the specified lineage checks; PointMaze retains 22 cross-role identical-image groups (Appendix S). Complete external pretraining lineage remains unresolved. Earlier Reacher benefits fade at the original terminal horizon (Appendix H), limiting cross-task persistence.

6 CONCLUSION

Feedback can be predictively insufficient even when current task readouts agree. Useful corrections survive additional pose constraints, a separately reserved evaluator and qualified motion constraints; the velocity-specific explanation remains unresolved. Recursive training improves free rollout, while the fixed probe does not confirm extra training-response information beyond ordinary errors. Neither native task confirms correction over free rollout, and the matched decision test does not confirm improved action selection. Prediction gains, diagnostic value and decision gains require separate evidence.

Reproducibility Statement. Code and analysis scripts are available at https://anonymous.4open.science/r/ICLR_HiddenFeedbackErrors-D443/README.md. The appendices specify data roles, readout fitting, losses, intervention constraints and complete statistical families. Companion evidence packages provide accepted recipient-level arrays, readout weights and normalizers, fixed bootstrap draws, and CPU verifiers for the earlier confirmations. New prospective tests retain their complete populations and independent numerical/statistical acceptance records; Appendix B.5 states measured compute and regeneration requirements. A separate demonstration generates projections and rollouts from raw images and three study-trained checkpoints on two predetermined development cases. It does not reproduce all training or all experimental populations. The reviewer README distinguishes saved-evidence checks, this limited demonstration and assets not redistributed. Editable vector figures and their exact plotting inputs accompany the LaTeX source.

AI Use Disclosure. OpenAI tools were used for language editing and to assist with implementing portions of analysis and visualization code specified by the authors. All outputs were reviewed and validated by the authors.

REFERENCES

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15619–15629, 2023. URL https://arxiv.org/abs/2301.08243.

Adrien Bardes, Quentin Garrido, Jean Ponce, Xinlei Chen, Michael Rabbat, Yann LeCun, Mahmoud Assran, and Nicolas Ballas. Revisiting feature prediction for learning visual representations from video. *arXiv preprint arXiv:2404.08471*, 2024. URL https://arxiv.org/abs/2404.08471.

Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. Scheduled sampling for sequence prediction with recurrent neural networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015. URL https://proceedings.neurips.cc/paper/2015/hash/e995f98d56967d946471af29d7bf99f1-Abstract.html.

Carles Gelada, Saurabh Kumar, Jacob Buckman, Ofir Nachum, and Marc G. Bellemare. DeepMDP: Learning continuous latent space models for representation learning. In *International Conference on Machine Learning*, 2019. URL https://proceedings.mlr.press/v97/gelada19a.html.

Christopher Grimm, Andre Barreto, Satinder P. Singh, and David Silver. The value equivalence principle for model-based reinforcement learning. In *Advances in Neural Information Processing Systems*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/3bb585ea00014b0e3ebe4c6dd165a358-Abstract.html.

Christopher Grimm, Andre Barreto, Greg Farquhar, David Silver, and Satinder P. Singh. Proper value equivalence. In *Advances in Neural Information Processing Systems*, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/400e5e6a7ce0c754f281525fae75a873-Abstract.html.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International Conference on Machine Learning*, 2019. URL https://proceedings.mlr.press/v97/hafner19a.html.

Stephane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pp. 627–635. PMLR, 2011. URL https://proceedings.mlr.press/v15/ross11a.html.

Max Schwarzer, Ankesh Anand, Rishab Goel, R Devon Hjelm, Aaron Courville, and Philip Bachman. Data-efficient reinforcement learning with self-predictive representations. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2007.05929.

Donna Vakalis. The intervention gap in latent world models. *arXiv preprint arXiv:2608.29998*, 2026. URL https://arxiv.org/abs/2608.29998.

Weirui Ye, Shaohuai Liu, Thanard Kurutach, Pieter Abbeel, and Yang Gao. Mastering atari games with limited data. In *Advances in Neural Information Processing Systems*, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/d5eca8dc3820cad9fe56a3bafda65ca1-Abstract.html.

Hanzhe You, Yonggang Zhang, Maohao Ran, Zhiqin Yang, Zhenyuan Zhang, Wei Xue, Jun Song, Xinmei Tian, and Yike Guo. A control theory of predictability in latent world models. *arXiv preprint arXiv:2607.10362*, 2026. URL https://arxiv.org/abs/2607.10362.

Yi Zhao, Wenshuai Zhao, Rinu Boney, Juho Kannala, and Joni Pajarinen. Simplified temporal consistency reinforcement learning. In *International Conference on Machine Learning*, 2023. URL https://proceedings.mlr.press/v202/zhao23k.html.

Gaoyue Zhou, Hengkai Pan, Yann LeCun, and Lerrel Pinto. DINO-WM: World models on pre-trained visual features enable zero-shot planning. *arXiv preprint arXiv:2411.04983*, 2024. URL https://arxiv.org/abs/2411.04983.

A Stronger-measurement confirmation and training study

This appendix documents the stronger-measurement PushT A/B/C experiments on 256-recipient banks. The original-readout experiments and earlier confirmations are preserved separately in Appendix D and the subsequent method appendices. Their populations, readouts, matching families, and confidence families must not be combined.

Core objective	Free	Actual	Donor	Full reset
Latent	6094.2	5771.6	6509.6	5224.0
Coordinate teacher	8448.1	7662.0	8305.8	6505.9
Physical labels	8583.8	7829.3	8381.0	6496.6

Table 2: Complete core terminal block-position MSE across all six groups, in pixels squared. Full reset changes the insertion readout and standardized displacement norm.

A.1 Complete prespecified contrasts

Table 3 contains all nine PushT A/B primary contrasts. A01–A05 use six model groups and four reference streams; A06–A07 use the pool 0's two architectures and four reference streams; B01–B02 use one native checkpoint and one reference stream on a separate bank. All use 256 recipients. The corrected error in the dual rows is scored with g_A. In Table 4, each training contrast averages three fixed Transformer pools and four reference streams. Every difference is left minus right. Lower error is better; a smaller G is interpreted only jointly with prediction quality.

ID	Comparison	Left mean	Right mean	Change [interval]
A01	Coordinate teacher: actual–free	7662.0	8448.1	-786.1 $[-998.8, -609.0]$
A02	Physical labels: actual–free	7829.3	8583.8	-754.5 $[-926.0, -607.8]$
A03	Latent: actual–donor	5771.6	6509.6	-737.9 $[-938.8, -546.9]$
A04	Coordinate teacher: actual–donor	7662.0	8305.8	-643.8 $[-811.9, -488.3]$
A05	Physical labels: actual–donor	7829.3	8381.0	-551.7 $[-699.6, -413.5]$
A06	Coordinate teacher: dual actual–free	7874.1	8361.2	-487.2 $[-667.6, -325.9]$
A07	Physical labels: dual actual–free	8722.5	9165.3	-442.8 $[-601.5, -293.1]$
B01	Native: actual–free	1439.4	1522.2	-82.9 $[-261.9, 63.8]$
B02	Native: actual–donor	1439.4	3965.8	-2526.5 $[-4729.4, -1035.5]$

Table 3: All nine PushT A/B primary contrasts, in pixels squared. Intervals have nominal 99.444% coverage (nine-contrast Bonferroni allocation).

ID	Comparison	Left mean	Right mean	Change [interval]
C01	Coordinate teacher: T1–T0	5635.6	7325.3	-1689.7 $[-2237.6, -1217.3]$
C02	Coordinate teacher: T2–T1	5603.7	5635.6	-31.8 $[-337.0, 284.0]$
C03	Coordinate teacher G: T1–T0	328.2	546.9	-218.7 $[-397.3, -43.6]$
C04	Coordinate teacher G: T2–T1	383.3	328.2	55.1 $[-89.9, 208.9]$
C05	Physical labels: T1–T0	5250.2	7603.0	-2352.8 $[-2978.2, -1802.5]$
C06	Physical labels: T2–T1	5120.8	5250.2	-129.4 $[-427.7, 165.2]$
C07	Physical labels G: T1–T0	364.2	661.8	-297.6 $[-467.9, -138.3]$
C08	Physical labels G: T2–T1	315.0	364.2	-49.2 $[-193.6, 86.8]$

Table 4: All eight prespecified training contrasts evaluated by g_A. Rows without G measure free-rollout terminal block-position MSE; G rows measure free-minus-actual error. Intervals have nominal 99.375% coverage.

All individual model points are reported below. They are descriptive components of the aggregate estimand, not independent samples for its confidence interval. Pool indices are zero-based; group order is pool 0 Transformer/GRU, pool 1 Transformer/GRU, pool 2 Transformer/GRU.

Table 5: Descriptive group-level values for the PushT pose and training contrasts. Left/right definitions follow Tables 3 and 4; Change is left minus right.

ID	Pool/architecture	Left	Right	Change
A01	0/Transformer	5938.8	6493.7	-555.0
A01	0/GRU	9391.8	10228.7	-836.9
A01	1/Transformer	7791.3	8546.9	-755.6
A01	1/GRU	8143.6	9330.6	-1187.0
A01	2/Transformer	6386.5	6818.6	-432.2
A01	2/GRU	8320.0	9270.0	-950.0
A02	0/Transformer	6145.8	6829.1	-683.3
A02	0/GRU	10617.1	11501.5	-884.4
A02	1/Transformer	7918.2	8668.9	-750.7
A02	1/GRU	8043.2	8722.4	-679.3
A02	2/Transformer	6402.0	6953.8	-551.8
A02	2/GRU	7849.6	8827.3	-977.7
A03	0/Transformer	5803.2	6495.4	-692.2
A03	0/GRU	5201.2	5588.9	-387.7
A03	1/Transformer	7390.3	8597.0	-1206.7
A03	1/GRU	5231.4	5813.2	-581.7
A03	2/Transformer	6290.9	7422.0	-1131.2
A03	2/GRU	4712.9	5141.0	-428.0
A04	0/Transformer	5938.8	6639.4	-700.6
A04	0/GRU	9391.8	9954.8	-562.9
A04	1/Transformer	7791.3	8410.6	-619.3
A04	1/GRU	8143.6	8775.9	-632.4
A04	2/Transformer	6386.5	7074.5	-688.0
A04	2/GRU	8320.0	8979.8	-659.8
A05	0/Transformer	6145.8	6784.0	-638.2
A05	0/GRU	10617.1	11053.9	-436.8
A05	1/Transformer	7918.2	8611.5	-693.4
A05	1/GRU	8043.2	8464.4	-421.2
A05	2/Transformer	6402.0	6938.2	-536.2
A05	2/GRU	7849.6	8433.9	-584.3
A06	0/Transformer	6073.1	6493.7	-420.6
A06	0/GRU	9675.0	10228.7	-553.7
A07	0/Transformer	6409.4	6829.1	-419.7
A07	0/GRU	11035.6	11501.5	-465.9
C01	0/Transformer	5074.5	6651.4	-1577.0
C01	1/Transformer	6128.6	8559.1	-2430.5
C01	2/Transformer	5703.7	6765.3	-1061.6
C02	0/Transformer	5003.2	5074.5	-71.3
C02	1/Transformer	6407.0	6128.6	278.4
C02	2/Transformer	5401.0	5703.7	-302.7
C03	0/Transformer	279.0	592.6	-313.6
C03	1/Transformer	402.0	668.8	-266.8
C03	2/Transformer	303.6	379.3	-75.7
C04	0/Transformer	298.9	279.0	19.9
C04	1/Transformer	529.1	402.0	127.1
C04	2/Transformer	321.9	303.6	18.3
C05	0/Transformer	4649.4	6933.5	-2284.1
C05	1/Transformer	5915.2	8772.6	-2857.4
C05	2/Transformer	5186.1	7102.9	-1916.8
C06	0/Transformer	4569.8	4649.4	-79.7
C06	1/Transformer	5936.7	5915.2	21.5
C06	2/Transformer	4856.0	5186.1	-330.1
C07	0/Transformer	289.3	708.5	-419.2
C07	1/Transformer	348.8	738.3	-389.5
C07	2/Transformer	454.4	538.5	-84.0
C08	0/Transformer	253.4	289.3	-35.9
C08	1/Transformer	362.5	348.8	13.7
C08	2/Transformer	329.2	454.4	-125.2

Table 6 reports the distribution of paired recipient changes after averaging over each contrast's fixed roster. The last column is the largest absolute recipient change divided by 256, its signed contribution to the aggregate mean. No recipient is trimmed or winsorized. The evidence package includes all recipient cells, all model points, the five largest contributions per contrast, and the exact bootstrap index arrays. Percentile intervals are nominal rather than exact finite-sample coverage guarantees.

ID	Q25	Median	Q75	Largest recipient	Mean contribution
A01	-1199.7	-518.7	-216.1	211	-37.7
A02	-1109.8	-569.7	-228.3	211	-28.5
A03	-1276.0	-474.5	-25.2	202	-24.3
A04	-1058.9	-450.5	-67.2	184	-18.3
A05	-834.5	-383.2	-72.5	14	-16.9
A06	-799.2	-312.4	34.1	211	-31.4
A07	-886.5	-295.1	62.6	211	-25.2
B01	-111.9	-6.7	79.1	24	-32.3
B02	-860.9	-162.1	42.8	220	-424.9
C01	-2812.5	-1351.5	15.6	184	-103.5
C02	-1021.4	-60.4	859.8	72	25.0
C03	-657.6	-129.4	321.5	184	-20.4
C04	-326.5	2.3	400.9	72	24.3
C05	-3745.1	-1594.3	-224.5	184	-95.9
C06	-991.8	-56.9	722.1	4	-28.7
C07	-583.8	-152.8	195.0	211	-22.8
C08	-308.9	-27.9	285.9	22	-15.0

Table 6: Recipient-change quartiles and largest absolute influence for each of the 17 A/B/C contrasts. Values are pixels squared; the recipient index refers to that contrast's own bank.

A.2 Readout training and quality

The fixed readout grid uses Adam with learning rate 10^{-4}, no weight decay, batch size 256, and 3,000 updates; validation is recorded every 250 updates. Expert and planner domains have equal training sampling weight. Nested frame budgets are 5k, 20k, and the available unique-frame maximum. The maxima for groups 0–5 are 32,866, 32,846, 33,984, 33,966, 32,730, and 32,712. The encoder, observation projector, and original latent/output normalizers are fixed. We minimize six-output normalized MSE; orientation has no additional decision-cost weighting.

For seed A, the best validation score among quality-qualified checkpoints determines g_A; ties favor smaller width, fewer frames, and earlier checkpoints. The second seed uses that selected width and data budget; its checkpoint is selected by validation alone. The second readout is not part of coordinate continuation. Complete metrics include position MSE and RMS distance, wrapped-angle MAE/MSE, individual sine/cosine MSE, and circle-norm quantiles. They and the 72 complete learning histories are in the accompanying evidence files; the compact table below gives the four gate quantities in both domains. Six-output MSE is normalized, position MSE is squared Euclidean distance in pixels squared, and angle MSE is radians squared.

Table 7: Validation errors for original and selected pose readouts in both domains. The asterisk marks the retained failed g_B quality gate; units and qualification rules are given in the text.

Group	Domain	Readout	Six MSE	Block MSE	Agent MSE	Angle MSE
0/Transformer	expert	Original	0.535	1667.1	12298.3	0.748
0/Transformer	expert	g_A	0.377	1167.9	8320.2	0.546
0/Transformer	expert	g_B	0.397	1178.7	8366.4	0.623
0/Transformer	planner	Original	1.116	3995.6	12476.2	2.633
0/Transformer	planner	g_A	0.784	2599.9	8325.4	1.952
0/Transformer	planner	g_B	0.780	2551.6	8101.9	2.046
0/GRU	expert	Original	0.491	1454.2	12056.4	0.624
0/GRU	expert	g_A	0.385	820.1	8723.2	0.578

Group	Domain	Readout	Six MSE	Block MSE	Agent MSE	Angle MSE
0/GRU	expert	$g_B{}^*$	0.416	886.4	9175.6	0.697
0/GRU	planner	Original	1.197	4970.4	18294.3	2.507
0/GRU	planner	g_A	0.845	2782.6	12603.5	1.848
0/GRU	planner	$g_B{}^*$	0.867	2934.2	13098.8	2.013
1/Transformer	expert	Original	0.572	2140.0	12799.0	0.773
1/Transformer	expert	g_A	0.406	1296.0	8973.7	0.586
1/Transformer	expert	g_B	0.431	1413.3	9215.7	0.524
1/Transformer	planner	Original	1.152	4728.6	12548.2	2.711
1/Transformer	planner	g_A	0.825	2869.7	9770.9	1.639
1/Transformer	planner	g_B	0.801	2809.1	9595.5	1.667
1/GRU	expert	Original	0.476	1257.8	11919.4	0.563
1/GRU	expert	g_A	0.341	797.2	7987.1	0.426
1/GRU	expert	g_B	0.357	799.4	8400.1	0.391
1/GRU	planner	Original	1.226	4203.8	16721.9	2.590
1/GRU	planner	g_A	0.856	2460.1	13020.7	1.658
1/GRU	planner	g_B	0.866	2310.3	12874.4	1.723
2/Transformer	expert	Original	0.529	1276.1	13385.5	0.730
2/Transformer	expert	g_A	0.361	1015.2	9740.3	0.381
2/Transformer	expert	g_B	0.360	929.0	9548.9	0.368
2/Transformer	planner	Original	1.152	4336.9	13782.4	2.224
2/Transformer	planner	g_A	0.672	1852.6	9623.8	1.256
2/Transformer	planner	g_B	0.684	1946.1	9402.6	1.251
2/GRU	expert	Original	0.424	994.0	11387.5	0.489
2/GRU	expert	g_A	0.320	546.7	8687.5	0.400
2/GRU	expert	g_B	0.333	514.8	8700.4	0.497
2/GRU	planner	Original	1.071	3506.2	15928.9	2.262
2/GRU	planner	g_A	0.724	1364.0	12723.9	1.276
2/GRU	planner	g_B	0.769	1380.9	12840.7	1.446

The asterisk marks the failed g_B quality gate in pool 0 GRU. Expert-angle MSE is 0.6972 versus 0.6236 for the original readout, a ratio of 1.11795. Its expert and planner block errors improve, but this does not override the angle gate. Both domain rows are marked because the gate applies to the whole readout. Every selected g_A passes all gate components.

A.3 Data roles and isolation

The frame audit found 8, 16, and 8 exact expert train/validation image overlaps in the three historical pools despite distinct episodes. Before any new readout fit, these images were removed from the new training side only. All validation examples were retained; the original encoders, readouts, normalizers, and old evidence were not rewritten. Consequently, historical encoder exposure to these images remains a limitation, even though the new readout fitting avoids their direct reuse.

The new A/C and B recipient banks and their separate donor banks each contain 256 admitted cases, following independent development banks of 64. All predetermined seed attempts and model-independent admission decisions are retained. Confirmation generation, action-source rules, and donor permutations were fixed before study-model outcomes. A/C uses the original four source-policy types for each group; all branches share their archived actions. The new bank audits find zero overlap under the checked parent, exact content, state-path, and readout-input image comparisons. External DINO-WM training metadata and the inspected initial/first-36-frame state paths were checked, but all upstream pretraining images and every possible internal training window are not available for a complete lineage proof. These generated banks are not labeled out-of-distribution on the basis of new seeds alone.

The scientific design and final protocol were frozen separately, before input construction and before study-model evaluation, respectively. The final protocol binds 25,823 input files and 487 evaluator source files. All expected shards were independently accepted before any main-effect statistics were opened; the frozen evaluator hashes remain unchanged. The release identifies these records by cryptographic hashes and supplies an anonymized protocol summary without machine-specific paths.

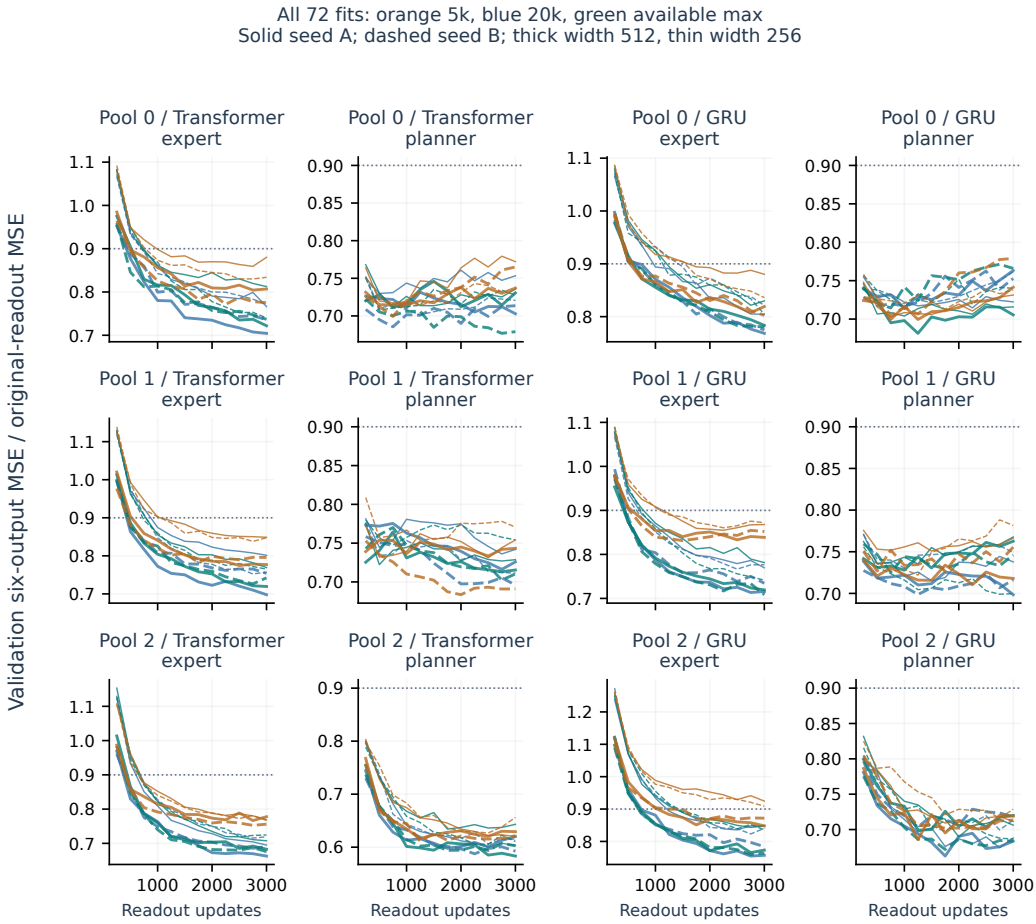

Figure 5: All 72 readout fits, with expert and planner validation shown separately. Curves divide six-output MSE by the corresponding original-readout MSE. Colors indicate nested frame budgets, line widths indicate capacity, and line styles indicate seed. The dotted 0.9 line marks the six-output portion of the quality gate; passing this line alone is insufficient. The complete machine-readable histories retain the other gate metrics as well.

A.4 NUMERICAL FEASIBILITY AND NORM MATCHING

Projection uses FP64 OSQP with absolute/relative tolerance 10^{-9}, at most 20,000 iterations and polishing. An SVD row-space basis with relative cutoff 10^{-10} represents the readout equalities. Activation signs are taken from the original token, including the specified active convention at zero. Constraints for both readouts are stacked for dual projection. Full corrections are retained before scaling, so matching is not confused with resolving a different optimization problem.

Every final FP32 insertion is independently tested using the full FP64 readout, inequalities for the original ReLU activation region, and common standardized displacement norm. Accepted normalized readout and region residuals are at most 10^{-6}; norm agreement uses absolute and relative tolerance 10^{-6}. The frozen numerical backoff policy can shrink an entire family by powers of two down to $1/256$, with complete revalidation; failures remain failures and cannot become zero-displacement cases. No confirmation family required backoff. Table 8 accounts for every matching family.

Family	Count	Mean	Min	Max	Zero	Backoff
A core	6144	3.289	1.005	7.857	0	0
A dual	2048	2.480	0.952	4.850	0	0
B core	256	130.327	74.800	467.449	0	0
C coordinate teacher	3072	3.193	1.212	8.267	0	0
C physical labels	3072	3.187	1.224	8.281	0	0

Table 8: Complete matching-family accounting and standardized displacement norms. Native norms are in a different dimension and normalization, so their sizes are not cross-model effect comparators. All families pass the original numerical constraints.

The native migration test distinguished solver reproducibility from transition reproducibility. Independently resolved QPs on one development case differed by 1.51×10^{-10} in full-direction coordinates and by at most 2.38×10^{-7} after FP32 insertion; both passed the unchanged constraints. With the *same* inserted FP32 token, native corrected rollout was bitwise identical. Thus we do not claim independent QP solves are bitwise identical.

A.5 DUAL-READOUT SECONDARY MEASUREMENTS

The following comparisons use a single twelve-direction matching family and the same g_A endpoint scoring. Their 95% intervals are exploratory, outside the nine primary slots. Positive dual-minus-single changes indicate a loss of correction benefit from adding the second constraint, which can coexist with the dual benefit over free rollout.

Objective	Comparison	Change [95% interval]
Coordinate teacher	dual actual–dual donor	-342.5 $[-450.6, -235.7]$
Coordinate teacher	dual actual–single actual	33.5 $[-26.4, 93.3]$
Coordinate teacher	single actual–free	-520.7 $[-652.3, -398.5]$
Latent	dual actual–dual donor	-192.5 $[-316.7, -71.3]$
Latent	dual actual–single actual	104.3 $[50.3, 159.9]$
Latent	single actual–free	-241.5 $[-346.7, -138.7]$
Physical labels	dual actual–dual donor	-283.7 $[-397.7, -174.6]$
Physical labels	dual actual–single actual	145.3 $[88.9, 205.6]$
Physical labels	single actual–free	-588.1 $[-705.7, -476.3]$

Table 9: Secondary comparisons within the matched single/dual experiment.

A.6 NATIVE MODEL: INTERFACE, READOUT, AND HORIZONS

The official source is fixed at commit `0a9492fa12044b852ae9e001cc74604b79c8bb0c`. The released PushT checkpoint's epoch field is 2; a configuration requesting 100 epochs does not establish that this payload completed them. DINOv2 source is fixed at `7764ea0f912e53c92e82eb78a2a1631e92725fc8`. The evidence package records the

Objective	Branch	Free	Corrected	Change
Latent	single actual	5784.0	5510.6	-273.3
Latent	dual actual	5784.0	5617.0	-166.9
Coordinate teacher	single actual	9020.1	8439.9	-580.2
Coordinate teacher	dual actual	9020.1	8504.5	-515.6
Physical labels	single actual	9983.0	9453.2	-529.8
Physical labels	dual actual	9983.0	9571.0	-412.0

Table 10: Descriptive endpoint scoring by g_B for the same twelve-direction experiment. These means do not replace the primary g_A measurement.

model and pretrained visual-weight hashes and official acquisition links, without redistributing external weights.

Raw images are divided by 255, passed through the official 224-pixel resize/crop and mean/scale 0.5 normalization, and resized to 196 inside the native model. Its representation has 196 visual patches of width 384 plus proprioceptive and known-action components of width 10 each per patch. The diagnostic flattens only the 75,264 visual coordinates for projection, then restores the exact patch layout. It uses the full visual space; no reduced search basis is fitted. Proprioceptive and action components are not projection degrees of freedom.

Three historical observations and five primitive actions per block are retained. Official rollout writes known future actions into the first four future tokens, then predicts its final extra token. The insertion is prediction index 3 and the scored endpoint index 7, corresponding to the first and fifth future blocks. The terminal predicted action entries are not overwritten by an invented future action. Identity replacement, native free-rollout parity, fixed action/proprioceptive insertion, and visual flatten/unflatten are checked explicitly.

The visual readout has architecture $75264 \to 256 \to 256 \to 6$. Its normalizers are fitted only on training encodings. A predetermined 2,000 training episodes with ten sampled frames each yield 19,721 unique training frames after removing 279 internal duplicates; all 2,514 validation frames are retained. There are no exact train/validation model-input image matches in this readout dataset. The fixed 3,000-update fit selects step 2,500 by validation, with block RMS distance 16.544 pixels. The native readout is a diagnostic measurement, not a replacement state input to the dynamics. The fixed action source uses ten approach controls to establish the observed history, followed by five uniformly sampled held-action blocks (25 primitive controls). Visibility admission uses only the historical prefix; all future outcomes are retained. This is a native-interface diagnostic, not a reproduction of the official planning benchmark.

Action	Free	Actual	Donor	Observed encoding	Visual reset
5	781.0	781.0	781.0	673.8	673.8
10	996.2	939.8	3291.2	725.6	912.0
15	1182.2	1118.3	3563.9	866.0	1103.5
20	1388.9	1278.2	3960.4	892.0	1268.3
25	1522.2	1439.4	3965.8	979.7	1427.4

Table 11: Complete native horizon means on 256 recipients, in pixels squared. Observed encoding decodes the actual image at each horizon; visual reset replaces only the first predicted visual component and then rolls freely. Actual and donor preserve the action-5 readout within tolerance. These are descriptive curves; action 25 remains the primary endpoint.

A.7 Training losses and joint diagnostics

Each example contains observations at $-10, -5, 0, 5, 10, 15$ and the five aligned action blocks. All three training conditions evaluate the final successor of each of three rolling transitions. For

coordinate-teacher or physical-label target q_{ij} in normalized six-output coordinates,

$$\mathcal{L}_{\text{coord}} = \frac{1}{3B \cdot 6} \sum_{i=1}^{B} \sum_{j=1}^{3} \|\tilde{g}_A(\hat{z}_{ij}) - q_{ij}\|_2^2, \tag{4}$$

$$\mathcal{L}_{\text{anchor}} = \frac{1}{3B \cdot 192} \sum_{i,j} \|(\hat{z}_{ij} - E(o_{ij})) \oslash \sigma_z\|_2^2. \tag{5}$$

T0 updates each history with observed encodings. T1 and T2 append predictions without detaching between transitions. T2 minimizes $\mathcal{L}_{\text{coord}} + 0.1\mathcal{L}_{\text{anchor}}$; this standardized anchor differs from the original raw-latent continuation loss. Targets are detached, but the frozen readout retains gradients with respect to predicted inputs. The summed objective receives the original complete-gradient normalization, not separate normalization of its terms.

All conditions use AdamW, learning rate 10^{-5}, weight decay 10^{-3}, batch 128, and 2,100 additional updates from the corresponding A coordinate checkpoint. A shared λ is chosen from $\{0.1, 1.0\}$ on pool 0 validation: the mean T0-normalized endpoint criterion is 0.86158 for 0.1 versus 0.96315 for 1.0. No intervention result participates. Profiling gives approximately 27.6–27.9 ms/update, maximum/minimum ratio 1.0073; this is measured compute similarity, not a claim of exactly equal FLOPs.

Objective	Train	Free	Observed-history prediction	G	Response	Variance range
Coordinate teacher	T0	7325.3	3504.9	546.9	0.381	1.024–1.039
Coordinate teacher	T1	5635.6	3539.6	328.2	0.309	0.851–0.885
Coordinate teacher	T2	5603.7	3521.6	383.3	0.337	0.563–0.602
Physical labels	T0	7603.0	3239.1	661.8	0.338	1.102–1.141
Physical labels	T1	5250.2	3181.9	364.2	0.235	0.876–0.935
Physical labels	T2	5120.8	3215.3	315.0	0.316	0.575–0.617

Table 12: Joint descriptive training diagnostics under g_A. Free-rollout MSE, observed-history prediction MSE, and G are in pixels squared. Response is mean latent MSE after the predetermined negated-action perturbation; variance range spans the three pools' terminal mean coordinate variances across recipients in standardized latent space. These magnitudes change and are not claimed invariant.

Observed-encoding terminal block-position MSE equals 3683.5045 in every row because perception, readout, recipients, and actions are shared. Observed-history prediction is a different quantity: it predicts the endpoint using observed earlier inputs. Mean action response is nonzero, although some individual response values are zero. Together with improved free rollout and nonconstant outputs, these checks argue against interpreting smaller G in isolation as a model improvement; they do not prove preservation of every behavioral sensitivity.

Under g_A, remaining correctability G for T1 is 328.2 [225.9,431.7] for the coordinate-teacher objective and 364.2 [260.6,475.8] for the physical-label objective under exploratory 95% intervals. The primary training conclusions use the eight differences in Table 4. Full corrections, shown below, use each condition's own larger standardized displacement norm and are descriptive rather than the between-training comparator.

A.8 Decision decomposition and reproducibility

The exploratory decision analysis retains all 9,216 pools (512 recipients, six groups, three objectives) and four saved branches, with no missing cost pools. Original score and physical-outcome file hashes are verified, and original selected-cost and rank metrics are reconstructed. Equation 3 and pair-margin MSE $= (64/31)V$ are checked separately in every 32-candidate pool. Exact ties retain their original semantics; undefined rank values remain explicitly counted in the evidence files. Pair agreement is exact sign agreement including ties; top-three incident pairs have at least one endpoint in the true top three, including ties at the third cost. Ground-truth ranking only defines offline evaluation subsets.

Objective	Train	Matched standardized displacement norm	Full actual standardized displacement norm	Full G
Coordinate teacher	T0	3.193	3.986	668.4
Coordinate teacher	T1	3.193	3.716	361.4
Coordinate teacher	T2	3.193	3.250	393.5
Physical labels	T0	3.187	4.149	823.4
Physical labels	T1	3.187	3.827	401.6
Physical labels	T2	3.187	3.239	321.8

Table 13: Norm-matched versus full actual-guided correction, evaluated by g_A. The first norm column is identical across T0/T1/T2 for every individual matched case, not only on average. Full G uses each condition's own projected direction.

Objective	Metric	Change [95% interval]
Latent	Cost MSE	-2.5163 $[-3.9336, -1.1347]$
Latent	Shared bias squared	-2.6802 $[-4.0528, -1.3395]$
Latent	Centered cost-error variance	0.1638 $[-0.0906, 0.4401]$
Latent	Pair agreement	0.0007 $[-0.0013, 0.0027]$
Latent	Top-3 incident agreement	0.0001 $[-0.0045, 0.0046]$
Coordinate teacher	Cost MSE	-27.3754 $[-33.4746, -22.1528]$
Coordinate teacher	Shared bias squared	-27.2970 $[-33.3641, -22.0862]$
Coordinate teacher	Centered cost-error variance	-0.0784 $[-0.3585, 0.1995]$
Coordinate teacher	Pair agreement	0.0011 $[-0.0008, 0.0030]$
Coordinate teacher	Top-3 incident agreement	0.0038 $[-0.0002, 0.0079]$
Physical labels	Cost MSE	-24.1493 $[-32.2739, -16.9636]$
Physical labels	Shared bias squared	-24.5944 $[-32.6258, -17.5096]$
Physical labels	Centered cost-error variance	0.4451 $[0.1397, 0.7528]$
Physical labels	Pair agreement	-0.0005 $[-0.0020, 0.0010]$
Physical labels	Top-3 incident agreement	-0.0021 $[-0.0055, 0.0013]$

Table 14: Complete objective-level exploratory decomposition. The first three metrics are in millions of squared cost units; agreement changes are fractions. Intervals are 95% recipient bootstraps, not candidate-pair bootstraps.

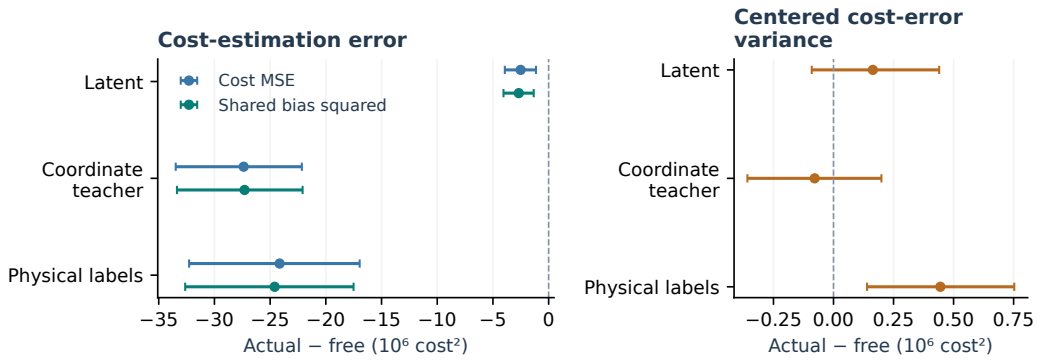

Figure 6: **Most cost-estimation improvement reduces shared bias squared.** Exploratory actual–free changes on the earlier 512-recipient bank, with 95% recipient-bootstrap intervals. All three objectives and all 9,216 candidate pools are retained. Axes are squared cost units, not position-error units. The right panel isolates the much smaller centered cost-error variance V; the two panels use different scales.

An independent NumPy verifier reconstructs all 17 A/B/C contrasts from accepted recipient cells. It regenerates the fixed 20,000 bootstrap draws and uses multinomial-count matrix multiplication, separately from the evaluator's index-gather implementation, to check means, intervals, relative ratios, and influential recipients. This establishes numerical reconstruction of the reported estimands; it does not replace the raw-token acceptance checks.

The companion end-to-end package supplies three self-trained model checkpoints, both readouts, all required normalizers, raw images/actions, and actual/donor assignments for a predetermined two-case development example. It creates new projections and free/corrected rollouts with six- and twelve-direction matching, rather than reading saved prediction means. A clean Python environment runs all four example configurations and 36 QPs; raw encodings and rollout tensors match their development fixtures. Independent NumPy checks also reject a deliberately corrupted inserted token even after its file checksum is updated. The new confirmation evidence package provides complete recipient arrays and statistics; the earlier evidence package preserves the original confirmations. These packages have distinct scopes and do not claim to regenerate the entire training corpus from the small demonstration.

B Prospective tests of measurement, diagnosis, and decisions

The following experiments extend the same feedback intervention. Their model rosters, data roles, endpoints and primary contrasts were fixed before the corresponding confirmation outcomes. Each test retains its own population and statistical family. They are not additional draws from the earlier confirmation banks.

B.1 An additional pose constraint with a reserved evaluator

The earlier evaluation-only analysis finds that preserving g_A does not preserve every fitted pose estimate. We therefore fit two new readouts per Pool 0 architecture. The additional constraint g_C is a two-hidden-layer ReLU network with width 512; the reserved evaluator g_D is a width-256 GELU network with two residual blocks, using the same architecture as the earlier evaluation-only readouts but separately fitted weights and normalizers. Both output the same six pose coordinates. Neither enters predictor training. Only g_C enters the additional intervention constraint; g_D never determines a displacement, a matching norm, or a predictor checkpoint.

The accepted measurement inputs are reused, with disjoint physical-parent assignments to the C and D roles within the existing fitting, validation and qualification roles. Each readout receives 512 expert and 512 planner fitting parents; validation uses 128 expert and 256 planner parents, and qualification uses the same respective counts on different parents. Assignments are shared across the

Table 15: A fresh test with an additional pose constraint and a reserved evaluator. Positive primary contrasts favor actual guidance. Both four-contrast families use 98.75% recipient-bootstrap intervals. Units are pixels squared; insertion rows concern squared output displacement, not prediction error.

Objective	Contrast	Estimate	98.75% interval
Coordinate teacher	U_{A+C}	676.6	$[398.2, 960.9]$
Coordinate teacher	S_{A+C}	733.0	$[459.8, 1033.4]$
Physical labels	U_{A+C}	670.7	$[421.6, 922.5]$
Physical labels	S_{A+C}	371.6	$[149.5, 601.4]$
Coordinate teacher	$U_A - U_{A+C}$	194.1	$[10.8, 383.6]$
Coordinate teacher	Insertion displacement reduction	65.6	$[6.2, 127.6]$
Physical labels	$U_A - U_{A+C}$	161.2	$[32.3, 294.0]$
Physical labels	Insertion displacement reduction	13.2	$[-45.3, 71.6]$

two architectures. Cross-role retained-image aliases are excluded by the fixed input gate; this is not a claim about external encoder pretraining. Input and target normalizers use the equally weighted expert/planner fitting mixture, with a standard-deviation floor of 10^{-6}.

All four fits use Adam for 8,000 updates, learning rate 10^{-4}, batch size 256 with equal domain counts, and no weight decay. Validation occurs every 250 updates; the lowest balanced six-coordinate normalized validation MSE selects the checkpoint, breaking ties in favor of the earlier update. All four selected checkpoints are locked before qualification. Each must stay within 1.1 times the corresponding g_A error on six-coordinate normalized MSE, block-position MSE, agent-position MSE and wrapped-angle MSE in both domains. All four qualify; the original qualification failures in other experiments remain separate.

The confirmation uses 256 fresh recipients and 256 separate donors, both Pool 0 architectures, the two coordinate objectives, and all four original reference streams. The insertion and terminal times remain primitive actions 5 and 25. For each architecture, recipient and stream, the two constraint sets $\{g_A\}$ and $\{g_A, g_C\}$, two objectives and two sources form one eight-direction matching family. Displacements use the g_A input-standardized metric. Each readout retains its own fixed normalizers. The common minimum projected norm is followed by the first accepted common numerical backoff after FP32 insertion from $1, 2^{-1}, \ldots, 2^{-8}$. Full nonlinear outputs and activation regions are checked in FP64 at tolerance 10^{-6} before scoring with g_D. All confirmation families pass; the largest checked output deviation is 2.5921×10^{-7} and no family has zero norm.

Let L_k^b be terminal block-position squared error measured by g_D, with constraint set $k \in \{A, A{+}C\}$ and branch b. The four primary contrasts are

$$U_{A+C} = \mathbb{E}[L^{\text{free}} - L^{\text{act}}_{A+C}], \qquad S_{A+C} = \mathbb{E}[L^{\text{don}}_{A+C} - L^{\text{act}}_{A+C}]$$

for the two coordinate objectives. Each recipient first averages both fixed architectures and four streams; 20,000 recipient-bootstrap draws then yield 98.75% intervals for the four-contrast family. Both contrasts are positive for each objective (Table 15). This establishes useful, source-sensitive correction under two fixed pose constraints when a separately reserved readout evaluates the future. The pooled result is not a separate confirmation in each architecture or stream.

The four prespecified secondary contrasts compare the two constraint sets at the same eight-way matched norm. Two are $U_A - U_{A+C}$; the other two measure the reduction in squared change of g_D's block-position output at insertion. The latter is output displacement, not insertion prediction error. This second four-contrast family also uses 98.75% intervals. The extra constraint attenuates terminal correction benefit for both objectives. Insertion output displacement is attenuated for the coordinate-teacher objective, while the physical-label interval crosses zero. Constraint changes also change the available projection directions, so these contrasts do not isolate a physical mediator.

B.2 Incremental diagnosis on fresh contexts for fixed model pairs

This test asks whether a source-sensitive feedback diagnostic predicts the response to recursive training beyond ordinary prediction errors. The six T0/T1 pairs are the already trained, fixed final checkpoints from the three Transformer pools and two coordinate objectives. There is no new predictor

training or checkpoint selection. Calibration uses 256 fresh recipients and 256 separate donors; testing uses another 512 recipients and 512 donors. Probe and response streams use the original random/full and CEM/full reference policies, respectively, with native batches of 300 candidates.

Only T0 enters the probe. At action 5, its two objectives and actual/donor sources share a four-direction g_A-preserving matching family. The existing pool-specific g_{eval} scores the probe and response; it is distinct from the reserved g_D above. For each recipient and pool/objective stratum, the signed probe feature is

$$s_{\text{probe}} = L^{\text{don}}_{\text{probe},25} - L^{\text{act}}_{\text{probe},25}.$$

The response is $Y = L^{T0,\text{free}}_{\text{response},25} - L^{T1,\text{free}}_{\text{response},25}$. Both use block-position squared Euclidean error without division by two. This feature differs from the earlier population quantity $G = \mathbb{E}[L^{\text{free}} - L^{\text{act}}]$; the implementation name `signed_g` refers to s_{probe} here.

The baseline uses T0 free-rollout errors at actions 5 and 25, observed-history error at action 25, and squared common displacement norm. Within-stratum calibration-only standardization precedes a full quadratic basis: four linear terms, four squares and six pairwise products. These basis columns are standardized again using calibration moments. Six unpenalized stratum intercepts account for the fixed pool/objective pairs, with slopes shared across strata. The response is scaled by one calibration RMS around the six stratum means. Ridge regularization is fixed at 0.01. The augmented predictor adds the calibration-standardized s_{probe} to the same design. Both test prediction vectors are fixed before computing the T0/T1 test-response stream.

The primary outcome is baseline minus augmented response-prediction MSE. Strata are averaged within each of the 512 test recipients before 20,000 paired bootstrap draws and a 95% percentile interval. Its units are pixels to the fourth power because Y is already a squared-position-error difference. This is an offline test on fresh contexts for known model pairs; it does not predict the outcome of training an unseen model or supply an observation-free online diagnostic.

Complete result. All 512 test recipients and all six fixed strata are retained. Baseline response-prediction MSE is 464391921.23 and augmented MSE is 463922578.82 pixels to the fourth power. Their paired reduction is 469342.41, with 95% interval $[-222573.08, 1649895.08]$. The descriptive reduction relative to baseline is 0.1011%. This does not confirm incremental predictive benefit and does not establish equivalence. Table 16 retains every stratum's descriptive change. The aggregate interval resamples recipients with their six strata, without refitting calibration coefficients or changing donors. Independent arithmetic reproduces calibration fitting, fixed test predictions and the full bootstrap statistics at the frozen absolute/relative tolerances $10^{-9}/10^{-10}$.

Table 16: **Incremental diagnosis: complete fixed strata.** Baseline minus augmented response-prediction MSE, in millions of pixels to the fourth power. Stratum estimates are descriptive; only the pooled primary contrast has a prespecified interval. Positive favors adding the feedback probe.

	Coordinate teacher	Physical labels
Pool 0	0.1142	−0.2232
Pool 1	−0.1563	0.0705
Pool 2	−0.7289	3.7397
Pooled primary	0.4693 [−0.2226, 1.6499]	

B.3 Matched decision after an executed prefix

The earlier decision experiment uses a different model generation from the recursive-training comparison. The new test uses all six fixed T0/T1 pairs on 512 fresh recipients and 512 separate donors. For each pool and recipient, the unchanged original reference policy supplies a native 300-candidate action population. Every candidate suffix is simulated after the same selected five-action prefix and complete physical execution history. The studied T0/T1 models share this proposal set rather than proposing model-specific candidates.

The fixed 32 suffixes consist of the selected proposal's remaining 20 actions, a zero-action suffix, and 30 source candidates sampled without replacement using the original deterministic candidate

rule. Duplicate action sequences and all physical outcomes are retained. Only the recipient has 32 candidate outcomes; the donor supplies its own observed token after its own executed prefix. The final horizon remains primitive action 25. This is one subsequent action choice, not a full closed-loop controller evaluation.

Each studied model predicts its own insertion token at native batch size 300. Its actual and donor branches share that model-specific anchor. The observed image is identical across candidate suffixes. All four model anchors in a pool—two objectives and T0/T1—and two sources form an eight-direction g_A-preserving matching family. A single donor permutation is shared across pools, objectives and conditions. Full observed-token reset is retained as an information reference, outside the matched displacement family. The free/actual/donor/reset branches then roll through the same 32 suffixes using the original ten-action context immediately before insertion. Numerical checks compare repeated native roots and free versus identity replacement on the same 32-candidate path; they do not assume bitwise equality between different batch sizes.

The same fixed pool readout g_A decodes each predicted terminal token and the goal-image token. Predicted cost is block-position squared distance plus 900 times squared wrapped-angle difference. Input validation checks the saved physical alignment before inference; physical outcomes are joined to predicted costs only after all candidate predictions are complete and fixed. Realized physical cost uses the same position/angle formula on simulator terminal and true goal states. Future physical states and true goal coordinates are absent from the candidate-scoring interface.

For an exact tie among predicted minimum-cost candidates, primary selected cost is the average realized physical cost of all tied candidates, equivalent to uniform tie resolution. The two primary contrasts are T1 actual minus free and T1 actual minus donor selected cost. Each recipient averages the three pools and two objectives before 20,000 paired bootstrap draws; the two-contrast family uses 97.5% intervals. Useful, source-sensitive decision improvement requires both upper bounds below zero. No crossing interval is interpreted as equivalence. Changes from T0 to T1, changes in the exact minimizing candidate set, and lowest-index tie resolution are descriptive secondary summaries. All previous decision results retain their original statistical identity.

Complete result. All 512 recipients, three pools, 12 fixed models and 32 candidate suffixes per pool/recipient are retained. Both primary contrasts are unresolved: T1 actual minus free selected physical cost is -50.6644 $[-116.4936, 11.4868]$, and T1 actual minus donor is -34.2978 $[-111.6705, 39.4363]$. These 97.5% intervals do not meet the joint criterion for useful, source-sensitive decision improvement. Table 17 retains every model and branch mean. The descriptive change in free selected cost from T0 to T1 is -146.5338; the change in free-minus-actual benefit is 15.1089. Actual guidance changes T1's exact minimizing candidate set in 45.9961% of pool/objective/recipient cases. These secondary summaries have no prespecified intervals. Lowest-index tie resolution yields the same branch means as uniform exact-tie resolution in this population.

All candidate predictions were fixed before the outcome join. Independent exact-tie selection and count-weighted NumPy resampling reproduce all recipient contrasts and 20,000 bootstrap means at absolute/relative tolerances $10^{-9}/10^{-12}$. This verifies saved predictions and statistics; it does not independently replay the world models or simulator. An initial execution failed before projection because the GPU environment lacked the solver dependency. Installing the pinned solver in an isolated runtime overlay allowed the unchanged frozen stage to complete; the failed records remain retained.

B.4 Complete historical study map

Table 18 keeps the earlier model rosters, readout roles, banks and inferential families separate from the three prospective tests above.

B.5 Computation and regeneration scope

The prospective experiments reuse the fixed encoders, normalizers, action encoders and predictor checkpoints. Only the four new pose readouts in the reserved-evaluator test are fitted; the diagnosis adds two small ridge regressions. No predictor is retrained for these extensions. Scientific settings,

Table 17: **Matched decision: complete selected-cost means.** Every fixed pool, condition, objective and branch is retained. Values are realized physical pose costs under uniform exact-tie selection, averaged over 512 recipients. Rows are descriptive; the pooled two-contrast primary family is reported separately. Full reset is an unmatched information reference.

Pool	Condition	Objective	Free	Actual	Donor	Full reset
0	T0	Coordinate teacher	3447.22	3413.63	3418.71	3331.25
0	T0	Physical labels	3584.72	3435.23	3609.72	3391.71
0	T1	Coordinate teacher	3206.32	3217.61	3327.92	3137.06
0	T1	Physical labels	3419.89	3325.11	3430.43	3286.70
1	T0	Coordinate teacher	3666.90	3613.05	3732.27	3596.83
1	T0	Physical labels	3780.28	3856.04	3831.95	3730.50
1	T1	Coordinate teacher	3618.54	3534.86	3558.13	3513.84
1	T1	Physical labels	3645.55	3628.58	3680.99	3549.73
2	T0	Coordinate teacher	3303.42	3303.47	3327.30	3269.44
2	T0	Physical labels	3479.13	3426.93	3382.47	3310.06
2	T1	Coordinate teacher	3279.59	3207.28	3199.90	3155.04
2	T1	Physical labels	3212.58	3165.05	3086.90	3124.25

Table 18: **Readout roles and evidence banks.** Training denotes coordinate supervision; latent-objective and native world-model training use no diagnostic readout. Pose confirmation and training share 256 recipients and four reference streams; M has 512 fresh recipients; B/P each have 256 separate recipients. Direction counts specify the matching family. † and ‡ mark post-hoc g_{eval} and g_B scoring. Earlier N/D/R use original models/readouts (Appendices L, K, H); D does not evaluate T1. Table 2 retains all core branch means.

Study / roster	Training	Preserved	Scored	Bank / dirs.	Inferential role
Core / 6 Transformer/GRU groups	g_A	g_A	g_A; $g^{\dagger}_{\text{eval}}$	Shared / 6	5 primary; rescoring†
Dual / Pool 0 Transformer/GRU	g_A	g_A, g_B	g_A; $g^{\ddagger}_B$; $g^{\dagger}_{\text{eval}}$	Shared / 12	2 primary; simultaneous preservation
Motion / Pool 0 Transformer/GRU	g_A	g_A; (g_A, h_c); $(g_A, h_{Q,c})$	g_A	M / 12	6 primary mechanism; 6 secondary usefulness
Training / 3 Transformer pools	g_A	g_A	g_A; $g^{\dagger}_{\text{eval}}$	Shared / 6	8 prespecified; 8 post-hoc†
Native PushT / DINO-WM	None	Native pose	Native pose	B / 2	2 primary; usefulness unresolved
Native PointMaze / DINO-WM	None	Native position	Native position	P / 2	2 primary; both unresolved
Nearest donor / 6 earlier groups	Original pose	Original pose	Original pose	N / 12	Earlier source control
Decision / 6 earlier groups	Original pose	Original pose	Pose; native latent score	D / 6	4 primary pose; secondary native score
Reacher / 6 models	Original circle	Circle, joint angles	Circle, joint angles	R / —	Earlier terminal boundary

source identities and data-role assignments are recorded before the corresponding confirmation outcomes; execution receipts retain all cases and failed engineering attempts.

Fresh PushT contexts and physical suffix outcomes are generated on CPUs. Encoding, action proposal and latent rollout use one NVIDIA RTX PRO 6000 GPU per allocated task. Projection and statistics use FP64 CPU arithmetic; world-model dynamics and inserted tokens use FP32. The matched decision execution uses at most three concurrent GPU tasks, each requesting four CPU cores and 24–32 GiB host memory. Its CPU stages request two to eight cores and 8–32 GiB. These are scheduler requests, not claims of measured minimum memory requirements.

The completed reserved-evaluator extension, including its fixed development run, used 0.78 allocated GPU-hours and 7.57 allocated CPU-core-hours. The completed incremental-diagnosis extension used 1.86 allocated GPU-hours and 10.61 allocated CPU-core-hours. The matched decision experiment used 1.72 allocated GPU-hours and 8.38 allocated CPU-core-hours. These are sums of successful allocation duration multiplied by allocated resources, exclude queue time and earlier model training, and are not wall-clock turnaround times. The initial missing-dependency failure in the decision stage consumed another 0.61 GPU-hours and 2.45 CPU-core-hours, reported separately from successful execution. For diagnosis, the complete 512-recipient test-response rollout took 988 seconds on one GPU; final regression and independent statistical verification took 107 seconds on four CPU cores. The three recovered decision inference jobs took 1522, 1539 and 1871 seconds; prediction sealing and final statistics took 26 and 15 seconds, respectively.

Regenerating the new experiments requires the fixed image encoder and predictor checkpoints, all readout weights and normalizers, the original action proposal implementation, and the PushT simulator version and fixed input manifests. The order is fresh-input generation and isolation checks, frozen-model inference, complete numerical acceptance, then the specified recipient-level statistics. Diagnosis additionally fixes calibration preprocessing and both test-prediction vectors before computing the test response. Decisions fix all candidate predictions before joining the physical costs. Recomputing accepted scalar contrasts or redrawing the supplied vector figures needs no new training or simulator run. A complete raw-input reproduction of every historical experiment was not performed for this revision; the reported correctness checks concern the new experiments and their saved evidence.

C From pose constraints to motion constraints

The added experiment asks whether a qualified measurement of motion explains part of the feedback difference left unconstrained by pose. It follows a separate development study, independent measurement test, and fresh intervention confirmation. The earlier bounded qualification failure in Appendix P is retained; its validation data are not relabeled as an independent test.

C.1 A conditional measurement that adds information

An estimate computed only from preserved pose outputs and fixed actions is already invariant under the original intervention. We instead fit a residual velocity readout with access to the full latent history. Its five targets are instantaneous agent v_x, v_y, block v_x, v_y, and block angular velocity, read from the simulator bodies after the last physics substep of each control step. Labels are obtained by exact replay of the original seeds and actions. Three-token contexts are separated by five primitive actions; six context endpoints per trajectory are retained. The latest five two-dimensional actions are known inputs.

Let x_1, x_2, x be the three standardized latent tokens, p the standardized concatenation of their six-output pose readouts, and u the standardized action vector. The standardized velocity estimate is

$$h(x_1, x_2, x, p, u) = B(p, u) + R(x_1, x_2, x, p, u). \tag{6}$$

The residual's own baseline B is frozen before its residual is fitted. It need not be the independently selected strongest pose/action baseline used for the qualification comparison. The baseline has 28 inputs; the full residual has 604. Latent normalization remains that of the original g_A; action, decoded-pose-feature and velocity normalizers use training data only. We retain all five targets throughout.

Development uses the predeclared nested budgets of 1,024, 4,096 and 16,384 independent training initial states, with a shared 1,024-state validation bank. Each architecture uses two fixed policy routes and six windows per initial state. The fitted input families are decoded pose/actions, privileged true pose/actions, direct latent history/actions, the full latent residual, and a residual omitting the latest latent token. A same-time RGB-history model supplies an additional empirical comparison. The full grid contains 120 MLP fits and six RGB fits. The RGB and MLP models have different architectures and compute budgets; this comparison is not an information-theoretic bound or a compute-matched architecture test.

The formal velocity readout is selected only from the 16,384-state full-residual candidates by the recorded validation criterion, rather than replacing it with the strongest input family. Both architectures select width 512 and the same predeclared fit-seed index. Each retains its own baseline and frozen normalizers. A separately selected strongest pose/action model and a separately selected past-latent residual are the measurement controls. Selection precedes the independent test and any new intervention output. Complete development curves and candidate records are included in the evidence archive.

Each MLP has two ReLU hidden layers of width 256 or 512. The two fixed fitting seeds, 12,000 Adam updates, batch size 512 and learning rate 10^{-4} are shared across input families at each data budget. All five standardized targets enter the training loss. Validation is evaluated every 500 updates; selection minimizes mean standardized block-motion MSE, with the earlier update winning an exact tie. Residual output layers are zero-initialized, so update zero recovers the frozen baseline. Across widths and seeds, the smallest validation loss selects the final full residual and each control independently; smaller width then smaller seed resolves ties.

C.2 Independent measurement qualification

The separate test contains 1,024 initial states, two routes and six time windows per architecture: 12,288 correlated contexts per architecture. All five targets must have nonnegative training-mean R^2 relative to the fixed training-mean predictor; the mean over the three block-motion targets must be at least 0.25. The full residual must reduce standardized block-motion MSE by at least 10% against the strongest selected pose/action baseline. Its paired block-loss difference from the selected past-latent residual must additionally have a one-sided 95% bootstrap upper bound below zero. Initial state is the resampling unit for 20,000 draws. These are operational measurement gates, separate from intervention-effect inference.

Table 19: **Independent five-target qualification.** Training-mean R^2 uses the fixed training mean, as in Appendix P. The final column is the ratio of mean standardized block-motion MSE to the strongest selected pose/action baseline. It is not a future-position error ratio. Both groups pass all gates.

	Agent v_x	Agent v_y	Block v_x	Block v_y	Block ω	Loss ratio
Transformer	.999	.998	.708	.759	.595	.780
GRU	.999	.999	.641	.722	.412	.847

The exact block-loss reductions are 22.0287% and 15.2700%. The full-minus-past loss upper bounds are -0.049019 and -0.033317, respectively. The test therefore establishes additional predictive motion information under the qualified distribution; it does not identify a sufficient physical state. Route, time-window and contact strata are all retained descriptively. A full-function FP64 evaluation and an independently implemented feature construction, network evaluation and initial-state bootstrap reconstruct the complete test.

C.3 Joint pose and velocity constraints and the rotated control

During insertion, the earlier tokens, actions and all three pose readouts remain fixed. Collect those fixed quantities into c, and write Equation 6 as $h_c(x)$. For each anchor x_0 and the preselected orthogonal matrix Q, define

$$h_{Q,c}(x) = h_c\big(x_0 + Q(x - x_0)\big), \qquad Q^\top Q = I. \tag{7}$$

The pose features inside c remain their anchor values in this control. Recomputing those features from the rotated token would define a different function. At the anchor, the conditional Jacobians obey $J_{h_Q} = J_h Q$ and therefore have equal full singular spectra. This equality does not extend to their restriction to the pose tangent, nor to their feasible geometry of the original ReLU activation region; both are recorded.

The three conditions preserve g_A alone, (g_A, h_c), or $(g_A, h_{Q,c})$. We fold the fixed arguments into the first-layer bias to solve the quadratic program within the original ReLU activation region, then recompute the complete nonlinear functions on each actual FP32 inserted token using FP64 arithmetic.

The original pose and added velocity output tolerances are 10^{-6} in their normalized coordinates. Region and norm checks also retain their frozen 10^{-6} tolerances. For each architecture, recipient and reference stream, all three conditions, two coordinate objectives and two sources share the minimum standardized displacement norm; any numerical backoff is applied synchronously to all 12 members. The minimum qualifying common norm is 10^{-4}.

The two architectures and two objectives are the original pool 0 Transformer/GRU coordinate continuations. Models, pose readouts and the group-specific rotation matrices are unchanged. A new bank contains 512 recipient initial states and a separate bank 512 donor initial states. Each architecture retains four fixed reference streams. A fixed one-to-one donor permutation is recorded independently for each architecture/reference stream before outcomes. The intervention remains at relative action 5, followed by four free transitions to action 25. No true velocity or later observation enters the predictor.

The independent measurement test uses two latent-policy reference streams per architecture. The formal intervention retains those reference streams and two state-policy reference streams. The frozen analysis therefore also reports insertion-time velocity-measurement errors on observed and predicted tokens for every formal reference stream. These transport checks are descriptive, not a rule for selecting reference streams or changing the primary analysis.

C.4 Populations, numerical gates and inference

For condition $k \in \{\mathrm{pose}, \mathrm{joint}, \mathrm{rotated}\}$, define

$$\Delta_k = \mathbb{E}[L_k^{\mathrm{don}} - L_k^{\mathrm{act}}], \tag{8}$$

where L is the unchanged terminal block-position squared Euclidean error. Positive Δ_k favors actual guidance. For each architecture, the primary contrasts are Δ_{pose}, $\Delta_{\mathrm{pose}} - \Delta_{\mathrm{joint}}$, and $\Delta_{\mathrm{rotated}} - \Delta_{\mathrm{joint}}$. Specific attenuation requires all three adjusted lower bounds above zero. Two objectives and four reference streams are averaged within each recipient before resampling the 512 recipient states. The six contrasts use 20,000 shared bootstrap draws and nominal 99.1667% two-sided percentile intervals. Six $\mathbb{E}[L^{\mathrm{free}} - L_k^{\mathrm{act}}]$ contrasts form a separate secondary family with the same adjustment. These estimates condition on the frozen models, policy rules and realized donor bank.

All 4,096 matching families must pass the full numerical, anchor, effective-rank and nonzero-norm checks before scientific inference. The protocol does not permit complete-case analysis, failure-to-zero imputation, replacement samples or outcome-dependent additions. The complete formal projection passed this gate: all 49,152 actual inserted tokens were independently checked. The original 300-candidate predictor arithmetic is retained, while only selected trajectories are archived. An old-development migration check reconstructs 192 insertions and all subsequent tokens exactly using an independently written step-by-step recurrence.

The velocity and rotated constraints each have rank five on the pose tangent in every checked context; their three block-motion outputs have rank three. Each architecture contributes 2,048 matching families and 6,144 geometry contexts (observed, coordinate-teacher-predicted and physical-label-predicted). Median common standardized displacement norms are 2.956 for the Transformer and 3.800 for the GRU; ranges are $[1.680, 5.975]$ and $[2.315, 6.374]$. No family needs numerical backoff. The full conditional singular spectra agree to 2.67×10^{-15}, while the observed ratios of corresponding pose-tangent singular values span approximately $[0.89, 1.06]$. Thus the rotated control preserves the full spectrum and effective constraint count but is not identical in all local geometry.

A second, local check of every saved FP32 insertion finds maximum normalized pose-readout deviation 2.59×10^{-7}, joint velocity-readout deviation 1.46×10^{-7}, region violation 7.26×10^{-8} and absolute common-norm discrepancy 1.27×10^{-7}. All are within the original tolerances. These checks concern the implemented intervention, separately from the later effect estimates.

Training, validation, measurement-test, recipient and donor roles are checked against the explicitly enumerated earlier evidence. One exact single-frame image occurs in both the new recipient and donor banks, at the final time of distinct physical trajectories. Their seeds, initial states, histories and complete three-frame contexts are distinct. Neither image is an insertion-time guidance frame, and the fixed donor permutation does not pair them. An input-only adjudication retained every sample and the original blocked audit. All checked overlaps between the new banks and training, validation,

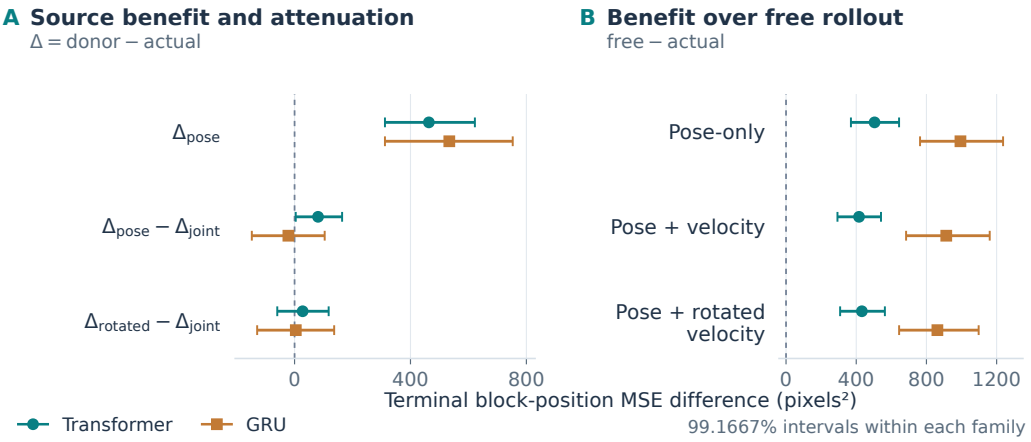

Figure 7: **A richer motion constraint leaves useful feedback corrections.** All six primary (A) and six prespecified secondary (B) contrasts use 99.1667% intervals within each family. Points average two fixed objectives and four reference streams within each of 512 recipients before bootstrapping. Positive Δ favors actual over donor; positive benefit in B favors actual over free rollout. Velocity-specific attenuation requires all three primary contrasts positive within an architecture; the rotated-control comparison is unresolved in both. Pose-only, joint pose and velocity, and rotated-control branches share a nonzero standardized displacement norm within the 12-direction matching family.

measurement-test and prior-reference collections are zero. A separate training–validation single-frame rendering alias is likewise retained and documented. These checks do not establish complete foundation-model pretraining provenance.

C.5 Complete formal comparisons

Figure 7 shows all primary and prespecified secondary contrasts. All 49,152 corrected five-step trajectories and 8,192 free trajectories are present. Independent reconstruction checks every decoded pose and velocity, raw simulator target, and all twelve planned contrast intervals. A second local reconstruction agrees to maximum pose and velocity differences of 5.12×10^{-13} and 8.66×10^{-15}. The first independent-verifier launch failed during module import, before result evaluation; a versioned search-path repair retains every numerical function unchanged. Original failures and both reconstruction records accompany the frozen evidence.

Table 20: **All six primary mechanism contrasts.** Positive values favor source benefit or attenuation in the stated comparison. Neither architecture meets all three conditions for specific attenuation. Units are pixels squared.

Architecture	Contrast	Difference	99.1667% interval
Transformer	Δ_{pose}	463.5	[311.9, 622.3]
Transformer	$\Delta_{\text{pose}} - \Delta_{\text{joint}}$	81.0	[4.0, 164.1]
Transformer	$\Delta_{\text{rotated}} - \Delta_{\text{joint}}$	27.9	[-59.7, 117.7]
GRU	Δ_{pose}	534.0	[311.9, 752.9]
GRU	$\Delta_{\text{pose}} - \Delta_{\text{joint}}$	-21.3	[-147.5, 103.9]
GRU	$\Delta_{\text{rotated}} - \Delta_{\text{joint}}$	4.3	[-129.1, 136.7]

Table 21: **All six prespecified secondary usefulness contrasts.** Benefit is free minus actual-guided terminal error. Every interval excludes zero; this is a distinct family from Table 20. Joint pose and velocity preservation retains a 6.0% reduction for the Transformer and 9.4% for the GRU, calculated from aggregate means.

Architecture	Constraint	Free	Actual	Benefit	99.1667% interval
Transformer	Pose-only	6908.6	6404.2	504.4	[369.8, 644.1]
Transformer	Pose + velocity	6908.6	6492.6	416.0	[293.1, 541.1]
Transformer	Pose + rotated velocity	6908.6	6476.4	432.1	[308.2, 563.2]
GRU	Pose-only	9667.3	8673.6	993.8	[763.7, 1236.7]
GRU	Pose + velocity	9667.3	8754.3	913.0	[684.1, 1160.6]
GRU	Pose + rotated velocity	9667.3	8804.9	862.5	[644.4, 1097.6]

Table 22: **Complete descriptive horizon means.** Every condition and source, averaged over recipients, objectives and reference streams. At action 5 all branches preserve the same complete pose readout; later errors reflect free rollout. Only action 25 is the primary endpoint.

Architecture	Branch	Action 5	10	15	20	25
Transformer	Free	2190.9	2795.8	3801.3	5210.7	6908.6
Transformer	Pose-only / actual	2190.9	2574.9	3464.8	4779.5	6404.2
Transformer	Pose-only / donor	2190.9	2802.2	3793.8	5191.2	6867.6
Transformer	Pose + velocity / actual	2190.9	2604.9	3520.0	4845.7	6492.6
Transformer	Pose + velocity / donor	2190.9	2783.0	3769.9	5172.0	6875.0
Transformer	Pose + rotated velocity / actual	2190.9	2615.6	3516.5	4835.0	6476.4
Transformer	Pose + rotated velocity / donor	2190.9	2807.7	3799.2	5197.1	6886.7
GRU	Free	2145.3	3254.1	5025.7	7188.0	9667.3
GRU	Pose-only / actual	2145.3	2800.3	4301.9	6282.3	8673.6
GRU	Pose-only / donor	2145.3	3092.2	4632.9	6712.5	9207.6
GRU	Pose + velocity / actual	2145.3	2853.2	4368.9	6350.5	8754.3
GRU	Pose + velocity / donor	2145.3	3133.6	4728.0	6814.4	9309.6
GRU	Pose + rotated velocity / actual	2145.3	2849.5	4389.3	6404.0	8804.9
GRU	Pose + rotated velocity / donor	2145.3	3126.2	4730.5	6855.0	9364.5

C.6 All fixed model and reference stream cells

The following descriptive means retain every architecture, objective, reference stream and constraint condition. The same free rollouts appear beside each constraint for comparison. These cells are not independent samples; inference uses recipient-level aggregation in Tables 20 and 21.

Table 23: **All fixed endpoint cells.** Mean block-position squared error for every architecture, objective, reference stream and constraint. Inference aggregates within recipients; these are descriptive cells.

Architecture	Reference stream	Objective	Constraint	Free	Actual	Donor
Transformer	0	Coordinate teacher	Pose-only	4739.1	4485.3	4842.1
Transformer	0	Coordinate teacher	Pose + velocity	4739.1	4492.2	4780.8
Transformer	0	Coordinate teacher	Pose + rotated velocity	4739.1	4518.9	4806.3
Transformer	0	Physical labels	Pose-only	4812.4	4340.1	4642.5
Transformer	0	Physical labels	Pose + velocity	4812.4	4394.6	4637.7
Transformer	0	Physical labels	Pose + rotated velocity	4812.4	4399.9	4608.4
Transformer	1	Coordinate teacher	Pose-only	6927.3	6517.7	7019.5
Transformer	1	Coordinate teacher	Pose + velocity	6927.3	6485.1	6927.7
Transformer	1	Coordinate teacher	Pose + rotated velocity	6927.3	6496.2	7027.6
Transformer	1	Physical labels	Pose-only	6651.0	6069.2	6430.4
Transformer	1	Physical labels	Pose + velocity	6651.0	6175.9	6529.7
Transformer	1	Physical labels	Pose + rotated velocity	6651.0	6121.4	6524.0
Transformer	2	Coordinate teacher	Pose-only	5786.6	5352.6	5834.9
Transformer	2	Coordinate teacher	Pose + velocity	5786.6	5432.4	5811.0

Architecture	Reference stream	Objective	Constraint	Free	Actual	Donor
Transformer	2	Coordinate teacher	Pose + rotated velocity	5786.6	5415.7	5790.5
Transformer	2	Physical labels	Pose-only	5953.1	5373.2	5745.3
Transformer	2	Physical labels	Pose + velocity	5953.1	5523.9	5775.8
Transformer	2	Physical labels	Pose + rotated velocity	5953.1	5443.2	5823.0
Transformer	3	Coordinate teacher	Pose-only	10028.6	9407.9	10154.2
Transformer	3	Coordinate teacher	Pose + velocity	10028.6	9598.0	10220.2
Transformer	3	Coordinate teacher	Pose + rotated velocity	10028.6	9545.0	10166.3
Transformer	3	Physical labels	Pose-only	10370.5	9687.3	10272.3
Transformer	3	Physical labels	Pose + velocity	10370.5	9838.8	10317.3
Transformer	3	Physical labels	Pose + rotated velocity	10370.5	9871.2	10347.8
GRU	0	Coordinate teacher	Pose-only	8057.0	7148.1	7398.5
GRU	0	Coordinate teacher	Pose + velocity	8057.0	7270.6	7599.1
GRU	0	Coordinate teacher	Pose + rotated velocity	8057.0	7276.1	7751.6
GRU	0	Physical labels	Pose-only	8489.4	7430.3	7724.6
GRU	0	Physical labels	Pose + velocity	8489.4	7573.0	7885.8
GRU	0	Physical labels	Pose + rotated velocity	8489.4	7629.3	8102.8
GRU	1	Coordinate teacher	Pose-only	11061.3	10016.4	10753.9
GRU	1	Coordinate teacher	Pose + velocity	11061.3	10147.0	10791.6
GRU	1	Coordinate teacher	Pose + rotated velocity	11061.3	10135.4	10746.1
GRU	1	Physical labels	Pose-only	11176.7	10204.2	10697.4
GRU	1	Physical labels	Pose + velocity	11176.7	10103.5	10873.1
GRU	1	Physical labels	Pose + rotated velocity	11176.7	10113.7	10745.2
GRU	2	Coordinate teacher	Pose-only	7774.8	6968.8	7450.0
GRU	2	Coordinate teacher	Pose + velocity	7774.8	7047.2	7563.7
GRU	2	Coordinate teacher	Pose + rotated velocity	7774.8	7169.9	7599.6
GRU	2	Physical labels	Pose-only	7800.5	6795.5	7331.3
GRU	2	Physical labels	Pose + velocity	7800.5	6848.2	7371.2
GRU	2	Physical labels	Pose + rotated velocity	7800.5	6937.6	7453.8
GRU	3	Coordinate teacher	Pose-only	11534.8	10462.5	11095.2
GRU	3	Coordinate teacher	Pose + velocity	11534.8	10562.2	11204.8
GRU	3	Coordinate teacher	Pose + rotated velocity	11534.8	10655.3	11223.2
GRU	3	Physical labels	Pose-only	11444.3	10362.6	11209.8
GRU	3	Physical labels	Pose + velocity	11444.3	10482.6	11187.9
GRU	3	Physical labels	Pose + rotated velocity	11444.3	10521.8	11293.8

C.7 Measurement transport and motion changes

Table 24 evaluates the same frozen readout on every formal insertion context. Observed-token block-component training-mean R^2 averaged over the three components ranges from 0.522 to 0.643 across the eight architecture/reference stream cells. For predicted tokens, the range is 0.093–0.385 across both objectives and all reference streams. Individual components remain visible in the table. The added constraint therefore preserves a learned motion measurement whose prediction-domain errors remain material, rather than certifying equality of the true physical velocity.

The evidence includes every recipient-level error, all five horizons, pre/post velocity values, contact/no-contact descriptions and geometry diagnostics. No additional hypothesis test is selected from these strata. The reported main comparison with the rotated control is unresolved in both architectures; it does not establish equivalence or the irrelevance of velocity. The positive joint actual-versus-free result establishes remaining correctability relative to the jointly preserved measurements.

D Earlier experiments with the original readout

These results use the original 256-wide readouts and earlier protocols. The 128-recipient prediction and 512-recipient decision confirmations are separate from the new 256-recipient stronger-readout confirmation. They are retained as earlier evidence, not pooled with the new sample.

Table 24: **Every formal measurement-transport cell.** Five columns give training-mean R^2 against the frozen training mean. Block loss is mean standardized MSE over block v_x, v_y, ω. Reference streams 0–1 use the policy types in measurement qualification; 2–3 use the retained state-policy reference streams. These are descriptive checks; no reference stream is dropped or used to reselect a readout.

Architecture	Reference stream	Token	v_{ax}	v_{ay}	v_{bx}	v_{by}	ω_b	Block loss
Transformer	0	Observed encoding	0.998	0.998	0.668	0.615	0.590	0.257
Transformer	0	Coordinate teacher	0.998	0.998	0.475	0.481	0.169	0.436
Transformer	0	Physical labels	0.998	0.998	0.463	0.470	0.150	0.445
Transformer	1	Observed encoding	0.999	0.998	0.707	0.669	0.469	0.289
Transformer	1	Coordinate teacher	0.999	0.998	0.495	0.436	0.096	0.494
Transformer	1	Physical labels	0.999	0.998	0.485	0.399	0.055	0.514
Transformer	2	Observed encoding	0.999	0.998	0.678	0.578	0.494	0.259
Transformer	2	Coordinate teacher	0.999	0.998	0.472	0.310	0.094	0.440
Transformer	2	Physical labels	0.999	0.998	0.481	0.310	0.104	0.435
Transformer	3	Observed encoding	0.999	0.999	0.638	0.682	0.498	0.432
Transformer	3	Coordinate teacher	0.999	0.999	0.213	0.257	0.047	0.909
Transformer	3	Physical labels	0.999	0.999	0.211	0.224	-0.002	0.938
GRU	0	Observed encoding	0.999	0.998	0.699	0.655	0.424	0.215
GRU	0	Coordinate teacher	0.998	0.998	0.521	0.434	0.200	0.325
GRU	0	Physical labels	0.998	0.998	0.523	0.438	0.190	0.326
GRU	1	Observed encoding	0.999	0.999	0.661	0.788	0.480	0.473
GRU	1	Coordinate teacher	0.999	0.999	0.406	0.306	0.132	0.979
GRU	1	Physical labels	0.999	0.999	0.372	0.265	0.114	1.023
GRU	2	Observed encoding	0.998	0.998	0.708	0.591	0.284	0.264
GRU	2	Coordinate teacher	0.999	0.998	0.391	0.366	0.068	0.417
GRU	2	Physical labels	0.998	0.998	0.384	0.335	0.011	0.434
GRU	3	Observed encoding	0.999	0.999	0.744	0.621	0.202	0.835
GRU	3	Coordinate teacher	0.999	0.999	0.275	0.194	-0.132	1.618
GRU	3	Physical labels	0.999	0.999	0.284	0.118	-0.124	1.661

Table 25: **Velocity readout changes and post-insertion measurement errors.** Mean normalized $\|\Delta h\|_2$ compares inserted with original predicted feedback. The five remaining columns are post-insertion standardized measurement MSE against the same true velocity. All architectures, conditions and sources are retained. Joint velocity changes are numerical residuals within the accepted tolerance.

Architecture	Constraint	Source	$\|\Delta h\|_2$	v_{ax}	v_{ay}	v_{bx}	v_{by}	ω_b
Transformer	Pose-only	actual	0.1267	0.002	0.002	0.541	0.365	0.630
Transformer	Pose-only	donor	0.09854	0.002	0.002	0.589	0.409	0.724
Transformer	Pose + velocity	actual	2.19e-08	0.002	0.002	0.596	0.413	0.720
Transformer	Pose + velocity	donor	2.07e-08	0.002	0.002	0.596	0.413	0.720
Transformer	Pose + rotated velocity	actual	0.1211	0.002	0.002	0.542	0.366	0.638
Transformer	Pose + rotated velocity	donor	0.0987	0.002	0.002	0.584	0.404	0.724
GRU	Pose-only	actual	0.1754	0.001	0.002	0.667	0.752	0.828
GRU	Pose-only	donor	0.1609	0.001	0.002	0.722	0.833	0.900
GRU	Pose + velocity	actual	2.576e-08	0.001	0.002	0.776	0.861	0.906
GRU	Pose + velocity	donor	2.55e-08	0.001	0.002	0.776	0.861	0.906
GRU	Pose + rotated velocity	actual	0.1707	0.001	0.002	0.684	0.752	0.831
GRU	Pose + rotated velocity	donor	0.1618	0.001	0.002	0.730	0.834	0.912

D.1 Improving future prediction while preserving the current readout

The readout-preserving intervention tests feedback quality within each model. At matched standardized displacement norm, actual-guided correction reduces terminal block-position MSE by 972.48 under coordinate-teacher training and 697.29 under physical-label training. These are relative improvements of **8.42% and 5.10%**. Their primary 99% intervals are $[-1264.39, -698.05]$ and $[-927.87, -468.33]$, respectively. Every one of the six model points improves under each coordinate objective.

Table 26 gives the absolute error scale. The pose at action 5 does not improve: it is preserved by construction. Across the confirmation, the largest normalized readout deviation for the scaled replacements is 2.4×10^{-7}. All earlier tokens and actions also remain fixed. The later position gains therefore reveal a useful change in the feedback input that the current pose readout cannot distinguish. They establish more than sensitivity to a latent perturbation: a targeted perturbation improves prediction against the same physical future.

The objective-dependent response survives norm matching. Relative to the latent objective, the coordinate-teacher and physical-label intervention effects are larger by 650.67 and 375.48 block-position MSE units, with secondary 95% intervals excluding zero. The stronger response distinguishes coordinate training from the latent objective under this diagnostic.

Endpoint evaluated using	Latent	Coordinate teacher	Physical labels
Observed encoding	5911.37	5911.37	5911.37
Observed-history prediction	5831.28	5924.17	3754.19
Free rollout	7386.40	11552.71	13665.47
Actual	7064.59	10580.23	12968.18

Table 26: Absolute terminal block-position MSE (pixels squared), averaged over 128 recipients, six model groups and four reference streams under the original readouts. Observed encoding decodes the actual endpoint image; observed-history prediction predicts the endpoint from earlier observed tokens. Matched feedback changes only the first predicted token and preserves its current readout.

D.2 Guidance source and local encoding neighborhoods

At the same standardized displacement norm and current pose, actual guidance outperforms donor guidance under all three objectives. The primary block-position MSE differences are -803.39 for the latent objective, -1069.29 for the coordinate-teacher objective, and -471.35 for the physical-label objective. Their 99% intervals all exclude zero (Figure 8). Thus the five frozen primary comparisons support both a benefit over free rollout and a benefit specific to the guidance source.

Actual guidance wins over donor guidance at 17 of the 18 model points. The remaining physical-label model favors the donor by 476.32 MSE units, while actual guidance still improves its prediction relative to free rollout. The aggregate source benefit therefore does not require uniform ordering at every model point.

These controls separate current pose correction, standardized displacement norm, and guidance source. The pose readout is fixed and the standardized displacement norm is matched, yet the source changes future accuracy. They identify distinctions between pose-equivalent feedback tokens whose modification improves recursive prediction.

Closer donors recover part of the benefit. On the same 128 recipients, we select donor encodings nearest to the current observation in normalized pose or standardized latent distance and jointly match all twelve standardized displacement norms. Both sources improve the coordinate objectives. Actual guidance remains better for the coordinate-teacher objective: actual-minus-nearest effects are -400.21 and -229.87 MSE units, with both exploratory 95% intervals below zero. Physical-label source comparisons remain unresolved. These controls support useful neighborhood information and an additional coordinate-teacher source benefit under these two proximity measures (Appendix L).

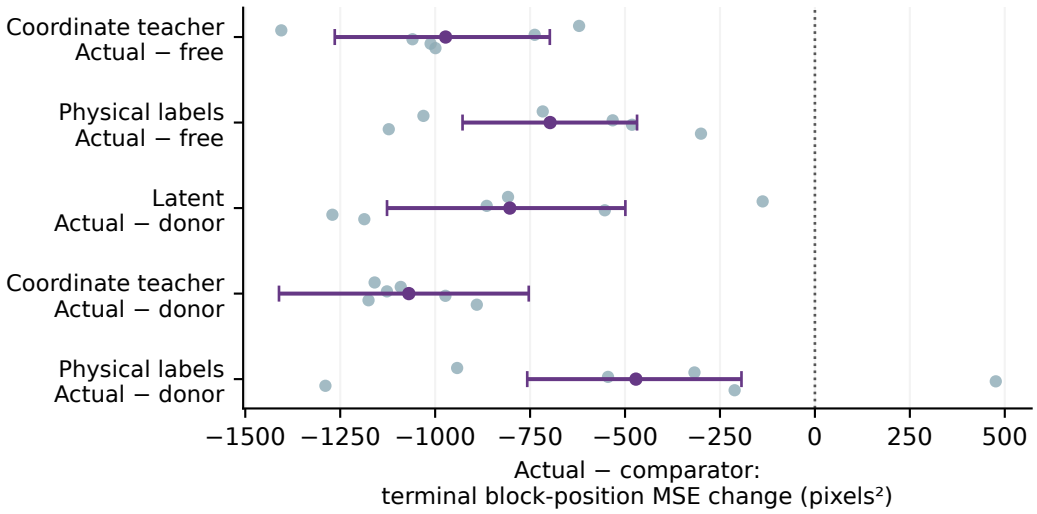

Figure 8: Independent confirmation of the feedback mechanism. Every row compares actual guidance with free rollout or donor guidance at action 25. All five primary 99% intervals lie below zero. Small points show all six model groups, including the physical-label donor comparison favoring the donor. All guidance branches preserve pose and use matched standardized displacement norm.

D.3 Supervision ordering and readout dependence

Under the original readouts, physical-label-minus-coordinate-teacher terminal block-position MSE is -2169.98 with observed-history prediction (secondary 95% interval $[-2873.03, -1544.85]$), but $+2112.76$ under free rollout ($[1080.43, 3174.09]$). This reversal uses the same actions and physical targets; only the source of the input history changes (Figure 9).

The ordering also depends on the readout. In a pilot gated on planner-validation error, a refitted readout and matched continuations reduce the free-rollout physical-label-minus-coordinate-teacher gap from 3239.10 to -270.35 in one Transformer; the paired change is -3509.45 (exploratory 95% interval $[-5579.23, -1426.16]$). Both free and observed-history prediction gaps now have intervals spanning zero. Readout calibration and geometry change together; planner and expert validation errors are reported in Appendix N.

The intervention improves both coordinate objectives without narrowing their gap. Physical labels minus coordinate teacher rises from 2112.76 to 2387.95 MSE units; the paired change is $+275.19$, with exploratory 95% interval $[48.62, 501.03]$. The reversal motivates the diagnosis, while the intervention identifies a within-model feedback effect rather than explaining the between-objective ordering. Appendix M reports absolute errors and the complete gap comparison.

The feedback effect survives without a resolved supervision reversal. We repeat the complete actual/donor/free intervention under the fitted readouts, freezing the design before its outcomes. Two configurations keep the pool 0 Transformer and GRU models fixed; a third uses the existing adapted Transformer continuations. All retain the three objectives, 128 recipients, four reference streams, exact nonlinear preservation and norm matching within the six-direction matching family. The GRU's calibration failure remains reported; its fixed-model sensitivity involves no additional continuation. All 9216 projections pass, with maximum normalized matched-readout deviation 3.91×10^{-7} (Appendix O).

In the adapted Transformer, actual-guided correction lowers coordinate-teacher and physical-label terminal block-position MSE by 713.10 and 701.98 (10.8% and 11.1%). All four actual-minus-free/donor exploratory intervals exclude zero (Table 27), although neither supervision ordering was resolved under this readout. With model weights fixed, coordinate-teacher benefits persist in both architectures, while several physical-label and latent comparisons remain unresolved. The useful feedback correction therefore survives a readout change where the motivating reversal weakens; its

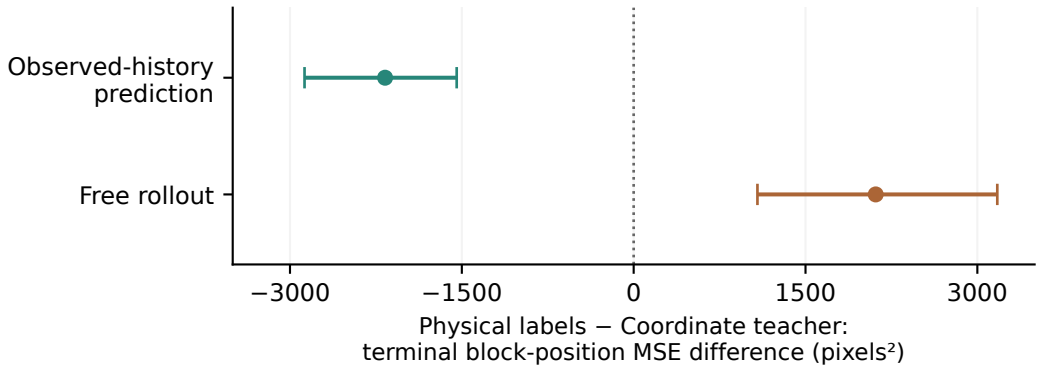

Figure 9: Better coordinate supervision reverses its advantage under free rollout. Physical-label training improves terminal block-position prediction relative to the coordinate-teacher objective with observed-history prediction, but worsens it in free rollout. Both estimates use the same 128 independent recipients, six model groups, and four reference streams; bars show prespecified secondary 95% intervals. Negative differences favor physical labels.

Configuration	Objective	Actual–free	Actual–donor
Transformer: fixed	Latent	-104.97 $[-399.31, 192.80]$	-680.70 $[-1200.59, -187.60]$
	Coordinate teacher	-1107.74 $[-1653.06, -597.95]$	-1499.10 $[-2169.17, -894.71]$
	Physical labels	-705.06 $[-1533.27, 102.06]$	-659.64 $[-1696.87, 366.54]$
GRU: fixed	Latent	-174.43 $[-497.16, 127.64]$	56.60 $[-270.63, 402.26]$
	Coordinate teacher	-3213.99 $[-4360.29, -2132.11]$	-1242.43 $[-2186.77, -290.97]$
	Physical labels	-1687.03 $[-2902.72, -486.84]$	761.66 $[-390.74, 1848.39]$
Transformer: adapted	Latent	-52.04 $[-366.30, 271.75]$	-632.35 $[-1120.19, -161.90]$
	Coordinate teacher	-713.10 $[-1122.93, -330.91]$	-881.83 $[-1251.58, -534.20]$
	Physical labels	-701.98 $[-1162.73, -272.24]$	-691.44 $[-1079.62, -318.75]$

Table 27: The complete feedback comparison under a second pose readout. Each configuration retains all 128 reused recipients, four fixed reference streams and three objectives; brackets are exploratory paired-recipient 95% intervals. The first two configurations change the readout and preserve model weights. The third uses the already adapted Transformer continuations; its supervision ordering was unresolved. Standardized displacement norms match across all six objective/source directions within each configuration. Negative values favor actual guidance.

size and source benefit still depend on the model–readout configuration. Readout changes also alter the feasible corrections and their matched standardized displacement norm, so this sensitivity does not isolate readout calibration alone.

D.4 Cost estimation, candidate ordering and realized decisions

On the 512-recipient visible-transport confirmation bank, both coordinate objectives reduce physical-cost estimation error at the same candidates chosen by free rollout. Yet their candidate rankings improve neither clearly nor uniformly, and all four primary selected-cost intervals span zero (Table 28). These are three distinct quantities measured on the same candidate pools: better estimates at fixed actions do not by themselves demonstrate better ordering or better decisions.

The coordinate-teacher objective's actual-minus-free point estimate is a 0.23% mean cost improvement, compared with 4.94% in development. This benefit is not confirmed on new recipients drawn by the same admission rules. Full observed-token reset likewise does not confirm lower cost under the pose score. Native latent-distance scoring retains distinct secondary benefits; it does not replace the four primary comparisons (Appendix K).

Quantity	Coordinate teacher	Physical labels
Fixed-candidate cost MAE change	-291.75 $[-373.98, -210.11]$	-137.69 $[-231.33, -46.40]$
Rank correlation: free $\rightarrow$ actual	$0.0723 \rightarrow 0.0767$	$0.0503 \rightarrow 0.0483$
Rank correlation change	0.0044 $[-0.0014, 0.0101]$	-0.0020 $[-0.0066, 0.0025]$
Selected cost: actual–free	-8.00 $[-74.07, 58.69]$	33.98 $[-27.22, 95.35]$
Selected cost: actual–donor	40.43 $[-35.33, 118.35]$	19.14 $[-47.92, 83.26]$

Table 28: Three evaluation levels on the same 512 recipients. Cost MAE is measured at each model's fixed free-selected candidates; rank is within-pool Spearman correlation with realized physical cost. Their changes use exploratory paired-recipient 95% intervals. The bottom four selected-cost comparisons are the frozen primary tests with 98.75% intervals (nominal familywise 95%). Negative cost differences favor actual guidance; positive rank differences favor it.

Decisions nevertheless depend on feedback distinctions invisible to current pose. Actual guidance changes the minimizing candidate set in 49.2%, 46.7% and 37.0% of latent, coordinate-teacher and physical-label comparisons, while maximum normalized readout deviation stays below 2.24×10^{-7}. Only 30.8–32.6% of changed choices have a free-score top-two margin below 5% of the pool's interquartile range. Near ties contribute to sensitivity but do not cover all changes. The latent objective's free rank correlation is also weak (0.1134), and its change remains unresolved. Appendix M reports all objectives, reset branches and tie thresholds. These diagnostics localize the missing evidence for decision improvement to candidate ordering and selection; they do not identify a unique reason for the weak rankings.

D.5 The effect has a temporal boundary

Simulated Reacher tests the same intervention principle in joint-angle prediction. Six independently trained models, four reference streams, and 256 recipients yield 6144 corrections preserving both the circle readout and decoded joint angles. Joint-angle MSE falls from 0.01027 to 0.00301 at action 10 and from 0.01774 to 0.00469 at action 15. By action 25, the mean difference is -0.000735 with a secondary 95% interval of $[-0.001691, +0.000910]$.

A single pose-preserving correction can therefore improve the intervening predictions without securing a lasting endpoint advantage. Observed-history prediction remains near 0.0030 MSE at the endpoint, while both single-correction and full-reset branches return toward free-rollout error. This limits what one intervention can accomplish over time and motivates evaluating its effect as a curve over the planning horizon. The complete curve and joint-metric details are in Appendix H.

E Readout-preserving projection

Let $x = (z - \mu)/\sigma$ be the standardized latent token, where the readout's fixed mean and scale are applied coordinatewise. With hidden preactivations $a = W_0 x + b_0$ and $b = W_1 \operatorname{ReLU}(a) + b_1$, define their activation masks D_0, D_1. Within this original ReLU activation region, the Jacobian of the normalized readout output with respect to x is

$$J = W_2 D_1 W_1 D_0 W_0. \tag{9}$$

The physical-unit readout g also applies the fixed output scale and offset; its Jacobian with respect to x is $\operatorname{diag}(\sigma_y)J$. All output scales are positive, so $Jd = 0$ preserves both normalized and physical-unit outputs. For standardized observed target x_o, solve

$$\min_d \quad \tfrac{1}{2}\|d - (x_o - x)\|_2^2, \tag{10}$$

$$\text{subject to} \quad Jd = 0, \tag{11}$$

$$S_0(a + W_0 d) \geq 0, \tag{12}$$

$$S_1(b + W_1 D_0 W_0 d) \geq 0, \tag{13}$$

where S_i contains the signs of the original preactivations. These affine inequalities preserve the original region. We use an SVD row-space basis for $Jd = 0$, with relative singular-value cutoff

10^{-10}, and normalize the inequality rows. OSQP solves the problem with absolute and relative tolerance 10^{-9}, maximum 20,000 iterations, and polishing. This is an convex projection within the original ReLU activation region, rather than a claim of the closest point over the entire nonlinear readout fiber.

The inserted token $(x+d)\sigma+\mu$ is cast to FP32. Its complete nonlinear readout is then recomputed in FP64. PushT accepts each normalized output deviation at most 10^{-6}. Reacher also requires decoded-angle deviation at most 10^{-5} radians, in addition to raw circle-output deviation at most 10^{-6}. The naive local-null projection can cross activation boundaries: every one of the 720 predetermined PushT feasibility pairs violated the nonlinear preservation criterion under that method, while all constrained pairs passed.

F Displacement and donor controls

Within a group, all three objectives share the same frozen perception and readout. Let d_c^s be an accepted full correction in standardized latent coordinates, where c indexes the continuation objective and $s \in \{\mathrm{act}, \mathrm{don}\}$ indexes the guidance source. For each recipient and reference stream, the confirmation takes $m = \min_{c,s} \|d_c^s\|_2$ over all six directions. When $m > 0$, the replacement is

$$z_{5,c}^s = z_{5,c} + \sigma \odot \left(\frac{m}{\|d_c^s\|_2} d_c^s \right). \tag{14}$$

Here σ is the fixed input scale and $\odot$ denotes coordinatewise multiplication. If $m = 0$, all replacements equal their original predictions. The earlier displacement-only control takes the minimum over the three actual-guided directions. Convex interpolation preserves the region in exact arithmetic. We check every final FP32 token, requiring normalized readout deviation and region violation at most 10^{-6} and norm agreement with relative and absolute tolerance 10^{-6}.

The consumed-bank donor experiment uses a single random derangement shared across all groups, objectives and reference-stream slots. Its recipient intervals condition on realized assignments and omit uncertainty from shared donor/recipient coupling. Confirmation uses a separate fixed donor bank and fresh recipients. No donor is selected using errors or future improvement.

G Numerical reconstruction and experimental units

Mechanism rollouts preserve the original planning batch and modify only the selected trajectory. The unchanged branch reproduces every token of its accepted reference trajectory exactly. This controls implementation differences between the reference rollout and the intervention evaluation.

Physical trajectories are replayed independently. Saved latent arrays are decoded and scored with NumPy. Cross-platform last-bit differences in FP64 physical metrics are recorded separately, using unchanged tolerances. Precision outcomes, latent-error metrics and input bindings agree exactly. Every model tensor and buffer required to remain fixed is checked after inference.

Recipient-level paired vectors average across the fixed model and reference stream roster before resampling. Thus more architectures, reference streams or horizons do not create additional independent test recipients. All reported model points are retained, including mixed or adverse responses. The present three training pools and two architectures establish repetition across this roster, rather than a distribution over all model families.

H Reacher: horizon and metric boundary

The native metric wraps the first joint angle and uses the fixed chart for the second. All evaluated true second-joint states lie inside that chart. One Transformer case has target second angle -2.85444 radians, free rollout $+2.30370$, and corrected prediction $+2.71840$. Both predictions already lie on the wrong side under the native metric. The second-joint squared-error change is $+4.45016$, contributing $+0.00072431$ to the overall equal-weight mean. This is comparable to the aggregate net improvement. The case and original metric are retained; no wrapped replacement metric is used to reverse the endpoint conclusion.

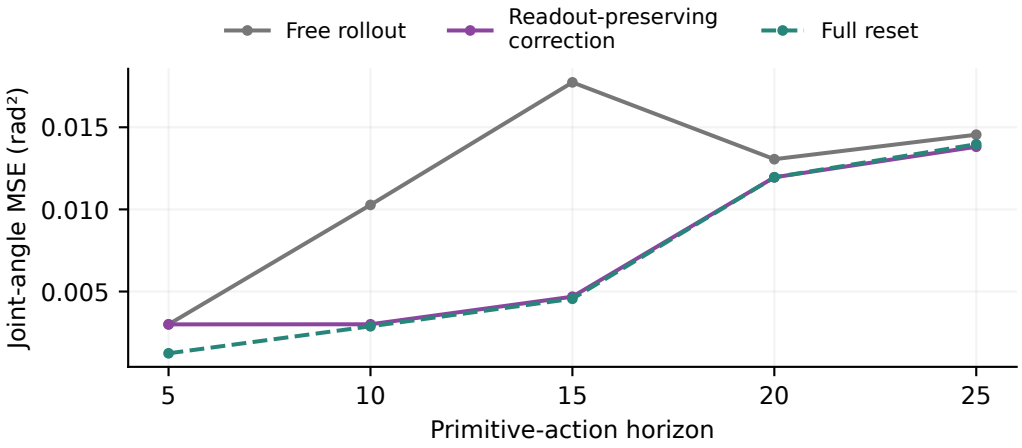

Figure 10: A single correction improves intermediate prediction, but its benefit fades. Curves average six Reacher models, four reference streams, and 256 recipients. The interval for the 25-action readout-preserving-minus-free-rollout MSE difference includes zero.

I Original architecture and continuation details

This section specifies the earlier one-forward continuation. The new three-transition training comparison is specified in Appendix A. The original models use 256 expert episodes per training pool and 21,000 updates. Continuations share 1280 training windows from the planner distribution. The fixed pose readout is already trained using physical labels, so the coordinate-teacher condition is not a label-free representation-learning system.

Images have resolution 224×224. The tiny vision transformer is initialized without external pretraining and produces 192-dimensional representations. The Transformer predictor has six layers, 16 attention heads, head dimension 64, and feedforward dimension 2048. The GRU predictor has two layers of width 384, consumes concatenated 192-dimensional observation and action embeddings, and maps its output back to 192 dimensions. Each predictor receives the same rolling three-token window; the GRU does not carry a separate hidden state beyond that window. Action embeddings encode five primitive two-dimensional actions per block.

The PushT readout has dimensions $192 \to 256 \to 256 \to 6$, with ReLU hidden activations. Reacher uses the same hidden dimensions and four circle outputs. Within each group the three PushT continuation objectives share the exact readout, image encoder, and observation projector. The continuation trains the predictor, action encoder and prediction projector for 2100 updates with batch size 128, AdamW learning rate 10^{-5} and weight decay 10^{-3}. Before each update the complete gradient is normalized to unit Euclidean norm, computed in FP64 and applied to FP32 gradients. This controls the gradient norm at the optimizer input while allowing objective-dependent AdamW updates.

Exact continuation losses. Each training window contains four observations, five primitive actions between observations, and three predicted successors. Write $z_k = E(o_{5k})$, $k = 0, \ldots, 3$, and let $(\hat{z}_1, \hat{z}_2, \hat{z}_3)$ denote the temporal predictor's three outputs from the first three observed tokens and their aligned action blocks. The rolling transition in the main text uses the final output of this same predictor. Define $\tilde{g}(z) = (g(z) - \mu_y) \oslash \sigma_y$ and $\tilde{y} = (y - \mu_y) \oslash \sigma_y$, where

$y = (x_a, y_a, x_b, y_b, \sin\theta_b, \cos\theta_b)$. For batch size B, the implemented losses are

$$\mathcal{L}_{\text{latent}} = \frac{1}{3B \cdot 192} \sum_{i=1}^{B} \sum_{k=1}^{3} \|\hat{z}_{i,k} - z_{i,k}\|_2^2, \tag{15}$$

$$\mathcal{L}_{\text{teacher}} = \frac{1}{3B \cdot 6} \sum_{i=1}^{B} \sum_{k=1}^{3} \|\tilde{g}(\hat{z}_{i,k}) - \tilde{g}(z_{i,k})\|_2^2, \tag{16}$$

$$\mathcal{L}_{\text{labels}} = \frac{1}{3B \cdot 6} \sum_{i=1}^{B} \sum_{k=1}^{3} \|\tilde{g}(\hat{z}_{i,k}) - \tilde{y}_{i,k}\|_2^2. \tag{17}$$

Latent targets use the raw 192-dimensional encoder/projector outputs, without additional loss standardization. Coordinate targets use the fixed readout's training-set mean and population standard deviation, separately for all six outputs. The readout also standardizes its latent inputs using its training-set means and standard deviations. A standard deviation at most 10^{-12} is replaced by one. There are no additional per-output weights; in particular, the decision score's orientation weight 900 is not a continuation-loss weight. Targets are detached, and no SIGReg or auxiliary term is added in continuation. The encoder, observation projector, readout and all running buffers remain fixed. All modules use evaluation-mode forwards: dropout is disabled and BatchNorm statistics do not update. Only predictor, action encoder and prediction-projector parameters receive gradients. Physical labels are available to the original readout fit as well as the physical-label continuation.

Original JEPA training. The source model uses the same three-successor latent MSE plus $0.09\,\mathcal{R}_{\text{SIGReg}}$. Both prediction and target encodings receive gradients during this original training; there is no separate EMA encoder or stop-gradient target. For four observation times and $M = 1024$ random unit directions $a_m \in \mathbb{R}^{192}$, the released regularizer computes

$$\mathcal{R}_{\text{SIGReg}} = \frac{B}{4M} \sum_{k=0}^{3} \sum_{m=1}^{M} \sum_{j=0}^{16} w_j \left| \frac{1}{B} \sum_{i=1}^{B} e^{\mathrm{i} t_j a_m^\top z_{i,k}} - e^{-t_j^2/2} \right|^2, \tag{18}$$

where $t_j = 3j/16$, $w_j = 2(3/16)e^{-t_j^2/2}$ at interior knots, and the two endpoint weights are halved. The complex modulus is implemented as the squared cosine discrepancy plus squared sine mean. Directions are resampled from normalized Gaussian vectors. The source models used here have no action auxiliary loss. Original training uses batch size 128, AdamW with peak learning rate 5×10^{-5}, weight decay 10^{-3}, a 10% linear warmup followed by cosine decay, and gradient clipping at norm one. It uses BF16 autocast; continuation and evaluation use strict FP32 dynamics, with the frozen pose readout recomputed in FP64. Images are scaled to $[0, 1]$ then standardized with channel means $(.485, .456, .406)$ and standard deviations $(.229, .224, .225)$. Each primitive action coordinate uses the source training dataset's mean and standard deviation, repeated across the five actions in a block.

Four fixed reference streams. Each group supplies (i) original JEPA with random shooting, (ii) original JEPA with CEM, (iii) original direct-state model with random shooting, and (iv) original direct-state model with CEM. These are the original 21000-update checkpoints, before any continuation. Both use decoded block-position and wrapped orientation cost to their decoded goal; the JEPA uses the frozen pose readout and the direct-state model de-normalizes its six native outputs. All optimize 25 two-dimensional primitive actions independently. Random shooting and CEM each score 30 populations of 300 candidates, initialized with zero mean and standard deviation .2 and clipped to $[-.35, .35]$. The first candidate is the current distribution mean. Random shooting keeps its initial distribution; CEM uses 30 elites, population standard deviation, and a 10^{-3} deviation floor. Both return the best scored candidate over all populations, with stable ties. The selected sequence is replayed in the simulator and held fixed across all three continuation objectives and feedback branches. These four reference streams are repeated contexts for each recipient, not independent statistical replicates.

J Development experiments supporting the frozen design

The intervention was developed on a separate 128-recipient bank with the same 18-model roster and four reference streams. Table 29 summarizes the sequence of controls. The earlier prediction

results use the independent confirmation for their primary claims; these development results explain why displacement and donor controls were frozen before confirmation. All entries use terminal block-position MSE at action 25.

Comparison	Latent	Coordinate teacher	Physical labels
Full correction minus free	-179.97	-1378.28	-1203.91
Norm-matched correction minus free	-181.42	-927.48	-686.69
Actual–donor (matched)	-580.77	-1018.01	-633.89

Table 29: Development-bank endpoint effects. Matching standardized displacement norms reduces, but does not remove, the larger coordinate-objective response. Actual trajectory guidance outperforms donor guidance at the same standardized displacement norm and readout. The three-objective matching experiment and six-direction donor experiment have separately matched standardized displacement norms.

In the development bank, full reset under physical-label training worsens immediate block-position MSE by 896.32 while improving terminal block-position MSE by 2822.02. This complementary result shows that immediate pose accuracy and subsequent predictive utility can move in opposite directions. The pose-preserving intervention gives the main experiment a stricter control on immediate accuracy.

The independent confirmation retains all five prespecified contrasts. Its primary intervals use 20,000 shared-recipient draws with seed 1368001; the protocol-consistent secondary horizon analysis uses the same draw count and seed. Recipient-level averaging uses six fixed model groups and four fixed reference streams, with 18 models across the three objectives. No model point is removed on the basis of its response.

K Shared-prefix action-selection diagnostic

For each model group, the original JEPA random-shooting reference supplies a selected five-block action sequence. We execute its first block and observe o_5. The candidate pool at this decision point contains 32 twenty-action suffixes: the original selected continuation, a zero-action continuation, and 30 suffixes sampled without replacement from the other 299 members of the selected search population. Their sampling seed is the NumPy SeedSequence of the case seed, reference-route index, and 1378001. The old candidates had different first blocks; appending their suffixes to one common executed prefix defines a new pool. No candidate is removed for its visibility, contact, score, or realized cost.

Every candidate is simulated after resetting and replaying all ten actions of the initial observation history and the common five-action prefix. Matching only object poses would not preserve the physical velocities and contact history. The original selected continuation must reconstruct its archived physical trajectory; an independently repeated zero suffix supplies an additional replay check. Inputs available to the scorer and future simulator outcomes are stored separately. True goal coordinates do not enter the model scorer, which encodes the goal image and applies the frozen pose readout.

All objectives reconstruct their predicted z_5 using the original 300-candidate arithmetic. Actual and donor corrections use the same standardized displacement norm matched across six directions matching and nonlinear-readout checks as the prediction experiment. Each of the free, actual, donor, and full-reset branches then evaluates all 32 suffixes with identical batch arithmetic. Its first action window is $U_5 = (u_{-5}, u_0, u_5)$, and all branches are free-running after insertion. The full observed reset is a non-preserving reference.

Write $\hat{p}_k^b, \hat{\phi}_k^b$ for a candidate's predicted endpoint block pose, and $\hat{p}_\star, \hat{\phi}_\star$ for the pose decoded from the goal image. The pose score is

$$\hat{C}_k^b = \|\hat{p}_k^b - \hat{p}_\star\|_2^2 + 900\ \mathrm{wrap}(\hat{\phi}_k^b - \hat{\phi}_\star)^2, \tag{19}$$

where orientation is decoded by $\mathrm{atan2}$ from the readout's sine and cosine outputs, positions are in pixels, and angles are in radians. Let $A^b = \arg\min_k \hat{C}_k^b$. The evaluated outcome is $|A^b|^{-1} \sum_{k \in A^b} C_k$, where C_k uses the same cost formula on the true simulated endpoint and goal.

This implements uniform choice among exact score ties. Lowest-index choice is a recorded sensitivity check. Candidate-set regret subtracts a branch-independent minimum and therefore gives the same paired effects, rather than another independent result.

Development. The 128-recipient development experiment comprises six physical reference contexts per recipient, 24576 candidate trajectories, and 2304 model-recipient comparisons. Its uncertainty unit is the recipient. We use 20000 paired recipient-bootstrap draws with seed 1379001 for exploratory 95% intervals. For scale, mean realized costs are 3971.13 for uniform random choice, 1632.10 for the best candidate in the pool, 4679.32 for zero action, and 3358.30 for the original selected continuation. About 8.11% of candidate trajectories leave the predefined visibility box and 65.18% make contact; all remain in the analysis.

Objective	Actual–free	Actual–donor
Latent	-77.83 $[-200.44, 32.54]$	-56.81 $[-169.99, 52.76]$
Coordinate teacher	-181.69 $[-334.69, -52.90]$	-82.31 $[-212.11, 45.70]$
Physical labels	-7.30 $[-137.84, 105.86]$	-41.49 $[-135.58, 53.82]$

Table 30: Development decision results on 128 reused recipients. Brackets are exploratory paired-recipient 95% intervals, without familywise adjustment.

The development coordinate-teacher mean gain is 4.94%, with five of six model points improving. Its recipient-level median difference is -9.48, and the most favorable recipient contributes 29% of the aggregate improvement. All recipients were retained.

Secondary scoring uses squared latent distance to the goal-image encoding. For the coordinate-teacher objective, actual minus free is -169.27 with 95% interval $[-274.47, -70.15]$, and actual minus donor is -141.96 with interval $[-259.39, -26.96]$. These secondary results do not replace the unresolved source comparison under the primary pose score. All model points, both scores, tie counts, and per-recipient outcomes are retained in the experiment artifacts.

Independent decision confirmation. We freeze the same models, suffix-pool rule, intervention time, pose score and physical cost before generating 512 new recipients. The prospective construction requires full-shape visibility in its construction branches and at least 40 pixels of goal block displacement. The development bank uses the same generator, simulator and admission parameters: their recorded source hashes match exactly. Development begins at seed 1371001 and confirmation at 1381001; admitted seeds are disjoint. The two banks reject 47 and 171 construction seeds, respectively, before reaching their fixed quotas. Thus confirmation does not introduce an additional population restriction. These model-independent admission conditions define a restricted visible-transport population; the experiment is not the official PushT benchmark. All 33 construction branches are independently replayed, and no accepted recipient is replaced based on a model outcome. The decision bank contains 98304 suffix trajectories: 512 recipients, six fixed reference contexts and 32 candidates. The 18 models yield 9216 paired model-recipient comparisons.

The independent 128-case donor bank remains fixed. A permutation with seed 1368001 is repeated four times over the recipient indices, with the same donor assignment across objectives. Recipient and donor seeds are disjoint. We retain all models, candidates, recipients, zero-norm cases and tail outcomes. Uncertainty conditions on the fixed model roster and donor bank. The four primary pose-score contrasts use 20000 paired recipient-bootstrap draws with seed 1389001 and two-sided 98.75% percentile intervals. The paired difference between physical-label and coordinate-teacher intervention effects is a prespecified secondary comparison with a 95% interval. Secondary native-score and latent-objective effects do not replace the primary comparisons.

The four primary results appear in Table 28. None confirms lower selected-action cost. The secondary physical-label-minus-coordinate-teacher interaction is 41.98 with 95% interval $[-31.42, 115.88]$, so the experiment does not establish an objective difference in intervention benefit. Under the pose score, the latent-objective actual-minus-free and actual-minus-donor effects are -9.22 $[-62.70, 46.36]$ and -4.06 $[-68.95, 60.82]$, respectively (secondary 95%).

Mean realized costs of uniform random, best-in-pool, zero and original-continuation selection are 3882.46, 1667.57, 4555.21 and 3436.48, respectively. About 6.33% of candidate trajectories leave

Objective	Actual–free	Actual–donor
Latent	-61.78 $[-113.46, -11.85]$	-99.42 $[-156.48, -42.48]$
Coordinate teacher	-34.67 $[-84.43, 13.02]$	-23.86 $[-72.85, 27.45]$
Physical labels	4.24 $[-35.10, 44.73]$	6.27 $[-33.12, 45.98]$

Table 31: Secondary native latent-distance scoring on the 512 fresh recipients. Entries remain realized selected-action cost differences, with paired-recipient 95% intervals. These score-dependent findings do not replace the four primary pose-score comparisons.

Pool / predictor	Latent	Coordinate teacher	Physical labels
1 / Transformer	$(-88.96, -116.71)$	$(5.46, -23.43)$	$(47.20, 23.92)$
1 / GRU	$(90.85, 148.08)$	$(71.98, 88.39)$	$(-21.73, 4.70)$
2 / Transformer	$(21.12, 15.68)$	$(-29.59, 19.68)$	$(8.53, 38.07)$
2 / GRU	$(14.03, 63.52)$	$(8.20, 45.60)$	$(65.21, 13.72)$
3 / Transformer	$(-74.72, -84.12)$	$(-102.72, 16.61)$	$(137.83, 58.25)$
3 / GRU	$(-17.62, -50.82)$	$(-1.35, 95.73)$	$(-33.16, -23.82)$

Table 32: All 18 model points in the fresh decision study. Each entry is the ordered pair (actual minus free, actual minus donor) in realized cost under the primary pose score, averaged over all 512 recipients. Pools 1–3 are the three fixed training pools. The rows are repeated contexts, not independent test samples.

the visibility box and 61.68% make contact; all remain included. The action-choice changes therefore coexist with unresolved primary mean-cost gains in this fixed-pool test. An interval spanning zero does not establish equivalence of feedback branches.

L NEAREST-DONOR SOURCE CONTROL

We test whether closer donor observations reproduce the guidance benefit. This focused sensitivity uses all 128 previously consumed recipients and 18 models. Within each group it follows the original JEPA reference trajectory, the first suffix in the decision bank, with all subsequent actions fixed. It is a prediction-error experiment, not another action-selection experiment or an independent confirmation. Donor selection and analysis were frozen before evaluating the new branches.

Let $x(z) = (z - \mu_z) \oslash \sigma_z$ be the fixed readout-input standardization and $r(z) = (g(z) - \mu_y) \oslash \sigma_y$ its six normalized outputs, with coordinatewise division. From the original independent bank of 128 observed encodings, pose-nearest minimizes $\|r(z^{\mathrm{don}}) - r(E(o_5))\|_2$ and latent-nearest minimizes $\|x(z^{\mathrm{don}}) - x(E(o_5))\|_2$. Ties use the smallest bank index. The donor indices are shared across objectives within each group. Neither selector uses future observations or outcomes; both use recipient information from the current observation. They are neighborhood-substitution controls, rather than information-free negative controls.

Mean standardized latent distance to the actual observation falls from 16.30 for the random donor to 10.00 for pose-nearest and 7.14 for latent-nearest. Their mean normalized readout distances are 3.675, 0.917 and 1.441, respectively. Nearest selection substantially narrows the geometric mismatch, but does not eliminate it or establish membership in a dynamically reachable manifold.

We project the four sources—actual, random, pose-nearest and latent-nearest— and jointly match the twelve standardized displacement norms across three objectives. All 4608 new nearest-source projections pass the fixed constraints. The new common norm is smaller than the original standardized displacement norm matched across six directions in 112 of 768 group-recipient contexts; actual and random branches are therefore reevaluated. Free and full-reset trajectories reproduce their accepted 32-candidate decision-bank anchors exactly; only the fixed first suffix contributes to this prediction estimand. The maximum normalized readout deviation is below 2.24×10^{-7}. Prediction error uses the same true physical position at actions 10, 15, 20 and 25. No recipient or model is removed.

Objective	Actual minus	Mean difference	95% interval
Latent	Free	-5.77	$[-210.40, 247.08]$
	Random donor	-438.49	$[-687.20, -188.14]$
	Pose-nearest	-114.60	$[-316.12, 100.55]$
	Latent-nearest	-204.83	$[-363.34, -33.08]$
Coordinate teacher	Free	-871.28	$[-1113.72, -629.28]$
	Random donor	-1067.45	$[-1343.47, -802.45]$
	Pose-nearest	-400.21	$[-655.51, -160.53]$
	Latent-nearest	-229.87	$[-395.18, -59.51]$
Physical labels	Free	-632.97	$[-886.47, -354.39]$
	Random donor	-422.40	$[-728.52, -103.77]$
	Pose-nearest	-107.15	$[-351.89, 161.88]$
	Latent-nearest	-154.27	$[-361.78, 79.65]$

Table 33: All planned endpoint contrasts in the nearest-donor sensitivity. Negative values favor actual guidance. Intervals are exploratory paired-recipient 95% intervals on reused recipients, without familywise adjustment.

We average six fixed model groups within each recipient and use 20000 shared-recipient bootstrap draws with seed 1390001. Table 33 reports every planned endpoint contrast. For the coordinate-teacher objective, pose-nearest minus free and latent-nearest minus free are -471.07 and -641.41 MSE units, with intervals $[-723.86, -214.17]$ and $[-869.91, -427.04]$. The corresponding physical-label effects are -525.82 $[-720.71, -328.68]$ and -478.70 $[-687.71, -266.84]$. For the latent objective, pose-nearest minus free is 108.83 $[-90.26, 316.64]$ and latent-nearest minus free is 199.06 $[15.36, 393.51]$. Nearness alone therefore does not guarantee useful feedback. Actual guidance retains an advantage over both nearest sources for the coordinate-teacher objective; the unresolved physical-label differences do not establish equivalence. All horizon values and individual model effects are retained with the accepted experiment records.

M Objective gaps, readout error, and decision diagnostics

These additional analyses use previously evaluated recipients and accepted arrays, without new model selection or exclusions. They were specified after the original results were known. All new intervals are exploratory 95% paired-recipient percentile intervals; they do not expand either primary confirmation family. Prediction analyses average six fixed groups and four reference streams within each of 128 recipients, then use 20000 shared draws with seed 1368001. Decision analyses average the six fixed groups within each of 512 recipients and use 20000 shared draws with seed 1391001. Undefined rank correlations are retained as missing, with their counts recorded; no corresponding cost outcome is removed.

The supervision gap and the within-model intervention. Table 34 reports endpoint error levels for every saved branch. "Observed encoding" evaluates the frozen readout directly on the actual endpoint image; "observed-history prediction" still predicts the endpoint from earlier observations. They are different diagnostics. Actual guidance improves both coordinate objectives but does not restore their observed-history prediction ordering. The physical-label-minus-coordinate-teacher gap increases by 275.19, with interval [48.62,501.03] (Table 35). Four of six group point estimates increase; their changes are -95.74, 374.24, 205.41, 578.11, -110.15, and 699.27. Thus the experiment identifies correctable feedback error within each model, without supplying a causal decomposition of the ordering between objectives.

Observed-encoding readout error. The frozen readout is not exact on the diagnostic trajectories (Table 36). Coordinate-teacher targets $g(E(o))$ are compatible with the observed token by construction; physical-label targets need not be. Predictor compensation for readout error is therefore a concrete alternative source of the supervision gap. We measure association using each recipient's action-5 six-output normalized readout MSE, averaged over groups and reference streams, and that recipient's endpoint gap or actual-minus-free intervention effect. Pearson correlations are -0.267 $[-0.437, -0.099]$ (free physical-label-minus-coordinate-teacher gap),

Branch	Latent	Coordinate teacher	Physical labels
Observed encoding	5911.37	5911.37	5911.37
Observed-history prediction	5831.28	5924.17	3754.19
Free	7386.40	11552.71	13665.47
Actual	7064.59	10580.23	12968.18
Donor	7867.98	11649.53	13439.53
Full correction	7079.38	10164.95	12397.12
Full reset	6775.31	9464.20	10699.46

Table 34: Absolute block-position MSE at action 25, in pixels squared. Full correction is the unscaled readout-preserving correction; full reset replaces the token with its observed encoding. All branches use the same physical targets.

Condition	Physical labels minus coordinate teacher: mean [95% interval]
Observed-history prediction	-2169.98 $[-2873.03, -1544.85]$
Free	2112.76 $[1080.43, 3174.09]$
Actual	2387.95 $[1388.33, 3397.02]$
Change: actual–free	275.19 $[48.62, 501.03]$

Table 35: Paired supervision gaps. The last row is a difference of paired intervention effects, not subtraction of confidence-interval endpoints.

0.137 $[-0.086, 0.344]$ (latent intervention), 0.187 $[0.024, 0.340]$ (coordinate-teacher intervention), 0.085 $[-0.069, 0.244]$ (physical-label intervention). Negative intervention differences denote improvement, so a positive correlation denotes a less favorable effect at higher readout error. These associations do not establish that calibration causes the gap or mediates the intervention. Recipient difficulty and distribution shift can affect both quantities.

Action	Observed block-position MSE	Six-output normalized MSE
5	3581.06 $[2786.87, 4613.52]$	1.055 $[0.964, 1.146]$
10	4217.68 $[3317.05, 5302.80]$	1.128 $[1.033, 1.222]$
15	4787.15 $[3800.92, 5944.31]$	1.192 $[1.098, 1.286]$
20	5426.19 $[4349.94, 6648.74]$	1.261 $[1.163, 1.358]$
25	5911.37 $[4764.82, 7220.23]$	1.348 $[1.247, 1.452]$

Table 36: Frozen-readout error on actual image encodings. The six-output metric averages squared errors after the original per-output normalization; it is not an error in complete physical state.

Full reset and absolute decision costs. The full reset uses the observation already available after the executed prefix. It is evaluated over exactly the same candidate pool as every other branch. Table 37 reports all branches under both scores. None of the pose-score reset-minus-free intervals excludes zero. Native latent scoring gives secondary reset benefits for the latent and physical-label objectives. This score dependence is consistent with the secondary actual-guided correction results in Appendix K; it does not replace the four primary pose-score tests or establish that no scoring rule can benefit from feedback correction.

Ranking and cost prediction. Actual guidance does not resolve the weak pose-score ordering of physical candidate costs (Table 38). All actual-minus-free rank intervals cross zero. At the same time, mean absolute physical-cost prediction error decreases over the entire candidate pool for all objectives (Table 39). At the candidates selected by the free branch, both coordinate objectives improve, whereas the latent objective worsens. The error at each branch's own selected candidates also decreases, but that comparison changes both the predicted scores and the candidates being evaluated. Keeping the free-selected candidates fixed separates those two changes. These distinctions explain how better cost estimates can coexist with unresolved ranking and selected-cost gains; they do not identify a unique cause of the latter.

Objective / score	Free	Actual	Donor	Full reset	Full reset minus free
Latent / pose	3340.43	3331.21	3335.27	3330.12	-10.31 $[-74.88, 52.05]$
Latent / latent	3143.49	3081.71	3181.14	3058.80	-84.69 $[-148.50, -21.94]$
Coordinate teacher / pose	3524.80	3516.79	3476.37	3478.19	-46.61 $[-110.44, 17.64]$
Coordinate teacher / latent	3508.18	3473.51	3497.37	3475.29	-32.89 $[-99.28, 32.18]$
Physical labels / pose	3630.59	3664.57	3645.43	3658.35	27.76 $[-36.77, 91.71]$
Physical labels / latent	3642.70	3646.94	3640.67	3578.02	-64.68 $[-124.67, -6.77]$

Table 37: Absolute mean realized selected cost and exploratory reset-minus-free 95% intervals. Scores select the action; every column reports the same physical cost with exact-tie expectation. Native latent scores themselves are not interpreted as physical cost predictions.

Objective	Free ρ	Actual ρ	Full reset ρ	$\Delta\rho$	$\Delta\tau_b$
Latent	0.1134	0.1150	0.1198	0.0016 $[-0.0042, 0.0075]$	0.0014 $[-0.0029, 0.0057]$
Coordinate teacher	0.0723	0.0767	0.0814	0.0044 $[-0.0014, 0.0101]$	0.0031 $[-0.0011, 0.0073]$
Physical labels	0.0503	0.0483	0.0410	-0.0020 $[-0.0066, 0.0025]$	-0.0015 $[-0.0049, 0.0018]$

Table 38: Within-pool Spearman and Kendall rank diagnostics under the pose score. Differences are actual minus free, with exploratory 95% intervals. Correlations are averaged within recipient over defined fixed-model comparisons.

Objective	All candidates	Fixed free selection	Each branch's selection
Latent	-152.43 $[-203.12, -101.38]$	154.84 $[93.07, 219.31]$	-117.47 $[-180.84, -53.23]$
Coordinate teacher	-681.74 $[-768.26, -596.27]$	-291.75 $[-373.98, -210.11]$	-566.96 $[-649.55, -484.29]$
Physical labels	-460.09 $[-557.42, -363.98]$	-137.69 $[-231.33, -46.40]$	-449.09 $[-544.13, -356.59]$

Table 39: Actual-minus-free change in mean absolute cost prediction error, in physical cost units, with exploratory 95% intervals. The three columns define different candidate distributions; only the first two hold the evaluated candidates fixed.

How close are the choices to a tie? Let $s_{(1)} \leq s_{(2)}$ be the two lowest free-branch pose scores and let $I = Q_{.75}(s) - Q_{.25}(s)$ be the pool's interquartile range. We summarize $(s_{(2)} - s_{(1)})/I$ at three fixed thresholds (Table 40). All 9216 pools have positive I. Changed choices have smaller margins more often, but are not restricted to near ties: at the 5%-of-IQR threshold, only 30.8–32.6% of changed choices qualify, versus 12.2–15.5% of unchanged choices. Exact minimum-score ties occur in only 0.52%, 0.26%, and 0.65% of latent, coordinate-teacher and physical-label cases. These are descriptive proportions of fixed model-recipient rows; the rows are not treated as independent evidence for an additional hypothesis test.

Objective	$\leq 1\%$ IQR	$\leq 5\%$ IQR	$\leq 10\%$ IQR
Latent	6.5 / 9.1	23.9 / 32.6	40.3 / 53.3
Coordinate teacher	5.8 / 8.3	22.9 / 31.7	38.7 / 51.2
Physical labels	4.6 / 8.1	19.1 / 30.8	34.9 / 55.0

Table 40: Near-minimum margins: percentage of all choices / percentage of changed choices. Thresholds scale by the within-pool score IQR, not by the minimum score or physical outcome.

N READOUT DEPENDENCE OF THE SUPERVISION ORDERING

We select the pool 0's Transformer and GRU by roster order, before fitting a refitted readout. Starting from each original 192–256–256–6 ReLU readout, we fit its weights on the 1280 planner-training windows' four observed tokens, retaining the original input and target normalization. We use 3000 Adam steps at learning rate 10^{-4}, no weight decay, and batches of 256 sampled with replacement (seed 1392001 plus group index). The final checkpoint is used; there is no validation checkpoint selection. A predeclared gate requires at least 10% lower six-output normalized MSE *and* block-position MSE on 320 separate planner-validation windows. Table 41 retains both groups. The GRU fails the six-output criterion despite its lower block-position MSE; no GRU continuation follows.

		Six-output MSE		Block-position MSE		
Validation	Predictor	Original	Refitted	Original	Refitted	Gate
Planner	Transformer	1.047	0.892	3544.17	2662.72	Pass
Planner	GRU	1.107	1.047	4205.21	2866.09	Fail
Expert	Transformer	0.489	0.973	1541.33	2304.61	—
Expert	GRU	0.458	0.929	1420.82	2078.62	—

Table 41: Readout calibration on separate validation windows. Block-position MSE is in pixels squared. The gate uses planner validation only; error increases on the separate 1441-window expert validation set for both models.

For the passing Transformer, we freeze the refitted readout and repeat both coordinate continuations from the original 21000-update checkpoint: 2100 updates with the same training windows, index schedule, optimizer, gradient normalization and frozen-module boundary as Appendix I. Evaluation uses all 128 previously evaluated recipients, all four original reference reference streams, and horizons 5, 10, 15, 20 and 25. Actions are unchanged; there is no new search. The original-readout runs reproduce all 2048 complete free-rollout/observed-history-prediction token trajectories bit for bit. Independent NumPy reconstruction verifies all saved decoded predictions. Table 42 reports endpoint results; all five horizons are retained in the experiment records. Uncertainty uses 20000 shared-recipient draws (seed 1393001), after averaging the four reference streams within each recipient. All intervals are exploratory 95%.

Readout	Evaluation	Coordinate teacher MSE	Physical labels MSE
Original	Free	9868.47	13107.57
Original	Observed-history prediction	5494.74	4071.47
Refitted	Free	6588.23	6317.88
Refitted	Observed-history prediction	3959.19	3840.25

Physical labels minus coordinate teacher	Mean [95% interval]
Original / free rollout	3239.10 [1256.39, 5193.77]
Refitted / free rollout	-270.35 [-756.27, 199.94]
Original / observed-history prediction	-1423.27 [-2508.24, -419.93]
Refitted / observed-history prediction	-118.94 [-309.60, 51.54]
Gap change / free rollout	-3509.45 [-5579.23, -1426.16]
Gap change / observed-history prediction	1304.33 [387.45, 2302.84]

Table 42: Fixed-action terminal block-position MSE and paired objective gaps for the single passing Transformer. Gap change is refitted minus original, with shared recipients.

Decoding the actual endpoint image gives MSE 5404.46 under the original readout and 4071.19 under the refitted readout: a paired change of -1333.26 with interval $[-2129.53, -587.83]$. The previously separated objective orderings become unresolved under the refitted readout. This establishes readout dependence in the tested Transformer. The improvement is specific to the planner distribution. Fitting the readout changes calibration and geometry together; it does not isolate calibration as a mediator. Appendix O separately tests the complete readout-preserving feedback intervention under the fitted readouts.

O SECOND-READOUT FEEDBACK SENSITIVITY

This focused experiment fixes its three configurations before constructing new feedback corrections. All readouts, model weights, recipients, donors and actions already exist. The pool 0's Transformer and GRU are selected by roster order. Configurations 0 and 1 use the three original continuation models with the fitted readout from Appendix N. This holds the predictive model fixed while changing the readout, its admissible correction set and the measurement of future pose. Configuration 2 uses the same Transformer readout with the two previously adapted coordinate continuations and the original latent continuation; latent training does not use the pose readout. No new training occurs.

The GRU's earlier calibration gate failure remains a failure. Its fixed-model readout sensitivity is not a calibrated-continuation study, and no GRU continuation is added. Both fitted readouts improve planner position error but worsen expert validation error (Table 41); this is not a comparison to universally better physical-state measurements.

We retain all 128 previously evaluated recipients and all four original reference streams. The donor bank is the independent 128-recipient bank, with permutation seed 1368001. For every objective and source, we solve the same convex projection within the original ReLU activation region under the second readout, then rescale all six directions to their minimum standardized displacement norm. The complete six-output readout is rechecked after FP32 conversion; constraints for the original ReLU activation region and norm matching use the original 10^{-6} tolerances. No recipient, zero-norm case or adverse outcome is removed. Free, actual, donor and full reset branches use the original batch-of-300 arithmetic and the same selected actions. Every complete free trajectory must exactly reproduce its original or calibrated-readout archive.

All 9216 projections are accepted, and all 4608 complete free-token trajectories match bit for bit. Independent local reconstruction checks source bindings, readouts, regions, displacements and all saved horizon predictions. Table 43 reports constraint residuals and scales. Because each readout induces a different feasible region, its common standardized displacement norm is recomputed: an across-readout effect change is not a pure calibration effect at identical perturbation directions.

Configuration	Mean standardized displacement norm	Max. normalized readout deviation	Max. region violation
Transformer: fixed	3.726	2.59×10^{-7}	7.26×10^{-8}
GRU: fixed	3.622	3.91×10^{-7}	7.96×10^{-8}
Transformer: adapted	3.586	2.93×10^{-7}	8.29×10^{-8}

Table 43: All three-objective configurations pass the original nonlinear constraint and norm checks. Norms use the fixed readout-input standardization; normalized readout deviations use its six-output standardization.

Configuration	Objective	Free	Actual	Donor	Full reset
Transformer: fixed	Latent	6194.32	6089.36	6770.06	5499.34
	Coordinate teacher	11430.67	10322.94	11822.04	8691.62
	Physical labels	20407.98	19702.92	20362.56	18892.85
GRU: fixed	Latent	6622.61	6448.18	6391.58	5851.78
	Coordinate teacher	22412.70	19198.72	20441.15	16408.21
	Physical labels	37310.15	35623.12	34861.46	32028.06
Transformer: adapted	Latent	6194.32	6142.28	6774.63	5499.34
	Coordinate teacher	6588.23	5875.13	6756.96	5568.30
	Physical labels	6317.88	5615.90	6307.34	5489.38

Table 44: Absolute terminal block-position MSE under the second readout. The full reset changes the current readout and is a separate reference.

The earlier-result contrasts in Table 27 average four reference streams within each recipient before 20000 paired recipient-bootstrap draws (seed 1394001). Intervals are exploratory 95%; these consumed recipients do not constitute a new confirmation. All five horizons are retained in the evidence package. Configurations are not pooled as independent replicates. Table 45 gives the original-readout effects for the same fixed models, avoiding substitution of six-group aggregate estimates for these individual configurations.

Original readout	Objective	Actual–free	Actual–donor
Transformer	Latent	-87.58 $[-364.57, 192.34]$	-808.11 $[-1200.28, -414.76]$
	Coordinate teacher	-620.96 $[-1047.37, -221.60]$	-1159.42 $[-1740.30, -633.00]$
	Physical labels	-716.69 $[-1095.83, -343.12]$	-942.10 $[-1549.16, -316.00]$
GRU	Latent	-163.49 $[-500.07, 161.08]$	-137.59 $[-441.23, 165.18]$
	Coordinate teacher	-1405.14 $[-1932.50, -880.10]$	-1090.62 $[-1616.51, -598.24]$
	Physical labels	-1030.90 $[-1449.80, -638.59]$	-317.16 $[-806.10, 134.23]$

Table 45: Original-readout context on the identical pool 0 models and recipients, with the same exploratory bootstrap draws as the second-readout sensitivity. Full paired changes of intervention effects are retained in the evidence package.

P INITIAL BOUNDED MOTION-READOUT QUALIFICATION

The pose-preserving intervention leaves a natural explanation open: useful changes may concern dynamical variables omitted by the pose readout. This appendix retains the initial bounded development qualification, which failed its measurement gate. It was completed after the A–D results; none of their protocols or outcomes was changed. The subsequent, separately designed study with larger fixed data budgets, independent qualification and fresh intervention confirmation is reported in Appendix C.

A richer target must add a meaningful constraint. Let h_1, h_2 denote the two fixed tokens preceding the token z at insertion, and let u denote the known actions. Any estimate of the form

$$\widetilde{v}(z) = \psi\big(g_A(h_1), g_A(h_2), g_A(z), u\big) \tag{20}$$

is already preserved whenever the existing pose readout, earlier history and actions are preserved. This includes finite differences of the decoded positions and any fitted function of that same pose history. Accuracy of such an estimate alone would not establish an additional constraint. To test whether omitted motion information accounts for the feedback effect, we sought a velocity readout from the full latent history and actions, and evaluated its accuracy against a pose-history/action baseline.

Labels and temporal alignment. The seven-entry archived state contains agent velocity but omits block velocity. We replayed the unchanged simulator from each archived seed using its exact 35-action sequence. At the end of each primitive control step, after all physics substeps, we read agent v_x, v_y, block v_x, v_y, and block angular velocity directly from the Pymunk bodies. These are instantaneous simulator velocities, not position differences. Every archived state and every saved RGB frame matched its replay exactly; the reconstructed agent velocities also equaled the archived state entries. Control and physics frequencies were 10 and 100 Hz. Translational labels use simulator coordinate units per second; angular labels use radians per second.

The roster was the pool 0's Transformer and GRU, matching the previous dual-readout roster. Each used its two existing planner training routes and the corresponding validation routes. Three-frame contexts ended at archived steps 10, 15, 20, 25, 30 and 35. Each architecture supplied 1,536 training and 384 validation contexts from 256 and 64 trajectory replays (128 and 32 distinct initial seeds), respectively; these contexts are correlated within trajectories and are not independent confirmation recipients. Encodings came from the existing permitted observed-image caches. The training and validation contexts had no exact pixel or simulator-seed overlap. No confirmation trajectories were used for fitting or selection.

Conditional readout and bounded calibration. The velocity readout received three 192-dimensional encodings and the last five two-dimensional actions, for 586 inputs. Earlier encodings and actions would remain fixed during insertion, leaving only the latest token variable. Thus temporal information was available to the measurement without injecting a true velocity into the model. Latent normalization was the original g_A input normalization; action and five-output normalizers used training data only.

Before fitting, we recorded two ReLU MLP widths (256 and 512), two hidden layers, 3,000 Adam updates, learning rate 10^{-4}, batch size 256, and validation every 250 updates. Selection minimized mean standardized five-output validation MSE. The pose/action baseline used the 18 outputs of g_A on the same three tokens plus the same ten action values, a width-256 MLP, and the same update budget. It received decoded pose, not ground-truth pose.

Both architectures failed the initial qualification below. We then recorded one bounded development calibration to check a lower-variance alternative: ridge regression on the same features, with penalties $10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100$ and an unpenalized intercept. With centered features and targets, it minimizes $N^{-1}\|X_cW - Y_c\|_F^2 + \lambda\|W\|_F^2$, where N is the number of training contexts. The same grid was applied to the pose/action baseline. We retained all candidates and selected the lowest joint validation loss independently for each input family, including the original MLPs. This recovery was specified after observing the initial validation errors but before any new intervention outcome; it is not a prospectively fixed eight-candidate test set. No target, sample, world-model checkpoint or measurement threshold was changed.

Qualification and outcome. The recorded operational gate required nonnegative training-mean R^2 for every component relative to its training-mean predictor, an average of at least 0.25 over the three block-motion components, and at least 10% lower mean standardized block-motion MSE than the pose/action baseline. These thresholds qualify a proposed measurement; they are not a statistical equivalence or noninferiority test. For component j, the reported quantity is

$$R^2_{\text{train mean},j} = 1 - \frac{\sum_{i\in\text{val}}(\hat{v}_{ij} - v_{ij})^2}{\sum_{i\in\text{val}}(v_{ij} - \bar{v}_{\text{train},j})^2}. \tag{21}$$

It uses the fixed training mean, not the validation mean. All values below are validation-selected descriptive metrics, without a new held-out effect estimate.

The selected full-feature readouts have block-motion MSE ratios of 1.297 and 1.129 relative to their selected pose/action baselines, rather than the required ratio at most 0.9. Neither selected

Table 46: **Selected velocity measurements fail qualification.** Component-wise $R^2_{\text{train mean}}$. Ridge with penalty 0.1 is selected for full latent history/actions in both architectures; the pose/action baselines select their MLP. Negative angular values indicate worse error than the fixed training-mean predictor.

Predictor	Readout inputs	Agent v_x	Agent v_y	Block v_x	Block v_y	Block ω
Transformer	Latents/actions	0.986	0.981	0.208	0.335	-0.084
Transformer	Pose/actions	0.990	0.991	0.382	0.414	0.206
GRU	Latents/actions	0.984	0.985	0.220	0.335	-0.133
GRU	Pose/actions	0.992	0.992	0.253	0.415	0.059

readout, nor any retained velocity candidate, passes all three gates. Independent NumPy evaluation reconstructs all 30 retained readout/baseline fits, their predictions, normalizers, selection decisions and qualification outcomes.

Following this study's recorded calibration stop rule, we did not generate confirmation recipients or run a velocity-preserving intervention matrix within this initial qualification. Its failure did not establish that velocity was absent from the representation or unlearnable by other readouts. The original archive retains its development configurations, every fitted candidate, observed-feature/label arrays and independent reconstruction code. It reproduces this initial measurement-limited outcome; it is separate from the later study in Appendix C.

Q Complete branch reporting and diagnostic context

This supplement reanalyzes accepted arrays at their original action-25 endpoint. It adds no training, rollouts, recipients, model selection or changes to the original primary families. The reporting choices were fixed after the original outcomes and are explicitly post-hoc. All fixed models and reference streams are averaged within recipient before resampling. The additional results have a separate audit and reproduction supplement; the original evidence archives are unchanged.

Q.1 All objectives in the core matching family

Table 47 uses all six groups, 256 recipients, four reference streams and the original six-direction matching family. Its latent actual–free contrast is an added exploratory comparison with a 95% interval, using the existing 20,000 recipient draws. The other five effects retain their original nine-primary-family intervals. Full reset replaces the complete token by its observed encoding. It changes the current readout and displacement and is a reference, not a norm-matched control.

Objective	Free	Actual	Donor	Full reset
Latent	6094.2	5771.6	6509.6	5224.0
Coordinate teacher	8448.1	7662.0	8305.8	6505.9
Physical labels	8583.8	7829.3	8381.0	6496.6

Objective	Actual–free	Actual–donor
Latent	$-322.6\ [-419.4, -226.6]^{\mathrm{E}}$	$-737.9\ [-938.8, -546.9]$
Coordinate teacher	$-786.1\ [-998.8, -609.0]$	$-643.8\ [-811.9, -488.3]$
Physical labels	$-754.5\ [-926.0, -607.8]$	$-551.7\ [-699.6, -413.5]$

Table 47: Complete A-core reporting with the original roster, score and matching family. Top: mean terminal block-position MSE. Bottom: paired effects and intervals; E marks the added exploratory 95% interval. All other intervals are the original 99.444% primary intervals. These values cannot be substituted for the twelve-direction dual study or the separate native bank.

Q.2 Separating actual and donor responses to constraints

The source benefit is donor minus actual error. A change in that source benefit can result from either branch, so it is not by itself the loss of actual correction benefit. Table 48 separates both branches within the same twelve-direction matching family. Its eight post-hoc paired comparisons use one exploratory family with 99.375% intervals and the original 20,000 recipient draws. For either reference constraint, actual's joint-minus-reference change minus donor's joint-minus-reference change equals the original reference-minus-joint source-benefit contrast. This identity is checked against the original results.

Predictor	Source	Pose + velocity minus pose-only	Pose + velocity minus pose + rotated velocity
Transformer	Actual	$88.4\ [15.0, 173.3]$	$16.2\ [-56.6, 95.8]$
Transformer	Donor	$7.4\ [-78.6, 100.5]$	$-11.7\ [-99.5, 78.8]$
GRU	Actual	$80.7\ [-35.4, 192.7]$	$-50.6\ [-185.9, 74.8]$
GRU	Donor	$102.1\ [-37.8, 244.9]$	$-54.9\ [-207.9, 91.5]$

Table 48: Post-hoc motion-constraint branch changes, in terminal block-position MSE units. Positive means joint pose and velocity constraints yield higher error for that same source. All 512 recipients, two objectives and four reference streams are retained in each architecture. These intervals do not replace the original six primary source-benefit contrasts or the six prespecified actual–free contrasts.

For the Transformer, the joint pose and velocity constraint raises actual error relative to pose-only while donor's change is unresolved. For the GRU, both changes from pose-only to joint pose and

velocity constraints have positive point estimates, and both intervals include zero. Neither architecture resolves a branch-level difference between joint pose and velocity constraints and the rotated control. The decomposition clarifies which branch contributes to a source-benefit change; it does not establish velocity-specific semantics or strengthen the original unresolved mechanism test.

Q.3 What the diagnostic adds to ordinary error measurements

Table 49 places all training conditions in one reporting frame. Free-rollout MSE measures the deployed recursive prediction. Observed-history prediction MSE replaces the histories at each step, and full reset replaces the entire insertion token. Both can change the current decoded state. Actual-guided correction preserves the specified g_A outputs; its advantage over free rollout therefore establishes an opportunity that correcting the current specified readout is not needed to obtain. The donor branch additionally tests the information source at a matched standardized displacement norm.

Objective	Train	Free	Actual	Full reset	Observed-history prediction	Observed encoding	G
Coordinate teacher	T0	7325.3	6778.4	5936.9	3504.9	3683.5	546.9
Coordinate teacher	T1	5635.6	5307.4	4763.7	3539.6	3683.5	328.2
Coordinate teacher	T2	5603.7	5220.4	4735.8	3521.6	3683.5	383.3
Physical labels	T0	7603.0	6941.2	5888.5	3239.1	3683.5	661.8
Physical labels	T1	5250.2	4886.1	4292.1	3181.9	3683.5	364.2
Physical labels	T2	5120.8	4805.8	4378.9	3215.3	3683.5	315.0

Table 49: Complete descriptive training context under g_A at action 25. The same 256 recipients, three Transformer pools and four reference streams are used throughout. Matched standardized displacement norms and action-response values remain in Appendix A. These means are not additional independent observations or new hypothesis tests.

Under g_A, T1 improves free-rollout MSE and reduces the benefit of the readout-preserving correction, while observed-encoding error is identical. It also changes action response and output variance. Thus G is the remaining correctability under a specified intervention, not an intrinsic error magnitude or an independently identified causal mediator. The present study does not validate G as a ranking rule for models with equal free-rollout MSE. Its additional information is the controlled readout-preserving contrast, which a full reset or an aggregate error gap does not impose.

Q.4 Absolute stronger-readout measurements

The following table retains all observed-encoding measurements underlying the stronger-readout qualification, separately from the rollout comparisons.

Table 50: **Stronger measurements in both validation domains.** Absolute block-position MSE on observed encodings (pixels squared); every group is retained. These are readout errors, not rollout errors or lower bounds on them. All g_A readouts also satisfy the six-output, agent, and angle gates. Complete g_B metrics and the one failed angle gate are in Appendix A.

		Expert validation		Planner validation	
Pool	Predictor	Original	g_A	Original	g_A
0	Transformer	1667.1	1167.9	3995.6	2599.9
0	GRU	1454.2	820.1	4970.4	2782.6
1	Transformer	2140.0	1296.0	4728.6	2869.7
1	GRU	1257.8	797.2	4203.8	2460.1
2	Transformer	1276.1	1015.2	4336.9	1852.6
2	GRU	994.0	546.7	3506.2	1364.0

Q.5 Paired uncertainty under the second pose readout

This additional post-hoc analysis scores the original dual-readout rollouts with g_B. It retains all three objectives, both pool 0 architectures, all 256 recipients and four reference streams, at the original action-25 endpoint, within the original twelve-direction matching family. The three actual–free contrasts form a separate exploratory family: the original 20,000 recipient draws give nominal 98.3333% intervals. Architectures and reference streams are averaged within recipient. No training, rollout, sample selection or change to the original primary family is involved.

Objective	Free	Dual actual	Dual donor	Full reset
Latent	5784.0	5617.0	5803.7	5088.8
Coordinate teacher	9020.1	8504.5	8879.3	7098.2
Physical labels	9983.0	9571.0	9795.5	7842.7

Objective	Dual actual–free [98.3333% interval]
Latent	-166.9 $[-281.7, -55.7]$
Coordinate teacher	-515.6 $[-691.7, -362.2]$
Physical labels	-412.0 $[-556.9, -277.9]$

Table 51: Complete second-readout endpoint scoring of the accepted dual-readout experiment. Top: terminal block-position MSE. Bottom: added exploratory paired intervals; all three objectives are retained.

All three paired intervals favor actual over free rollout under g_B. This readout did not train the coordinate continuations, but it did constrain the insertion and shares the encoder and fitting setup with g_A. The GRU readout that failed its angle gate is retained. These results quantify robustness to endpoint scoring within this measured system; they do not eliminate shared measurement bias. A separate statistical reproduction supplement independently reconstructs every mean and interval from the original accepted arrays.

R Evaluation with a separately fitted readout

We evaluate the original saved rollouts with g_{eval}, a separately fitted pose readout excluded from predictor training, intervention constraints, standardized displacement norms, donor assignment and branch selection. It shares the frozen encoder with g_A and g_B. No predictor, inserted token, action, matching family or horizon changes. This is a post-hoc measurement-sensitivity analysis on the original 256-recipient A/C bank, not a fresh confirmation population.

Fixed fitting and data roles. Each of the six encoder groups receives one predetermined fit of a smooth residual network. For its own standardized token $\xi = (z - \mu_{\mathrm{eval}}) \oslash \sigma_{\mathrm{eval}}$, it computes

$$h_0 = \mathrm{GELU}(W_{\mathrm{in}}\xi + b_{\mathrm{in}}), \tag{22}$$

$$h_{j+1} = h_j + \tfrac{1}{2}\big(B_j\,\mathrm{GELU}(A_j h_j + a_j) + b_j\big), \quad j = 0, 1, \tag{23}$$

$$g_{\mathrm{eval}}(z) = \nu_{\mathrm{eval}} + \tau_{\mathrm{eval}} \odot (W_{\mathrm{out}} h_2 + b_{\mathrm{out}} + W_{\mathrm{skip}}\xi). \tag{24}$$

The input has 192 dimensions, every hidden layer has width 256, and the six outputs retain the original pose order. GELU uses its exact error-function definition. The linear skip has no bias. These readout normalizers are separate from the original intervention standardization. Input and target moments use only fitting data: each domain contributes half of its population first and second moments, and each standard deviation has a 10^{-6} floor.

Each fit uses Adam with learning rate 10^{-4}, zero weight decay, $\beta = (0.9, 0.999)$ and $\epsilon = 10^{-8}$ for exactly 8,000 updates. Each batch contains 128 expert and 128 planner observations. Fitting uses FP32 with TF32 and automatic mixed precision disabled. Validation runs every 250 updates; the minimum equally weighted expert/planner six-output normalized MSE selects the checkpoint, with an exact tie resolved by the earlier step. There is one fixed initialization per group and no architecture or seed search. Selected FP32 parameters and FP64 normalizers are evaluated in FP64. Separately implemented error-function and PyTorch forwards are checked before qualification.

Expert data use 1,024 fitting, 256 validation and 256 qualification parents, with whole-parent exclusions for the original pose-readout training/validation parents and the fixed exact-image overlap checks. Roles are assigned before fitting; identical images within a role retain a deterministic first occurrence, with continuous-label aliases recorded. The planner roles contain 1,024/512/512 initial states and use predetermined banks from the motion-readout study. Their prior use is retained in the provenance record; they are not globally unseen paper data. All six selected readouts are frozen before qualification. Neither qualification nor intervention outcomes cause refitting or roster changes.

Observed-data qualification. For each domain, qualification requires g_{eval} to have at most 1.10 times g_A's error on each of four metrics: six-output normalized MSE, block-position MSE, agent-position MSE and wrapped-angle MSE. Position errors are squared Euclidean distances. Only the common six-output qualification metric uses g_A's target scales. All six readouts pass all eight checks (Table 52). This point-estimate engineering gate concerns observed encodings; it does not certify the readout's accuracy on predicted tokens.

Group	Domain	Six outputs	Block	Agent	Angle
0	Expert	0.701	0.559	0.910	0.384
0	Planner	0.611	0.521	0.806	0.402
1	Expert	0.778	0.755	0.870	0.598
1	Planner	0.714	0.630	0.830	0.568
2	Expert	0.694	0.460	0.879	0.405
2	Planner	0.656	0.546	0.802	0.483
3	Expert	0.700	0.549	0.827	0.533
3	Planner	0.685	0.562	0.818	0.507
4	Expert	0.705	0.578	0.847	0.440
4	Planner	0.657	0.589	0.784	0.455
5	Expert	0.680	0.621	0.864	0.410
5	Planner	0.748	0.710	0.847	0.557

Table 52: All qualification error ratios, g_{eval}/g_A; the gate requires every ratio to be at most 1.10. Groups 0/2/4 are the three Transformer pools; 1/3/5 are their GRU counterparts. Expert qualification has 32,161 unique frames from 256 parents in every group. Planner qualification has 6,613 frames for Transformer groups and 6,582 for GRU groups, from 512 parents in each. Ratios use the same observed frames and labels for both readouts.

Complete comparisons. Rescoring retains all 18 core model cells, 12 dual model/constraint combinations and 18 matched training cells, with 256 recipients, four reference streams, all four branches and all five horizons. Action 25 remains the endpoint. The seven readout comparisons and two PointMaze comparisons occupy a separate nine-slot family with nominal $1 - 0.05/9 = 99.444\ldots\%$ intervals. The eight training comparisons form a separate secondary family with 99.375% intervals. Both use 20,000 paired recipient-bootstrap draws after averaging fixed groups and streams within recipient; no slot is reallocated. These analysis choices were fixed before rescoring, while the recipient population is reused.

All seven primary intervals favor actual guidance. Thus future benefit is also measured by a readout that did not define the correction. The three actual–donor contrasts retain source sensitivity; the four coordinate-objective actual–free contrasts retain useful correction. This distinction is unchanged by the choice of evaluator.

Insertion changes under evaluation-only measurement. The intervention preserves the specified g_A outputs, or both g_A and g_B in the dual comparison; it does not constrain g_{eval}. Table 54 reports its insertion block-position displacement and all four branches' insertion errors. For the two coordinate objectives, actual guidance changes the block-position estimate by about 30 pixels RMS in the core experiment and lowers its insertion error. Fixed outputs under one pose readout can therefore coexist with improved pose estimates under another. These are changes in decoded estimates, not changes to the physical recipient trajectory. The insertion and terminal error reductions do not identify how much of the later benefit is mediated by the insertion estimate.

ID	Objective	Comparison	Difference [interval]
A. Seven primary comparisons: 99.444… % intervals			
GE01	Coordinate teacher	Core: actual–free	-1175.7 $[-1458.5, -909.5]$
GE02	Physical labels	Core: actual–free	-1053.5 $[-1278.2, -850.3]$
GE03	Latent	Core: actual–donor	-858.2 $[-1082.1, -653.9]$
GE04	Coordinate teacher	Core: actual–donor	-837.6 $[-1092.7, -596.1]$
GE05	Physical labels	Core: actual–donor	-760.7 $[-981.1, -552.2]$
GE06	Coordinate teacher	Dual: actual–free	-570.1 $[-818.2, -345.0]$
GE07	Physical labels	Dual: actual–free	-571.1 $[-825.6, -324.8]$
B. Eight secondary training comparisons: 99.375% intervals			
CE01	Coordinate teacher	Free-rollout MSE: T1–T0	-1466.0 $[-2055.1, -917.3]$
CE02	Coordinate teacher	G: T1–T0	-186.6 $[-405.0, 19.0]$
CE03	Coordinate teacher	Free-rollout MSE: T2–T1	-1344.7 $[-1801.1, -860.4]$
CE04	Coordinate teacher	G: T2–T1	-81.0 $[-278.9, 121.1]$
CE05	Physical labels	Free-rollout MSE: T1–T0	-2486.4 $[-3119.2, -1886.2]$
CE06	Physical labels	G: T1–T0	-281.7 $[-488.4, -75.4]$
CE07	Physical labels	Free-rollout MSE: T2–T1	-1221.4 $[-1621.7, -824.0]$
CE08	Physical labels	G: T2–T1	-213.1 $[-397.8, -27.6]$

Table 53: Complete comparisons under g_{eval}, in pixels squared. For MSE contrasts, negative values favor the left branch or condition; for G, they mean less remaining correctability. Core actual–free reductions are 8.91% and 8.68%; dual reductions are 4.89% and 5.08%. Dual rows preserve both g_A and g_B and retain their original twelve-direction matching family. For training, $G = \mathbb{E}[L^{\text{free}} - L^{\text{act}}]$ uses g_{eval} scoring and the original six-direction matching family. All contrasts, including intervals crossing zero, are retained.

Study	Objective	RMS change	Free	Actual	Donor	Full reset
Core	Latent	28.08	1819.0	1948.0	2776.8	1194.1
Core	Coordinate teacher	30.21	2984.1	2150.7	2923.9	1194.1
Core	Physical labels	30.09	2961.9	2081.0	2802.5	1194.1
Dual	Latent	23.79	1860.7	2001.2	2437.7	1274.5
Dual	Coordinate teacher	25.07	3176.0	2463.6	2939.1	1274.5
Dual	Physical labels	25.40	3287.1	2474.9	2970.9	1274.5

Table 54: Descriptive insertion results under g_{eval} at action 5. RMS change is the square root of the mean squared Euclidean difference between actual and free block-position readouts, in pixels. The last four columns are block-position MSE against the recipient's true insertion pose, in pixels squared. Core averages six groups; dual averages the two Pool 0 architectures with both original pose readouts preserved. All objectives and branches are retained.

Training conclusions depend on the evaluation readout. The four original T2–T1 contrasts remain unresolved under g_A. Under g_{eval}, T2 reduces free-rollout MSE for both objectives (CE03 and CE07). Its change in G is negative for physical labels (CE08) and unresolved for coordinate teacher (CE04). T1–T0 likewise reduces free-rollout MSE for both objectives, but its change in G is resolved only for physical labels under this evaluator (CE02 and CE06). Thus the original conclusion about no resolved additional anchoring benefit is specific to g_A, and reductions in remaining correctability are specific to the objective and readout. These comparisons do not establish a readout-independent measure of latent error or a model-selection rule.

The additional 18 full-feasible C cells retain each source's own displacement norm and are descriptive only; they do not replace matched training contrasts. Complete per-recipient arrays retain all branch curves, observed-encoding references, signed and normalized six-output insertion changes, and position/angle displacements. The controlled claim remains useful feedback correction at fixed specified readouts; neither qualification nor rescoring establishes equal physical states or isolates a missing dynamical variable.

S External PointMaze protocol and complete results

Fixed model and temporal interface. We use the official DINO-WM PointMaze checkpoint (Zhou et al., 2024), whose stored epoch is 10. Task and checkpoint selection preceded PointMaze outcome evaluation; all model parameters remain frozen. One native control advances five MuJoCo substeps of 0.01 seconds; one prediction block contains five native controls. A trajectory spans absolute control indices 0–35. Images and proprioception at indices 0, 5 and 10 form the observed history, with prediction origin at 10. We intervene at absolute index 15, or relative control 5, and score relative controls 5, 10, 15, 20 and 25 (absolute indices 15–35). The rollout retains the native three-token context and terminal prediction indexing.

Inputs and channels. RGB inputs follow the official preprocessing: division by 255, resize and center crop to 224, and channel normalization with mean and standard deviation 0.5, followed by the encoder's resize to 196. The frozen encoder produces 196 visual patches with 384 channels. The readout and projection use all 75,264 visual coordinates without pooling. Native actions use scale 1 and are standardized once; observed proprioception contains two positions and two velocities. Deployment statistics reproduce the official FP32 mean and sample standard deviation over all 200,000 valid publisher frames, before the upstream split. Readout normalization is separate and uses only its fitting data.

Position readout and qualification. A fixed 75,264–256–256–2 network with two ReLU hidden layers predicts position from visual coordinates alone. A frozen episode permutation assigns 1,600 publisher episodes to fitting and 400 to validation, sampling 10 and 20 evenly spaced frames per episode, respectively. Deterministic image deduplication, including exclusion of fitting images duplicated in validation, leaves 15,999 fitting and 8,000 validation frames. FP64 fitting-data means and population standard deviations standardize visual inputs and position targets, with scale floors 10^{-6} and 10^{-3}.

Adam uses learning rate 10^{-4}, zero weight decay, batch size 128 and 3,000 updates. Validation every 250 updates selects minimum normalized two-output MSE, with earlier exact ties preferred; the selected update is 2,500. All 12 checkpoint weights are retained. Qualification requires each coordinate's $R^2 = 1 - \mathrm{MSE}_{\mathrm{model}}/\mathrm{MSE}_{\mathrm{constant}}$ to reach 0.5 on both publisher validation and a separate 256-trajectory simulator bank. The constant is the fitting-data position mean. The simulator bank supplies eight observed frames per trajectory. Coordinate R^2 values are $(0.997865, 0.997502)$ on validation and $(0.997045, 0.996892)$ on simulator qualification; squared Euclidean position errors are 0.004475 and 0.006146 in native position units squared. Independent FP64 reconstruction checks validation predictions. Qualification does not select another head or certify physical-state equality on predicted visual features.

Banks and isolation. The unchanged native PointMaze simulator supplies 64 development recipients, 64 development donors, 256 qualification trajectories, and 256 confirmation recipients with 256 separate donors. Role-specific SHA256 namespaces with root seed 20260922 fix simulator and action seeds. Official initialization samples valid initial states; its returned goal is unused. Seven

independent two-dimensional controls sampled uniformly from $[-1, 1]$ are each repeated five times and cast to FP32. A fixed permutation assigns donors. Two exact replays verify state, proprioception, RGB and controls. Neither admission nor assignment uses prediction errors or intervention responses.

The isolation audit covers these 896 trajectories, two authenticated technical trajectories, retained readout inputs, and 130,000 comparable publisher state–action windows. It finds no audited seed, initial-state or full-path identity collision; publisher comparisons use canonical FP32 metadata. No retained readout-input image matches the new banks. However, 22 cross-role identical-image groups remain, so complete frame isolation is not established. All cases and future frames are retained; image coincidences do not trigger replacement or deletion. Historical world-model episode assignment, publisher generating seeds and DINO pretraining membership remain unresolved. The new head split is not claimed to be held out from world-model training.

Intervention and continuation. Four branches are fixed: free rollout, actual guidance, donor guidance and full visual reset. Guidance uses only the recipient's or assigned donor's visual encoding at absolute index 15. After native known-action replacement, two full-dimensional QPs find the closest actual- and donor-guided displacements preserving the position readout and its original ReLU activation region. The two directions share their minimum standardized displacement norm. FP64 OSQP uses absolute and relative tolerances 10^{-9}, 20,000 maximum iterations, polishing and SVD cutoff 10^{-10}. A fixed numerical shrink sequence 2^{-i}, $i = 0, \ldots, 8$, applies jointly. Independent FP64 checks of the actual FP32 inserted token require normalized output and region residuals at most 10^{-6}, with norm absolute and relative tolerances 10^{-6}. Valid zero norms remain; numerical failure blocks acceptance rather than becoming a zero correction.

At insertion, earlier history and predicted proprioceptive and known-action embeddings are unchanged. Subsequent proprioceptive embeddings evolve separately through each branch's native predictions; aligned known actions remain fixed. No further observation is injected. Full visual reset replaces only the current visual encoding and is a descriptive comparison, without readout or norm preservation.

Endpoints and inference. The primary endpoint is squared Euclidean position error at relative control 25, in native position units squared without division by two. Actual-minus-free and actual-minus-donor occupy two reserved slots in the nine-contrast family containing seven independent-readout comparisons. Paired recipient-level bootstrapping uses 20,000 draws and two-sided $99.444\ldots\%$ percentile intervals. Both upper bounds must be below zero to support useful source-related correction. All four branches and five horizons are retained, with descriptive 95% intervals for complete curves. Fresh PointMaze recipients remain separate from the reused PushT recipients and historical native results. Analysis requires all 256 recipients and numerical gates, without sample additions, case filtering, checkpoint substitution, horizon changes or alpha reallocation.

Complete numerical acceptance and findings. All 256 recipients and 512 actual/donor insertions pass; no family has a zero norm, numerical backoff or missing case. Maximum normalized output deviation is 4.66×10^{-8} and maximum region violation is 9.15×10^{-7}, below the frozen 10^{-6} tolerance. Complete saved-token checks reconstruct every branch and horizon; an independent implementation reproduces both primary contrasts, all branch curves and all 20,000 bootstrap draws.

Both primary intervals span zero (Table 55). The actual–free relative interval ranges from a 9.6% reduction to a 15.0% increase; the actual–donor interval ranges from an 18.4% reduction to a 10.7% increase. Thus this fixed checkpoint, action distribution and endpoint confirm neither useful correction nor source sensitivity. This is not an equivalence result. Table 56 retains all four branches at all five horizons, including the unconstrained full visual reset. No earlier-horizon contrast replaces the frozen primary endpoint.

Table 55: Both PointMaze primary contrasts at relative control 25, in native position units squared. Changes are actual minus comparator; negative values favor actual guidance. Paired recipient-bootstrap intervals use $99.444\ldots\%$ coverage within the nine-slot family. All 256 recipients are retained; intervals spanning zero do not establish equivalence.

Comparison	Actual MSE	Comparator MSE	Change	Interval
Actual − free	0.119259	0.116572	0.00268692	$[-0.0107516, 0.0184502]$
Actual − donor	0.119259	0.125025	−0.00576584	$[-0.0237829, 0.0127212]$

Table 56: Complete PointMaze error curves for all four branches and five relative native-control horizons. The endpoint is squared Euclidean position error in native position units squared, without division by two. Intervals are descriptive 95% recipient-bootstrap intervals, not additional primary tests. Full visual reset replaces visual features only.

Branch	Relative control	Position MSE	95% interval
Free	5	0.0138689	$[0.0114445, 0.016681]$
Free	10	0.0300273	$[0.0189812, 0.0477029]$
Free	15	0.0679895	$[0.0332377, 0.129122]$
Free	20	0.0800448	$[0.0486491, 0.13088]$
Free	25	0.116572	$[0.0840407, 0.155168]$
Actual	5	0.0138689	$[0.0114445, 0.016681]$
Actual	10	0.0296824	$[0.0188369, 0.046939]$
Actual	15	0.0666107	$[0.0319104, 0.127961]$
Actual	20	0.0789682	$[0.0476149, 0.129379]$
Actual	25	0.119259	$[0.0842427, 0.161557]$
Donor	5	0.0138689	$[0.0114445, 0.016681]$
Donor	10	0.0333718	$[0.021168, 0.0505538]$
Donor	15	0.0692911	$[0.0368092, 0.124372]$
Donor	20	0.0896269	$[0.0560542, 0.141278]$
Donor	25	0.125025	$[0.0899141, 0.166564]$
Full visual reset	5	0.00585695	$[0.00509769, 0.00666738]$
Full visual reset	10	0.0296237	$[0.0188275, 0.046773]$
Full visual reset	15	0.0670242	$[0.0321267, 0.128637]$
Full visual reset	20	0.0786445	$[0.0474263, 0.128971]$
Full visual reset	25	0.118609	$[0.0835182, 0.160731]$