SAME MODEL, DIFFERENT INTERACTIONS: HIDDEN HISTORY IN MULTI-TASK LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-task learning routinely characterizes the interaction between tasks from an instantaneous snapshot of the current model, using quantities such as task losses, gradients, or pairwise gradient alignment. Such snapshots describe how tasks relate at a given moment, but treating them as sufficient to determine how those relations evolve under continued training is an assumption, not a guarantee. We show it fails in the strongest possible sense: for any current model, one can construct distinct optimization states that implement it exactly, are indistinguishable under any model-derived instantaneous representation, yet evolve into different future task–response maps. What is missing is therefore not a better observable but the optimization history itself. That decisive history lives in the optimizer's momentum, which carries a hidden, task-attributed record of past training rather than unstructured noise. As a consequence, the same pair of visible tasks can help each other under one hidden history and harm each other under another. More strongly, two indistinguishable states can call for different next tasks to train: we quantify this with an Aliasing Decision Gap and prove that any policy acting on the snapshot alone incurs an irreducible regret, which we confirm across model families and task suites over many seeds. Task interaction is therefore not an intrinsic property of a task pair but a history-conditioned relation carried behind an identical model and exposed only through history.

1 INTRODUCTION

Multi-task learning (MTL) is fundamentally concerned with how tasks interact (Kendall et al., 2018; Zamir et al., 2018). Prevailing approaches characterize task interaction from the current model state, forming an instantaneous picture of how tasks relate and using it to guide how they should be optimized next (Chen et al., 2018; Yu et al., 2020; Li et al., 2024). This move is more consequential than choosing a specific interaction metric—it implicitly treats the current snapshot as an identity for the task-interaction state itself. We challenge the assumption that an instantaneous task snapshot can serve as the state of task interaction. The snapshot can be accurate about the present and still fail to identify the state that governs what happens next. This raises our central question:

Does instantaneous task geometry identify future task-response relations and next-task decisions in stateful multi-task learning?

This snapshot-based identification fails exactly where description becomes prediction: a single instantaneous task geometry can admit multiple incompatible futures. We show that snapshot-equivalent training states can develop different task-response relations under the same subsequent optimization. Task interaction is therefore not, in general, a function of the current geometry alone. This is not a measurement error, but an identifiability failure. The snapshot acts as a many-to-one representation of the training process, collapsing states that are indistinguishable in the present but dynamically inequivalent in what they imply next. Any geometry-only account must assign such states the same future, even when their futures disagree.

What distinguishes these aliases is not merely that the optimizer remembers the past, but that the hidden state records which tasks produced it. By intervening on optimizer-carried state while holding the instantaneous task snapshot fixed, we show that historical state can causally alter subsequent task responses. More importantly, this hidden difference can be resolved by past-task provenance: contri-

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.

butions written by different tasks are not interchangeable, and selectively changing one provenance component changes the future response. The missing state is therefore not generic optimization memory, but task-structured history carried across training.

The deeper implication is that task interaction is not an intrinsic property of a task pair. Knowing which task is trained next and which task is evaluated is still insufficient to determine their interaction: the effect also depends on which past task wrote the hidden historical state that the future update acts upon. This turns the usual pairwise relation into a history-conditioned one, $[k_{\text{past}} \times i_{\text{future}} \times j_{\text{response}}]$. Under this view, a pairwise interaction matrix is only a projection of a higher-dimensional dynamical relation in which training history has been marginalized out. The same pair of tasks can therefore express different, even qualitatively different, interactions at snapshot-equivalent states. This distinction is consequential for control: if the interaction law itself is not identified by the snapshot, neither is the action chosen from it. We show that snapshot-equivalent states can reverse the locally optimal next-task choice under the same visible objective. Thus, the failure is not noisy prediction or insufficient capacity: the snapshot omits the hidden state variable that makes the interaction identifiable for control.

Our main contributions are summarized as follows:

- **Snapshot identifiability failure.** We challenge the basic pairwise snapshot view of task interaction in MTL, showing that instantaneous task geometry does not identify future task-response relations: exact parameter-matched states can be indistinguishable under any model-derived snapshot, while still evolving into different task-response maps.
- **History-conditioned task interaction.** We trace the missing state to a task-structured record in optimizer momentum, which decomposes by the past task that wrote it; rewriting one task's component while holding the visible model fixed changes how future tasks respond, showing that task interaction is conditioned on past-task provenance rather than pairwise snapshots alone.
- **Irreducible decision regret.** We formalize the cost of snapshot aliasing as the *Aliasing Decision Gap*, and prove that any snapshot-only task-selection policy incurs irreducible regret: aliased states can flip the helpfulness of the same visible task pair and induce incompatible locally optimal next-task choices. The resulting gap persists beyond the momentum horizon and against strong snapshot-conditioned baselines.

2 The Current Model Does Not Identify Future Task-Response Relations

The claim of the introduction has a direct empirical test: if the current model identifies the state of task interaction, then fixing that model must also fix how the tasks respond to the same subsequent training. We state the condition a sufficient snapshot must satisfy (Sec. 2.1), remove all current-model differences and measure the residual future-response differences (Sec. 2.2), and then follow this ambiguity over one hundred continuation steps (Sec. 2.3).

Setup. We evaluate on two instruction families, SuperNI (Wang et al., 2022) and P3 (Sanh et al., 2022) plus a controlled synthetic task set (Controlled), across three backbones—Llama-3.1-8B (Grattafiori et al., 2024), Qwen2.5-1.5B, and Qwen2.5-7B (Yang et al., 2024). We also evaluate dense-prediction transfer on **NYUv2** (Silberman et al., 2012) in a cross-domain setting, described in Sec. 4.3. Details of the task sets and data splits, exact-alias construction, matched-continuation protocol, metrics, and hyperparameters are provided in Appendix A.

2.1 Current-State Sufficiency Requires a Unique Future Response

Write the full training state as $S = (\theta, \xi)$, with model parameters θ and any additional process state ξ. For future task i, let $\mathbf{r}_{i,h}(S)$ be the vector of response-task loss changes after h matched continuation steps, and $\mathbf{R}_h(S)$ be their collection over all candidate future tasks. Let $\mathcal{G}(\theta)$ be any deterministic snapshot representation—losses, gradients, gradient norms, pairwise alignments or conflicts, and richer instantaneous geometry. The snapshot is *sufficient* iff

$$\mathcal{G}(\theta_A) = \mathcal{G}(\theta_B) \quad \Longrightarrow \quad \mathbf{R}_h(S_A) = \mathbf{R}_h(S_B). \tag{1}$$

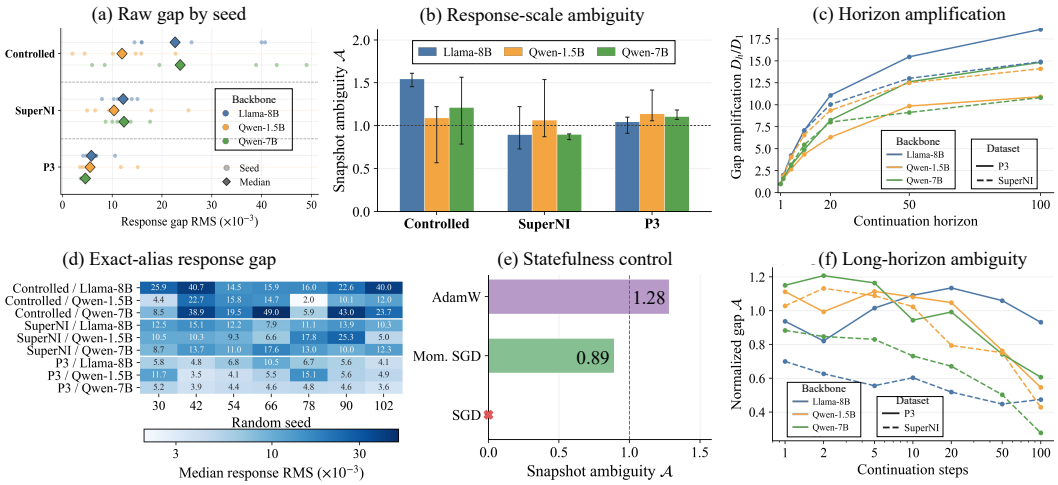


Figure 1: **The same current model admits response-scale, stateful, and persistent future ambiguity.** (a) Raw response gaps across seeds (large markers: medians). (b) One-step snapshot ambiguity $\mathcal{A}_{i,1}$ with interquartile ranges; dashed line: $\mathcal{A} = 1$. (c) Raw-gap amplification D_h/D_1 through 100 matched continuation steps. (d) Exact-alias response gaps by configuration and seed (cells: median RMS). (e) Optimizer-state control: AdamW and momentum SGD retain a gap; memoryless SGD reaches the numerical floor. (f) Long-horizon normalized ambiguity $\mathcal{A}_{i,h}$ remains substantial.

When equation 1 fails, two states share one observation $\mathcal{G}(\theta_A) = \mathcal{G}(\theta_B)$ yet demand different futures. Every snapshot-only predictor $\widehat{\mathbf{R}}_h = f_h(\mathcal{G}(\theta))$ then receives identical input for both states, so by the triangle inequality it cannot lie close to both targets at once:

$$\max_{X\in\{A,B\}} \left\| f_h(\mathcal{G}(\theta_X)) - \mathbf{R}_h(S_X) \right\|_F \;\geq\; \tfrac{1}{2} \left\| \mathbf{R}_h(S_A) - \mathbf{R}_h(S_B) \right\|_F. \tag{2}$$

The right-hand side is set by the two futures alone, so the error is *irreducible*: no added model capacity lowers it. The question is therefore not predictive accuracy but identifiability—whether the snapshot determines a unique future at all.

2.2 Same Parameters, Different Future Responses

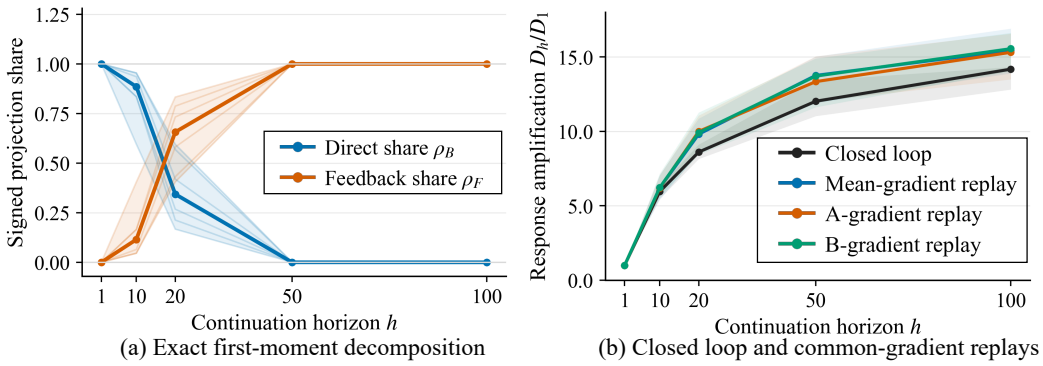
We test equation 1 in the strongest setting: rather than matching selected losses or gradients, we remove all differences in the current model. Under AdamW we construct two states $S_A = (\theta, m_A, v, c)$ and $S_B = (\theta, m_B, v, c)$ that share parameters, second moment, and optimizer clock, differing only in a hidden first-moment coordinate (used here only to place two full states behind one model). Exact parameter equality makes all snapshot-derived quantities identical; we then apply the same future task, batches, randomness, and continuation protocol.

The parameter-matched pairs still diverge. Fig. 1(a)(d) shows parameter-matched pairs producing different task-response maps under identical continuation—a direct violation of equation 1. What remains is whether this residual is merely detectable or large relative to the response itself.

The hidden gap rivals the response in magnitude. For future task i, define *snapshot ambiguity*

$$\mathcal{A}_{i,h}(S_A, S_B) := \frac{\left\| \mathbf{r}_{i,h}(S_B) - \mathbf{r}_{i,h}(S_A) \right\|_{\mathrm{RMS}}}{\frac{1}{2}\left(\|\mathbf{r}_{i,h}(S_A)\|_{\mathrm{RMS}} + \|\mathbf{r}_{i,h}(S_B)\|_{\mathrm{RMS}} \right)}, \tag{3}$$

the unresolved future distance normalized by the response magnitude, so $\mathcal{A} \approx 1$ means the two compatible futures are as far apart as the response itself. Across the six SuperNI/P3 task-world–backbone configurations (Fig. 1(b)), median one-step ambiguity ranges from 0.898 to 1.141 over 1,260 exact contexts, exceeding one in four of the six, and every context clears five times the matched numerical floor. Since the current model is identical, what differs is the future relation the observation was meant to identify.



(a) Exact first-moment decomposition (b) Closed loop and common-gradient replays

Figure 2: **A decaying hidden perturbation leaves a persistent dynamical footprint.** (a) Signed projection shares in the exact first-moment decomposition: direct carryover is replaced by newly generated trajectory feedback. (b) Closed-loop and three common-gradient replays. Bold lines aggregate the six natural-task configurations; faint lines and bands expose configuration variation.

2.3 Snapshot Insufficiency Generalizes and Persists

The gap is driven by statefulness, not the choice of optimizer. Replacing AdamW with momentum SGD preserves nonzero parameter-matched future-response gaps across SuperNI, P3, and all three backbones, whereas memoryless SGD collapses to the matched numerical floor (Fig. 1(e))—the zero point at which parameters are the complete dynamical state and identical parameters force identical futures. What matters is cross-step memory, not the specific update rule.

Rather than fading, the gap widens over training. Let $D_h = \|\mathbf{r}_{i,h}(S_B) - \mathbf{r}_{i,h}(S_A)\|_{\mathrm{RMS}}$. By the 100-step horizon, D_h exceeds its one-step value in all 180 reported contexts, with median amplification D_{100}/D_1 from 10.8 to 18.6 (Fig. 1(c)), while the normalized ambiguity stays substantial (0.277–0.931, Fig. 1(f)). Under AdamW, the first-moment difference propagates as

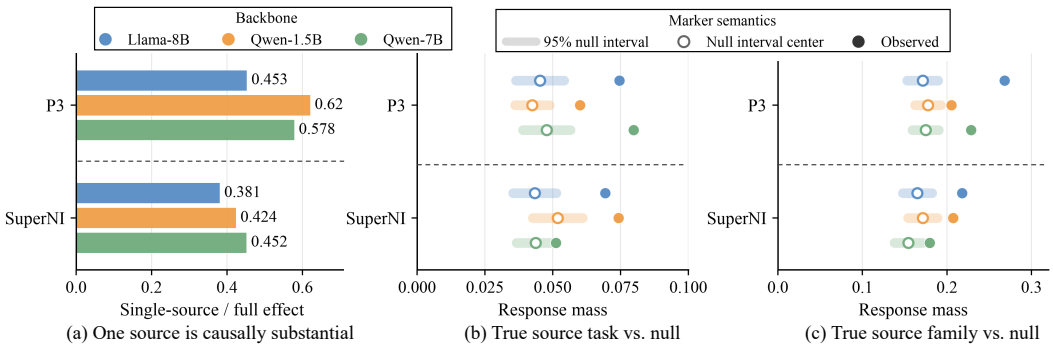
$$\Delta m_h = \beta_1^h \, \Delta m_0 + (1 - \beta_1) \sum_{r=1}^{h} \beta_1^{h-r} \Delta g_r. \tag{4}$$

The first term is the initial hidden difference inherited directly, and decays geometrically ($\beta_1^{100} \approx 2.7 \times 10^{-5}$). The second term is trajectory feedback: once the hidden state alters an early update, the two runs enter different parameters, the same future training produces divergent gradients Δg_r, and these regenerate the difference downstream. The exact first-moment decomposition confirms this mechanism directly (Fig. 2): the direct-carryover share collapses while the trajectory-feedback share takes over, yet the raw gap keeps amplifying. A rapidly decaying perturbation can thus leave an expanding, persistent footprint.

Fixing the current model therefore fails to fix the task-response relation at any horizon, and the unresolved difference is comparable to or larger than the response itself. Whether the omitted state carries structure specific to multi-task learning, rather than generic optimization history, is the question of Section 3.

3 Hidden Optimization History Retains Information About Source Tasks

Section 2 showed that the same current model can hide dynamically different training states, but a stateful optimizer could carry generic information about the past without retaining anything about *which tasks* produced it. We therefore ask whether the hidden distinction behind a fixed snapshot preserves task identity. We answer this question in three steps: we first identify a hidden-state coordinate attributable to individual tasks (Section 3.1), then intervene on one such component while holding the current model fixed (Section 3.2), and finally test whether the resulting effect stays tied to its true source (Section 3.3).



(a) One source is causally substantial (b) True source task vs. null (c) True source family vs. null

Figure 3: **Task-attributed hidden history is causally active and retains source-task structure.** All configurations use the matched five-source provenance protocol. (a) Median selective-to-complete response-magnitude ratio q: one source is causally substantial. (b, c) Observed source-task and source-family response mass (colored markers) against the central 95% source-label permutation interval (gray): the observed mass exceeds the null in every configuration.

3.1 Hidden History Can Be Exactly Attributed to Past Tasks

Once many optimization steps interleave, their contributions are generally entangled and cannot be separated by originating task. The AdamW first moment is the exception: it is a linear, exponentially weighted sum of past gradients, so the summation splits cleanly along the task that wrote each term, yielding an *exact* task-wise decomposition. This is what makes it the natural provenance coordinate. From the recursion $m_r = \beta_1 m_{r-1} + (1-\beta_1) g_r$, for a history of length H let $m_H^{(k)}$ collect the task-k terms; for two matched histories A and B,

$$\begin{aligned} m_H^{(k)} &= (1-\beta_1) \sum_{r:a_r=k} \beta_1^{H-r} g_r, \\ \Delta m^{(k)} &:= m_B^{(k)} - m_A^{(k)}, \qquad m_B - m_A = \sum_k \Delta m^{(k)}. \end{aligned} \tag{5}$$

The last equality turns the hidden first-moment difference into task-indexed coordinates, and closes to a relative residual below 3×10^{-16} across the evaluated configurations. Each $\Delta m^{(k)}$ is therefore not a post hoc label on an unstructured vector, but exactly the part of the between-history difference written by past task k.

3.2 A Single Task-Attributed Component Can Causally Alter Future Responses

Equation 5 gives a clean intervention target. From the matched state $S_A = (\theta, m_A, v, c)$ we insert only source-k history, $S_A^{[k]} = (\theta, m_A + \Delta m^{(k)}, v, c)$, with the complete difference $S_A^{[\text{all}]} = (\theta, m_B, v, c)$ as reference; both leave θ, v, the optimizer clock—and hence every current-model quantity—unchanged. After the same future task i, we measure $\mathbf{e}_{i,k} = \mathbf{r}_i(S_A^{[k]}) - \mathbf{r}_i(S_A)$ against $\mathbf{e}_i^{[\text{all}]}$ through the magnitude ratio

$$q_{k,i} := \frac{\|\mathbf{e}_{i,k}\|_{\text{RMS}}}{\|\mathbf{e}_i^{[\text{all}]}\|_{\text{RMS}}}, \tag{6}$$

a calibration of intervention magnitude rather than an additive share. Under the matched five-source protocol, the median $q_{k,i}$ ranges from 0.381 to 0.621 across the six SuperNI/P3 configurations (Fig. 3(a): changing the history attributed to a *single* past task moves future responses substantially, while every current-model quantity is held fixed.

3.3 The Causal Effect Retains a Signature of Its Source Task

We next ask whether each effect $\mathbf{e}_{i,k}$ stays tied to its source label. We measure how much of its absolute mass falls on its source task and family, and compare both to a response-preserving

permutation null—the distribution expected if the effects carried no source-specific information. The null holds the realized effect vectors fixed and only permutes their source labels (Appendix A), destroying the effect–source correspondence while leaving the entire response geometry intact.

The signature is not task-local memory: only 5.1%–8.0% of the absolute effect lies on the source task and 18.0%–26.8% within its family, so most of the consequence is distributed elsewhere. Yet it is not source-agnostic either—observed source-task and source-family mass exceeds the null in every configuration (Fig. 3(b)(c); weakest task-level $p = 0.0171$, all others $p < 10^{-4}$, all family-level $p \leq 0.0012$). Provenance thus gives hidden history an origin and biases where its effect appears, but does not fix a self-contained downstream meaning: that meaning is completed only when subsequent optimization acts on the component. Section 4 examines how the future training task reads task-attributed history and turns it into a realized task interaction.

4 Hidden History Changes Task Interaction and Local Decisions

Section 3 gave hidden history an origin; this section shows how that history becomes an interaction. A component written by one past task is transformed by the task trained next and expressed through the response of another task. We formalize these three roles (Sec. 4.1), show that future context reshapes the response profile of the same component rather than merely rescaling it (Sec. 4.2), that history can reverse whether a fixed visible task pair is helpful or harmful (Sec. 4.3), and that two snapshot-equivalent states can require incompatible next-task decisions (Sec. 4.4).

4.1 Hidden History Adds a Third Role to Task Interaction

A historical effect is produced by three distinct operations: past task k writes a component into the hidden optimization state, future task i acts on it through the subsequent update, and response task j reads the resulting parameter displacement through its loss. Recall the selective state $S_A^{[k]}$, obtained by inserting the component $\Delta m^{(k)}$ attributed to past task k while leaving the current model unchanged. After one continuation step on future task i,

$$\mathbf{e}_{i,k} := \mathbf{r}_i(S_A^{[k]}) - \mathbf{r}_i(S_A), \qquad E_{j,i,k} := [\mathbf{e}_{i,k}]_j. \tag{7}$$

Because $S_A^{[k]}$ and S_A share their current parameters, second moment, optimizer clock, and task-i gradient, a first-order expansion (derived in Appendix B.1) gives

$$E_{j,i,k} = -\eta\kappa_c \nabla_\theta L_j\big(\theta_i^+(S_A)\big)^\top D_i \Delta m^{(k)} + O\big(\|\delta\theta_{i,k}\|^2\big)\,, \tag{8}$$

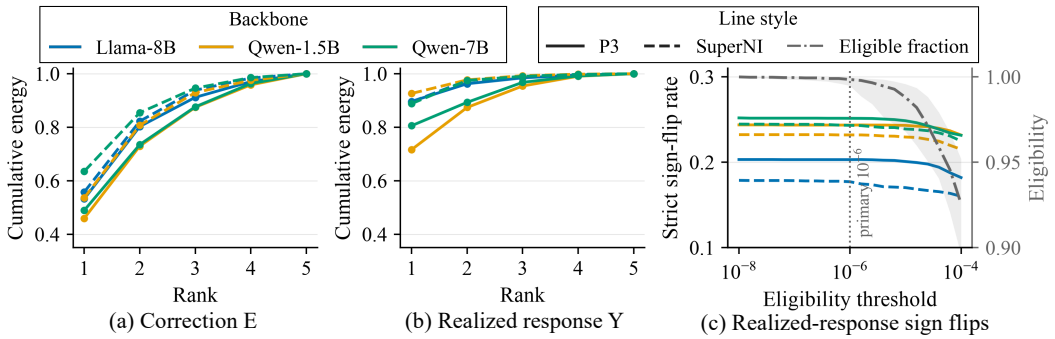
where D_i is the diagonal update operator induced by the future task-i gradient and κ_c collects the first-moment and bias-correction factors set by the optimizer clock; the finite-horizon analogue replaces D_i with the Jacobian of the matched continuation. Each role enters through a different object: $\Delta m^{(k)}$ supplies the historical direction, D_i transforms it under the future update, and $\nabla_\theta L_j$ converts the displacement into an observed response. The indices $(k_{\text{past}}, i_{\text{future}}, j_{\text{response}})$ are therefore not interchangeable labels on one pairwise effect, but distinct operators in the process that produces it.

4.2 Future Tasks Reshape Historical Effects, Not Merely Rescale Them

If provenance fixed a single downstream meaning, the response vectors produced from one component across future tasks would differ only in strength,

$$\mathbf{e}_{i,k} = a_{i,k}\mathbf{u}_k, \tag{9}$$

leaving the response matrix $\mathbf{E}^{(k)} = [\mathbf{e}_{1,k}, \ldots, \mathbf{e}_{|\mathcal{I}|,k}]$ rank one even when the gains $a_{i,k}$ are signed. Holding provenance k fixed and varying only the future task i, we measure directional agreement (cross-future cosine), magnitude variation (amplitude max/min), and the squared response magnitude captured by one fixed profile (rank-one share). Cross-future cosine is only 0.020–0.165 while the effect magnitude varies by 1.679–$2.045\times$, and the leading singular component captures just 0.460–0.635 of the squared response magnitude (Fig. 4(a)(b)). A single provenance-specific direction with an arbitrary signed gain cannot recover the remaining variation: future context selects the response profile through which a component is expressed, so the meaning of hidden history is completed by the training that follows rather than stored as one fixed downstream profile.



(a) Correction E (b) Realized response Y (c) Realized-response sign flips

Figure 4: **The same historical component is reshaped rather than merely rescaled, and its realized sign depends on history.** (a,b) Rank-wise cumulative energy of the correction $E^{(k)}$ and realized response Y. (c) Strict realized-response sign-flip rate (IFlip, left) and eligible fraction (right) versus threshold; the dotted line marks the primary $\tau = 10^{-6}$.

Table 1: **History-conditioned interaction extends to NYUv2 dense prediction.** Separation, fraction $>100\times$ the floor, and IFlip (also with the largest-effect condition dropped).

Setting	Median sep.	Min. sep.	$>100\times$	IFlip	$\text{IFlip}_{\text{drop}}$
Primary	3536×	58.4×	91.4%	49.8%	48.8%
Shared Adam burn-in	441×	67.0×	97.1%	35.6%	34.6%
Eval-128 panel	6504×	19.7×	88.6%	51.1%	50.2%
Independent base seed 42	38415×	110.3×	100.0%	53.3%	52.0%
Independent base seed 54	36200×	117.0×	100.0%	52.4%	51.8%

4.3 Hidden History Can Flip a Fixed Task Pair from Helpful to Harmful

Pairwise MTL quantities—gradient alignment, conflict, affinity, transfer—describe whether training task i helps or harms task j at a given state. We test whether that qualitative sign survives when the current model and the visible pair (i, j) are held fixed and only the inserted component k varies. Writing $Y_{j,i}^{[k]} = [\mathbf{r}_i(S_A^{[k]})]_j$ for the realized loss change of response task j (negative means help, positive means harm), we count a strict interaction flip when two eligible sources place the responses on opposite sides of $\pm\tau$, with $\tau = 10^{-6}$ as the primary threshold (eligibility rule and threshold sensitivity in the appendix). Strict helpful–harmful reversals occur in 18.5%–25.9% of eligible cases and persist across the threshold range (Fig. 4(c), while the current model, future task i, response task j, and continuation protocol all remain unchanged. The realized sign therefore belongs to the full training state: a visible pair (i, j) is only one history-specific slice of a larger relation,

$$\boxed{k_{\text{past}} \times i_{\text{future}} \times j_{\text{response}}}. \tag{10}$$

The same reversal appears in dense vision. We repeat the exact-alias intervention on NYUv2 across semantic segmentation, depth, and surface-normal prediction, covering 35 conditions and 315 ordered training–response interactions (Table 1). Changing only hidden history keeps the effect above the numerical floor and reverses the helpful–harmful sign in 49.8% of cases, robustly across thresholds, evaluation panels, and base checkpoints. Thus, history-conditioned interaction is not specific to LLM instruction tuning but also appears in dense-prediction multi-task learning.

4.4 Snapshot-Equivalent States Can Require Incompatible Next-Task Decisions

The decision-level claim is stronger than response prediction: differing futures can reorder the candidate actions, so we measure action compatibility directly. For a fixed horizon H and response-task objective j, let $C_{H,j}(S, i)$ be the loss of task j after selecting candidate future task i from state S, with state-aware optimum $a_{H,j}^*(S) = \arg\min_i C_{H,j}(S, i)$. For an exact snapshot-alias pair

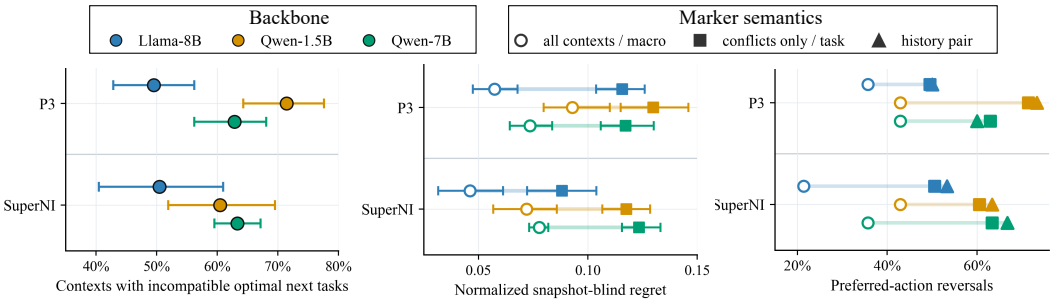


Figure 5: **Snapshot-equivalent states require incompatible next-task actions and impose nonzero snapshot-blind regret.** (a) Fraction of contexts with no common optimal action ($\Delta_{H,j} > 0$). (b) Normalized snapshot-blind regret of the best common action, overall and conditional on a conflict; bars are seed-bootstrap 95% intervals. (c) Preferred-action reversal rate under macro, task-conditioned, and history-pair aggregation: aggregating across tasks hides reversals.

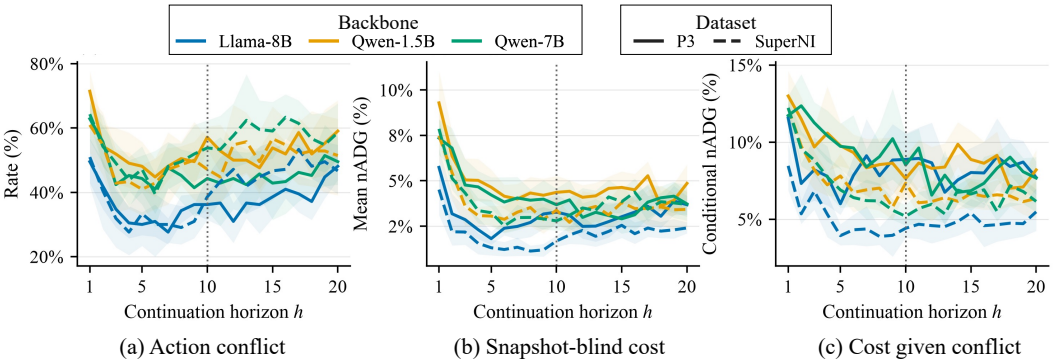


Figure 6: **Snapshot-blind conflict and regret persist across the continuation horizon.** (a) Action-conflict rate, (b) snapshot-blind cost (MeanNReg), and (c) cost conditional on conflict for $h = 1$–20. Bands are seed-bootstrap 95% intervals.

(S_A, S_B) a full-state policy may choose a different action for each state, whereas a snapshot-only rule must commit to one common action. The *Aliasing Decision Gap* is the cost of that restriction:

$$\Delta_{H,j}(A,B) := \min_i \frac{C_{H,j}(S_A,i) + C_{H,j}(S_B,i)}{2} - \frac{1}{2}\left[\min_i C_{H,j}(S_A,i) + \min_i C_{H,j}(S_B,i)\right] \geq 0, \tag{11}$$

the best cost under a forced common action minus the cost under separate choices, so $\Delta_{H,j} > 0$ exactly when no common action is optimal for both states.

Theorem 4.1 (Snapshot-Blind Impossibility)**.** *Let S_A and S_B satisfy $\mathcal{G}(\theta_A) = \mathcal{G}(\theta_B)$. For any policy π_j whose action depends only on the common current-model representation and the response-task objective j,*

$$\frac{1}{2}\sum_{X\in\{A,B\}} \mathbb{E}_{i\sim\pi_j(\mathcal{G}(\theta_X))}\left[C_{H,j}(S_X,i) - \min_a C_{H,j}(S_X,a)\right] \geq \Delta_{H,j}(A,B). \tag{12}$$

Equality is attained by the common action minimizing the first term of equation 11.

Proof. See Appendix B.2.

Whenever $\Delta_{H,j} > 0$, no added capacity in a snapshot-restricted policy can recover the missing distinction.

Across the six configurations, no action is locally optimal for both hidden states in 49.5%–71.4% of contexts (Fig 5(a), and the best common action still incurs normalized regret 0.0461–0.0929,

bounded away from zero across seeds and robust to task-conditioned versus macro aggregation (Fig 5b,c). This persists as the horizon grows past the ≈ 10-step first-moment memory (Fig 6), so it is no one-step artifact. Hidden history is thus the missing state variable: the same model, carried forward under a different history, meets a different future, a different interaction, and a different best move.

Snapshot-conditioned baselines. We add the weighted updates prescribed by DB-MTL and PiKE (Lin et al., 2026; Li et al., 2025) as candidate shared actions at the common snapshot. The expanded nADG remains strictly positive in all 24 combinations (95% CIs; Appendix C.5, Table 12).

5 Related Work

Task interaction and multi-task optimization. MTL has long studied which tasks transfer, cooperate, or interfere through task relatedness, transfer structure, and task grouping, including Taskonomy (Zamir et al., 2018), task-grouping methods (Standley et al., 2020; Jeong & Yoon, 2025), negative-transfer analysis (Wu et al., 2020; Li et al., 2023), and gradient-based task-affinity estimation (Li et al., 2024). Optimization-centric methods instead balance losses or learning rates, e.g., uncertainty weighting (Kendall et al., 2018), GradNorm (Chen et al., 2018), DWA (Liu et al., 2019), DB-MTL (Lin et al., 2026), RLW (Lin et al., 2022), and FAMO (Liu et al., 2023), formulate MTL as multi-objective optimization, e.g., MGDA (Sener & Koltun, 2018), or modify conflicting gradients through PCGrad (Yu et al., 2020), GradVac (Wang et al., 2021), IMTL (Liu et al., 2021b), CAGrad (Liu et al., 2021a), Nash-MTL (Navon et al., 2022), FairGrad (Ban & Ji, 2024), and ConsMTL (Qin et al., 2025b) or resolve conflict at the representation level (Wu et al., 2026). More recent approaches make task relations adaptive through learned task weights, training progress, excess risk, dynamic data mixtures, or learned optimization policies, including Auto-λ (Liu et al., 2022), ExcessMTL (He et al., 2024), PiKE (Li et al., 2025), NTKMTL (Qin et al., 2025a), and TaskForce (Choi et al., 2026). Yet these interaction estimates remain primarily snapshot-conditioned rather than history-resolved, leaving open whether the same visible model can realize different future task relations.

History-aware learning and training memory. Optimizers accumulate information across updates—momentum (Sutskever et al., 2013), Adam (Kingma & Ba, 2015), AdamW (Loshchilov & Hutter, 2019)—and training trajectories depend on the order of examples, through curriculum learning (Bengio et al., 2009), random reshuffling (HaoChen & Sra, 2019; Rajput et al., 2020; Gürbüzbalaban et al., 2021), and data-order studies (Dini et al., 2026). In continual learning, EWC (Kirkpatrick et al., 2016), GEM (Lopez-Paz & Ranzato, 2017), cyclic-forgetting (Evron et al., 2023; Swartworth et al., 2023), task-similarity (Hiratani, 2024), and task-order results (Li & Hiratani, 2025) show earlier tasks shape later learning and forgetting. A related line asks what remains encoded after an update: data reshapes later use (Krasheninnikov et al., 2024), order recency is recoverable from activations (Krasheninnikov et al., 2026), and SGD non-Markovianity ties causally to optimizer state (Sevetlidis & Pavlidis, 2026). Yet this history is typically read through diverging trajectories, endpoints, or forgetting, and none of it resolves history by task provenance or traces its consequences for pairwise interaction and task-selection decisions.

6 Conclusion and Discussion

We challenge a basic assumption in multi-task learning: that a model's current parameters are a sufficient state for how it interacts with future tasks—and show they are not. The missing state variable is the hidden optimization history: it decomposes by source task into causally active components, is reshaped rather than rescaled by the task trained next, and can flip a fixed visible pair between helpful and harmful or make two snapshot-equivalent states demand incompatible next-task actions under an irreducible snapshot-blind regret (Thm. 4.1).

Persistence and statefulness. Conflict and regret persist beyond the AdamW first-moment horizon (Fig. 6), while parameter-matched response gaps appear under AdamW and momentum SGD but collapse under memoryless SGD (Fig. 1).

Design implication. Whenever the aliasing decision gap is positive, no snapshot-only policy can match a full-state policy, regardless of capacity. Task selection, weighting, and data mixing should therefore condition on optimizer-carried history, especially the source components $\Delta m^{(k)}$ and their future-conditioned readout.

AI USE STATEMENT

In this work, we used generative AI tools solely to improve the grammar, clarity, and readability of the manuscript. Generative AI tools were not used to generate research ideas, data, results, analyses, conclusions, or references, and other disclosure categories are not applicable to this work. All AI-assisted revisions were reviewed and edited by the authors to ensure their accuracy and consistency with the original scientific meaning. We take full responsibility for the final content of this work.

REFERENCES

Hao Ban and Kaiyi Ji. Fair resource allocation in multi-task learning. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pp. 2715–2731. PMLR / OpenReview.net, 2024.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML 2009, Montreal, Quebec, Canada, June 14-18, 2009*, volume 382 of *ACM International Conference Proceeding Series*, pp. 41–48. ACM, 2009.

Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In Jennifer G. Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pp. 793–802. PMLR, 2018.

Wonhyeok Choi, Kyumin Hwang, Jihun Park, Kyoungmin Lee, Seunghun Lee, Jaeyeul Kim, Minwoo Choi, and Sunghoon Im. Taskforce: Cooperative multi-agent reinforcement learning for multi-task optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 36871–36880, 2026.

Luca Dini, Lucia Domenichelli, Dominique Brunato, and Felice Dell'Orletta. On the impact of pretraining data ordering in transformer encoder- and decoder-only language models. *Knowl. Based Syst.*, 342:115850, 2026.

Itay Evron, Edward Moroshko, Gon Buzaglo, Maroun Khriesh, Badea Marjieh, Nathan Srebro, and Daniel Soudry. Continual learning in linear classification on separable data. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 9440–9484. PMLR, 2023.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Mert Gürbüzbalaban, Asuman E. Ozdaglar, and Pablo A. Parrilo. Why random reshuffling beats stochastic gradient descent. *Math. Program.*, 186(1):49–84, 2021.

Jeff Z. HaoChen and Suvrit Sra. Random shuffling beats SGD after finite epochs. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2624–2633. PMLR, 2019.

Yifei He, Shiji Zhou, Guojun Zhang, Hyokun Yun, Yi Xu, Belinda Zeng, Trishul Chilimbi, and Han Zhao. Robust multi-task learning with excess risks. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pp. 18094–18114. PMLR / OpenReview.net, 2024.

Naoki Hiratani. Disentangling and mitigating the impact of task similarity for continual learning. In *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024.

Wooseong Jeong and Kuk-Jin Yoon. Selective task group updates for multi-task optimization. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025.

Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pp. 7482–7491. Computer Vision Foundation / IEEE Computer Society, 2018.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun (eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

James Kirkpatrick, Razvan Pascanu, Neil C. Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *CoRR*, abs/1612.00796, 2016.

Dmitrii Krasheninnikov, Egor Krasheninnikov, Bruno Kacper Mlodozeniec, Tegan Maharaj, and David Krueger. Implicit meta-learning may lead language models to trust more reliable sources. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pp. 25534–25559. PMLR / OpenReview.net, 2024.

Dmitrii Krasheninnikov, Richard E Turner, and David Krueger. Fresh in memory: Training-order recency is linearly encoded in language model activations. In *International Conference on Learning Representations*, volume 2026, pp. 62177–62197, 2026.

Dongyue Li, Huy L. Nguyen, and Hongyang Ryan Zhang. Identification of negative transfers in multitask learning using surrogate models. *Trans. Mach. Learn. Res.*, 2023, 2023.

Dongyue Li, Aneesh Sharma, and Hongyang R. Zhang. Scalable multitask learning using gradient-based estimation of task affinity. In Ricardo Baeza-Yates and Francesco Bonchi (eds.), *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*, pp. 1542–1553. ACM, 2024.

Zeman Li, Yuan Deng, Peilin Zhong, Meisam Razaviyayn, and Vahab Mirrokni. Pike: Adaptive data mixing for large-scale multi-task learning under low gradient conflicts. In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2025, NeurIPS 2025, San Diego, CA, USA, December 2-7, 2025 / Mexico City, Mexico, November 30 - December 5, 2025*, 2025.

Ziyan Li and Naoki Hiratani. Optimal task order for continual learning of multiple tasks. In *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*, volume 267 of *Proceedings of Machine Learning Research*. PMLR / OpenReview.net, 2025.

Baijiong Lin, Feiyang Ye, Yu Zhang, and Ivor W. Tsang. Reasonable effectiveness of random weighting: A litmus test for multi-task learning. *Trans. Mach. Learn. Res.*, 2022, 2022.

Baijiong Lin, Weisen Jiang, Feiyang Ye, Yu Zhang, Pengguang Chen, Ying-Cong Chen, Shu Liu, Ivor W. Tsang, and James T. Kwok. Dual-balancing for multi-task learning. *Neural Networks*, 195:108317, 2026.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 18878–18890, 2021a.

Bo Liu, Yihao Feng, Peter Stone, and Qiang Liu. FAMO: fast adaptive multitask optimization. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural*

Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023, 2023.

Liyang Liu, Yi Li, Zhanghui Kuang, Jing-Hao Xue, Yimin Chen, Wenming Yang, Qingmin Liao, and Wayne Zhang. Towards impartial multi-task learning. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021b.

Shikun Liu, Edward Johns, and Andrew J Davison. End-to-end multi-task learning with attention. In *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 1871–1880. IEEE, 2019.

Shikun Liu, Stephen James, Andrew J. Davison, and Edward Johns. Auto-lambda: Disentangling dynamic task relationships. *Trans. Mach. Learn. Res.*, 2022, 2022.

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 6467–6476, 2017.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

Aviv Navon, Aviv Shamsian, Idan Achituve, Haggai Maron, Kenji Kawaguchi, Gal Chechik, and Ethan Fetaya. Multi-task learning as a bargaining game. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato (eds.), *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pp. 16428–16446. PMLR, 2022.

Xiaohan Qin, Xiaoxing Wang, Ning Liao, and Junchi Yan. NTKMTL: mitigating task imbalance in multi-task learning from neural tangent kernel perspective. In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2025, NeurIPS 2025, San Diego, CA, USA, December 2-7, 2025 / Mexico City, Mexico, November 30 - December 5, 2025*, 2025a.

Xiaohan Qin, Xiaoxing Wang, and Junchi Yan. Towards consistent multi-task learning: Unlocking the potential of task-specific parameters. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pp. 10067–10076. Computer Vision Foundation / IEEE, 2025b.

Shashank Rajput, Anant Gupta, and Dimitris S. Papailiopoulos. Closing the convergence gap of SGD without replacement. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7964–7973. PMLR, 2020.

Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Arun Raja, Manan Dey, M Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal V. Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Févry, Jason Alan Fries, Ryan Teehan, Teven Le Scao, Stella Biderman, Leo Gao, Thomas Wolf, and Alexander M. Rush. Multitask prompted training enables zero-shot task generalization. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pp. 525–536, 2018.

Vasileios Sevetlidis and George Pavlidis. Process-tensor tomography of SGD: Measuring non-markovian memory via back-flow of distinguishability. In *Greeks in AI*, 2026. URL https://openreview.net/forum?id=mvBSa4kuTv.

Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *European conference on computer vision*, pp. 746–760. Springer, 2012.

Trevor Standley, Amir Zamir, Dawn Chen, Leonidas J. Guibas, Jitendra Malik, and Silvio Savarese. Which tasks should be learned together in multi-task learning? In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 9120–9132. PMLR, 2020.

Ilya Sutskever, James Martens, George E. Dahl, and Geoffrey E. Hinton. On the importance of initialization and momentum in deep learning. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*, volume 28 of *JMLR Workshop and Conference Proceedings*, pp. 1139–1147. JMLR.org, 2013.

William Swartworth, Deanna Needell, Rachel A. Ward, Mark Kong, and Halyun Jeong. Nearly optimal bounds for cyclic forgetting. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, Atharva Naik, Arjun Ashok, Arut Selvan Dhanasekaran, Anjana Arunkumar, David Stap, Eshaan Pathak, Giannis Karamanolakis, Haizhi Gary Lai, Ishan Purohit, Ishani Mondal, Jacob Anderson, Kirby Kuznia, Krima Doshi, Kuntal Kumar Pal, Maitreya Patel, Mehrad Moradshahi, Mihir Parmar, Mirali Purohit, Neeraj Varshney, Phani Rohitha Kaza, Pulkit Verma, Ravsehaj Singh Puri, Rushang Karia, Savan Doshi, Shailaja Keyur Sampat, Siddhartha Mishra, Sujan Reddy A, Sumanta Patro, Tanay Dixit, and Xudong Shen. Super-naturalinstructions: Generalization via declarative instructions on 1600+ NLP tasks. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (eds.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pp. 5085–5109. Association for Computational Linguistics, 2022.

Zirui Wang, Yulia Tsvetkov, Orhan Firat, and Yuan Cao. Gradient vaccine: Investigating and improving multi-task optimization in massively multilingual models. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

Chengfeng Wu, Tao Zou, Yanru Wu, and Jingge Wang. CORE-MTL: Rethinking gradient balancing via causal orthogonal representations. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=vIIk6uur7Z.

Sen Wu, Hongyang R. Zhang, and Christopher Ré. Understanding and improving information transfer in multi-task learning. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report. *CoRR*, abs/2412.15115, 2024. doi: 10.48550/ARXIV.2412.15115. URL https://doi.org/10.48550/arXiv.2412.15115.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Amir R. Zamir, Alexander Sax, William B. Shen, Leonidas J. Guibas, Jitendra Malik, and Silvio Savarese. Taskonomy: Disentangling task transfer learning. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pp. 3712–3722. Computer Vision Foundation / IEEE Computer Society, 2018.

A Experimental Setup

Task worlds. We study two 15-task instruction-tuning worlds, P3 and SuperNI, each organized into five task families with three tasks per family. Table 2 lists the fixed panels used throughout the paper. For P3, we deterministically sample at most 1,000 training and 100 evaluation examples per task and retain only closed-set examples whose target belongs to the provided answer choices. For SuperNI, we train on the valid instances in the official training files and evaluate on the development split, after verifying that the two splits have no overlapping instance identifiers. Inputs are truncated to 512 tokens; targets are truncated to 96 tokens for P3 and 128 tokens for SuperNI. Language-model loss is computed only on target tokens. All response comparisons use the same fixed evaluation panel within a matched context.

Controlled task world. The controlled synthetic task set, Controlled, consists of five fixed task pools spanning different domains: Chinese text, code, mathematics, instruction following, and web text, denoted by chinese, code, math, instruction, and web. Each task is stored as fixed pre-tokenized training and evaluation pools containing input IDs, attention masks, and causal-language-model labels; no resplitting is performed during an experimental run. All sequences have length 512. At each optimizer step, we use 16 microbatches of four sequences, giving an effective batch size of 64. Training examples are sampled deterministically from the corresponding task pool using a key derived from the experimental seed, task identity, task-occurrence index, and microbatch index. Evaluation uses up to the first 64 examples from each fixed evaluation pool and records the per-task causal-language-model loss. Because the controlled pools do not share a unified raw-text generation or annotation protocol, we use loss-based response vectors and do not report native accuracy or generation metrics for this environment.

Models and optimization. The language experiments use Llama-3.1-8B, Qwen2.5-1.5B, and Qwen2.5-7B. We freeze the base model and train LoRA adapters with rank 16, scaling 32, and zero dropout on the attention projections and MLP projections. Unless stated otherwise, optimization uses AdamW with learning rate 10^{-4}, $(\beta_1, \beta_2) = (0.9, 0.95)$, $\epsilon = 10^{-8}$, weight decay 0.1, and an effective batch size of 64. We use bfloat16, a constant learning rate, and no gradient clipping. Each experiment begins from a common adapter checkpoint obtained after 1,000 optimizer steps of shared multi-task warm-up.

Exact snapshot aliases. For each seed, we generate two balanced 30-step histories in which every task appears twice. The histories use different task orders, while a given task occurrence uses the

Table 2: The fixed task panels used in the language experiments. Each row contributes three tasks from one family.

Family	P3	SuperNI
Sentiment / dialogue	Amazon Polarity; IMDb; Rotten Tomatoes	task639 (MultiWOZ); task1590 (Diplomacy); task1729 (PersonaChat)
Paraphrase / extraction	MRPC; QQP; PAWS	task181 (outcome); task748 (cause–event); task1510 (relation)
Topic / question answering	AG News; DBpedia-14; TREC	task002 (Quoref); task073 (CommonsenseQA); task591 (SciQ)
Commonsense / summarization	SocialIQA; CoS-E; CosmosQA	task511 (Reddit TIFU); task1290 (XSum); task1572 (SAMSum)
Science QA / sentiment	SciQ; QASC; QuaRTz	task363 (SST-2); task875 (Emotion); task1687 (Sentiment140)

same batch and random seed in both runs. Their endpoints generally differ in parameters and optimizer state. Writing the two endpoints as $(\theta_A, m_A, v_A, c_A)$ and $(\theta_B, m_B, v_B, c_B)$, we construct the exact alias pair

$$S_A = (\theta_A, m_A, v_A, c_A), \qquad S_B = (\theta_A, m_B, v_A, c_A). \tag{13}$$

Thus, the two states share model parameters, buffers, AdamW second moments, optimizer clock, and parameter-group settings; only the first moment differs. All quantities derived from the current model are consequently identical.

For provenance interventions, the first-moment difference is decomposed exactly by the task that wrote each gradient term. We insert one unmodified source component as

$$S_A^{[k]} = (\theta_A, m_A + \Delta m^{(k)}, v_A, c_A), \tag{14}$$

without rescaling, normalization, or clipping. The relative reconstruction residual of the complete task-wise decomposition is below 3×10^{-16} in the reported runs.

Matched continuations and evaluation. Branches compared within a context receive the same future task, mini-batches, batch order, random stream, continuation length, and evaluation examples. For decision experiments, each candidate first action is followed by the same deterministic balanced tail. Evaluation is performed in inference mode. The primary study uses seven seeds, $\{30, 42, 54, 66, 78, 90, 102\}$, two alias pairs per seed, and 15 candidate future tasks, yielding $7 \times 2 \times 15 = 210$ contexts per task-world–backbone configuration and 1,260 contexts across the six configurations. Long-horizon response experiments continue to 100 steps; decision experiments report horizons through 20 steps.

The matched-resolution provenance analysis selects one source task from each family before observing the results, giving five source anchors in every configuration. Qwen2.5-1.5B is additionally evaluated with all 15 source tasks as a higher-resolution robustness check. Results from this full-source sweep are reported separately and are not mixed into the cross-backbone aggregates.

Numerical controls and uncertainty. We estimate a numerical floor by repeating an identical one-step update eight times from the same warmed state with matched data and randomness, then taking the 95th percentile of the response distance from the reference execution. An exact-alias response is considered reliably separated when its distance exceeds five times this floor. Confidence intervals use 10,000 seed-level bootstrap resamples: contexts are first aggregated within each seed, after which the seven seeds are resampled with replacement. Source-localization nulls use 10,000 within-pair permutations of source labels while holding the realized response vectors fixed. The reported permutation p-values are configuration-wise descriptive statistics and are not adjusted for multiple comparisons. Sign-flip analyses use $\tau = 10^{-6}$ as the primary eligibility threshold and sweep logarithmically from 10^{-8} to 10^{-4}.

Optimizer-state controls. To isolate the role of cross-step state, we repeat the matched intervention with AdamW, momentum SGD with momentum 0.9, and memoryless SGD. These runs share the nominal learning rate, batches, random stream, and continuation protocol; they are a mechanism control rather than an optimizer-performance comparison, and their update norms are not matched. We also interpolate individual source components as $m(\lambda) = m_A + \lambda \Delta m^{(k)}$ for $\lambda \in \{0, 0.1, \ldots, 1\}$ while fixing parameters, second moments, data, and randomness.

NYUv2. The dense-prediction study uses a dilated ResNet-50 shared encoder with separate DeepLab/ASPP heads for 13-class semantic segmentation, depth, and surface normals. We use the standard 795/654 train/validation split at resolution 288×384 and disable data augmentation during the diagnostic runs. The three losses are cross-entropy, masked ℓ_1 depth error, and one minus masked cosine similarity, respectively. Adam uses learning rate 10^{-4}, $(\beta_1, \beta_2) = (0.9, 0.999)$, $\epsilon = 10^{-8}$, weight decay 10^{-5}, batch size 8, and FP32.

As in the language experiments, two 30-step balanced histories generate different first moments, after which only the first moment from one history is transplanted into the common parameter-and-second-moment carrier. We evaluate seven intervention seeds and five fixed history pairs. Combining the resulting 35 conditions with three future tasks and three response tasks gives 315 ordered interactions per setting. Effects are normalized by the corresponding task loss at the base checkpoint.

Additional checks use a shared 300-step burn-in, a larger nested evaluation panel, and independently trained base checkpoints.

B Proofs and Derivations

B.1 Proof of equation 8

Proof. We give the one-step derivation; the finite-horizon case follows by applying the same linearization to the matched-continuation map.

Consider two states $S_A = (\theta, m_A, v, c)$ and $S_A^{[k]} = (\theta, m_A + \Delta m^{(k)}, v, c)$ that share the same current parameters, second moment, and optimizer clock. Under the matched continuation on future task i, the gradient $g_i = \nabla_\theta L_i(\theta)$ is the same in both branches because the current parameter vector and mini-batch are identical. The AdamW first-moment update therefore differs only through the inserted component:

$$m_i^+(S_A^{[k]}) - m_i^+(S_A) = \beta_1 \Delta m^{(k)}.$$

Holding the shared second-moment branch and optimizer clock fixed, the corresponding first-order change in the parameter update can be written as

$$\delta\theta_{i,k} = -\eta\kappa_c D_i \Delta m^{(k)} + O(\|\Delta m^{(k)}\|^2),$$

where D_i is the diagonal operator induced by the AdamW normalization under the future task-i continuation and κ_c collects the first-moment and bias-correction factors fixed by the optimizer clock.

Let $\theta_i^+(S_A)$ denote the parameter after the matched continuation from S_A. The response of task j is the change in its loss after this continuation. Applying a first-order Taylor expansion of L_j around $\theta_i^+(S_A)$ gives

$$L_j(\theta_i^+(S_A) + \delta\theta_{i,k}) - L_j(\theta_i^+(S_A)) = \nabla_\theta L_j(\theta_i^+(S_A))^\top \delta\theta_{i,k} + O(\|\delta\theta_{i,k}\|^2).$$

Substituting the expression for $\delta\theta_{i,k}$ yields

$$E_{j,i,k} = -\eta\kappa_c \nabla_\theta L_j(\theta_i^+(S_A))^\top D_i \Delta m^{(k)} + O(\|\delta\theta_{i,k}\|^2),$$

which proves the stated one-step form.

For a continuation of length $h > 1$, fix the future task sequence, mini-batches, batch order, random stream, optimizer clock, and evaluation examples. The matched continuation then defines a deterministic map from the inserted first-moment component to the final parameter displacement. Linearizing this map around $\Delta m^{(k)} = 0$ replaces the one-step operator D_i by the Jacobian of the matched-continuation map, giving the finite-horizon analogue. □

B.2 Proof of Theorem 4.1

Proof. Write

$$C_X(a) := C_{H,j}(S_X, a), \qquad b_X := \min_a C_{H,j}(S_X, a), \qquad X \in \{A, B\}.$$

By assumption, the two states have the same current-model representation:

$$\mathcal{G}(\theta_A) = \mathcal{G}(\theta_B).$$

Therefore any policy whose action depends only on the current-model representation and the fixed response-task objective j must induce the same distribution over candidate actions at S_A and S_B. Denote this common distribution by

$$p(a) := \pi_j(a \mid \mathcal{G}(\theta_A)) = \pi_j(a \mid \mathcal{G}(\theta_B)).$$

The average regret of this snapshot-blind policy is then

$$\frac{1}{2} \sum_{X \in \{A,B\}} \mathbb{E}_{i \sim \pi_j(\mathcal{G}(\theta_X))} \left[C_{H,j}(S_X, i) - \min_a C_{H,j}(S_X, a) \right]$$

$$= \sum_a p(a) \left[\frac{1}{2} \sum_{X \in \{A,B\}} C_X(a) \right] - \frac{1}{2} \sum_{X \in \{A,B\}} b_X.$$

Table 3: **Exact snapshot aliases produce different future responses.** Each row aggregates states with identical current parameters and identical protocol coordinates except for optimizer-carried history. *Response RMS* is the root-mean-square difference between the two future response operators. *Separated* counts contexts whose response difference exceeds five times the matched numerical floor. Results are aggregated over seven random seeds.

Task world	Backbone	Exact contexts	Response RMS	Separated
SUPERNI	Llama-3.1-8B	210	1.220×10^{-2}	210/210
SUPERNI	Qwen2.5-1.5B	210	1.035×10^{-2}	210/210
SUPERNI	Qwen2.5-7B	210	1.235×10^{-2}	210/210
P3	Llama-3.1-8B	210	5.811×10^{-3}	210/210
P3	Qwen2.5-1.5B	210	5.534×10^{-3}	210/210
P3	Qwen2.5-7B	210	4.608×10^{-3}	210/210
Total	–	**1260**	–	**1260/1260**

Since the first term is an average, under p, of the common-action costs

$$\frac{1}{2} \sum_{X \in \{A,B\}} C_X(a),$$

it is lower bounded by their minimum:

$$\sum_a p(a) \left[\frac{1}{2} \sum_{X \in \{A,B\}} C_X(a) \right] - \frac{1}{2} \sum_{X \in \{A,B\}} b_X \geq \min_a \left[\frac{1}{2} \sum_{X \in \{A,B\}} C_X(a) \right] - \frac{1}{2} \sum_{X \in \{A,B\}} \min_{a'} C_X(a').$$

By the definition of the aliasing decision gap,

$$\Delta_{H,j}(A,B) := \min_a \frac{1}{2} \sum_{X \in \{A,B\}} C_{H,j}(S_X, a) - \frac{1}{2} \sum_{X \in \{A,B\}} \min_{a'} C_{H,j}(S_X, a'),$$

the desired inequality follows:

$$\frac{1}{2} \sum_{X \in \{A,B\}} \mathbb{E}_{i \sim \pi_j(\mathcal{G}(\theta_X))} \left[C_{H,j}(S_X, i) - \min_a C_{H,j}(S_X, a) \right] \geq \Delta_{H,j}(A,B).$$

Equality is attained by any deterministic snapshot-blind policy that places all mass on an action

$$a^\star \in \arg\min_a \frac{1}{2} \sum_{X \in \{A,B\}} C_{H,j}(S_X, a),$$

because this action minimizes the common-action average cost appearing above. □

C Additional Experimental Results

This appendix reports full carrier intervention analysis, the complete direct history-dependence test, and a matched-resolution provenance comparison.

C.1 Full Carrier Intervention Analysis

Table 5 extends the carrier summary in Table 4 with effect-magnitude ratios, directional-change diagnostics, and interpolation quality across all six settings. This analysis characterizes the AdamW first moment as a control handle on hidden training history, quantifying its effect magnitude relative to the second moment, the parameter displacement it induces, and the smoothness and monotonicity of interpolation along it.

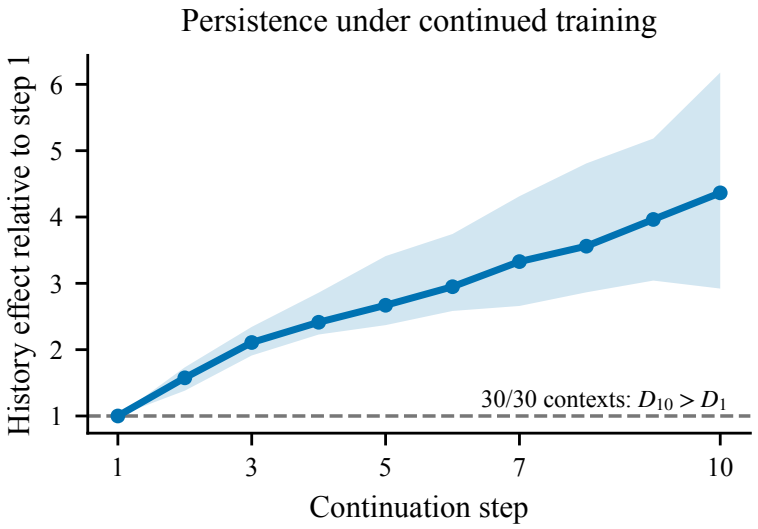


Figure 7: **The hidden-history effect persists under continued training.** For each context, the response-profile distance is normalized by its value after the first continuation step. The line and shaded region show the median and interquartile range over 30 contexts. In all 30 contexts, the distance after ten steps exceeds that after one step. The result establishes persistence over the tested horizon, but does not require every trajectory to increase monotonically at every intermediate step.

Table 4: **The AdamW first moment provides a controllable historical coordinate.** Values are ranges across three backbones. The first-moment effect is substantially larger than the second-moment effect, and all tested interpolation paths are monotone and nearly linear.

Task world	$\|E_m\|/\|E_v\|$	Monotone	Interp. R^2
SUPERNI	6.492–8.110	1.000	0.994–0.999
P3	4.951–9.273	1.000	0.986–0.991

Table 5: Full carrier intervention analysis. m/v compares the response-effect magnitude of intervening on the AdamW first moment versus the second moment. $\cos \Delta m(\theta)$ and $\mathrm{amp}\, \Delta m(\theta)$ measure the directional and magnitude change that the first-moment intervention induces in the parameters. Interp. R^2 and ρ assess smoothness and monotonicity of first-moment interpolation. Across all settings, the first moment is substantially stronger than the second moment and provides a smooth, monotone, and controllable historical coordinate.

			Directional change		Interpolation	
Backbone	World	m/v	$\cos \Delta m(\theta)$	$\mathrm{amp}\, \Delta m(\theta)$	R^2	ρ
Llama-3.1-8B	P3	4.951	0.825	1.340	0.991	1.000
Llama-3.1-8B	SuperNI	7.542	0.884	1.270	0.999	1.000
Qwen2.5-1.5B	P3	7.609	0.748	1.367	0.986	1.000
Qwen2.5-1.5B	SuperNI	8.110	0.944	1.299	0.994	1.000
Qwen2.5-7B	P3	9.273	0.619	1.417	0.990	1.000
Qwen2.5-7B	SuperNI	6.492	0.888	1.424	0.995	1.000

C.2 PROVENANCE RESULTS AT MATCHED RESOLUTION

Table 7 compares all backbones using five anchor provenance tasks, removing the resolution difference between the Qwen2.5-1.5B full-provenance results (1050 contexts) and the anchors5 results (350 contexts) reported for the other backbones in Tables 6 and 8.

Table 6: **Task provenance is exactly recoverable and causally active.** *Closure error* is the residual after reconstructing the complete first-moment difference from its task-wise provenance components. *Selective effect* reports the magnitude induced by inserting one provenance component, normalized by the effect of inserting the complete historical difference. Every intervention preserves the current parameters and therefore leaves the instantaneous task snapshot unchanged. Qwen2.5-1.5B uses the full provenance set; the other configurations use five anchor provenance tasks.

Task world	Backbone	Mode	Contexts	Closure error	Selective effect
SUPERNI	Llama-3.1-8B	anchors5	350	2.65×10^{-16}	0.381
SUPERNI	Qwen2.5-1.5B	full	1050	2.68×10^{-16}	0.414
SUPERNI	Qwen2.5-7B	anchors5	350	2.76×10^{-16}	0.452
P3	Llama-3.1-8B	anchors5	350	2.53×10^{-16}	0.453
P3	Qwen2.5-1.5B	full	1050	2.35×10^{-16}	0.627
P3	Qwen2.5-7B	anchors5	350	2.39×10^{-16}	0.578

Table 7: Provenance results at matched resolution. All rows use five anchor provenance tasks, enabling direct cross-backbone comparison. Qwen2.5-1.5B rows are the anchors5 subset (350 contexts); other backbones use anchors5 throughout. Columns follow Table 6 and Table 8: selective effect, same-task and same-family enrichment, cross-future cosine, amplitude ratio, and rank-one share. The Qwen2.5-1.5B anchors5 subset yields even lower cross-future cosine on SuperNI (0.020 vs. 0.044 in the full set), strengthening the fixed-meaning rejection at matched resolution.

Backbone	World	Sel. effect	Same-task	Same-family	Cross-fut. cos ↓	Amp. max/min	Rank-1 ↓
Llama-3.1-8B	P3	0.453	0.074	0.268	0.071	1.809	0.532
Llama-3.1-8B	SuperNI	0.381	0.069	0.218	0.165	1.866	0.558
Qwen2.5-1.5B	P3	0.621	0.060	0.205	0.045	1.765	0.460
Qwen2.5-1.5B	SuperNI	0.424	0.074	0.207	0.020	1.918	0.536
Qwen2.5-7B	P3	0.578	0.080	0.228	0.057	1.679	0.489
Qwen2.5-7B	SuperNI	0.452	0.051	0.180	0.136	2.045	0.635

Table 8: **Different future tasks read the same provenance differently.** For each fixed past-task provenance k, we vary the future task i and compare the resulting response vectors $e_{i,k}$. Values are ranges across the three tested backbones. Cross-future cosine measures directional agreement between response vectors; amplitude max/min measures variation in effect magnitude; and rank-one share measures the fraction of variation explained by a single fixed response profile. The low directional agreement and partial rank-one fit reject the hypothesis that future tasks merely rescale a fixed provenance meaning.

Task world	Cross-future cosine ↓	Amplitude max/min ↑	Rank-one share ↓
SUPERNI	0.044–0.165	1.866–2.045	0.534–0.635
P3	0.026–0.071	1.679–1.809	0.483–0.532
Overall	0.026–0.165	1.679–2.045	0.483–0.635

C.3 DIRECT HISTORY-DEPENDENCE TEST: FULL RESULTS

Table 9 reports the complete set of metrics for the direct test that fixes the future task i and response task j while varying only the past-task provenance k (Figure 4 in the main text shows the cosine and sign-flip columns).

C.4 CROSS-DOMAIN VALIDATION ON NYUV2

Setup. We apply the exact-alias intervention of Section 2.2 to NYUv2 with a shared-backbone architecture and three heads—semantic segmentation, depth estimation, and surface-normal prediction. From a common base checkpoint we construct snapshot-equivalent optimizer states that differ only in task-attributed history, then continue training under 5 frozen candidate schedules and read out the response of each task. Interactions are fixed-normalized per task and compared against a

Table 9: Direct history-dependence test: fix the future task i and response task j, vary only the past-task provenance k. Icos measures cosine similarity of provenance inputs across k; Ecos measures cosine similarity of the realized response effects. NormDiff and q90Diff quantify the normalized and 90th-percentile effect magnitude differences. $R^2_{\text{hist}}(E)$ and $R^2_{\text{hist}}(I)$ report how much variance in effects versus inputs is explained by past-task identity. EFlip is the fraction of comparisons whose history-induced correction changes sign; IFlip is the stricter fraction where the same (i, j) pair is helpful under one history and harmful under another. Inputs remain aligned ($0.765 \leq \text{Icos} \leq 0.946$) while effects diverge ($0.380 \leq \text{Ecos} \leq 0.514$), and history explains 44–64% of effect variance versus only 8–24% of input variance.

Backbone	World	K	Icos	Ecos	NormDiff	q90Diff	$R^2_{\text{hist}}(E)$	$R^2_{\text{hist}}(I)$	EFlip	IFlip
Llama-3.1-8B	SuperNI a5	5	0.945	0.380	0.613	4.082	0.643	0.079	0.356	0.185
Qwen2.5-1.5B	SuperNI a5	5	0.906	0.462	0.749	4.353	0.535	0.157	0.333	0.214
Qwen2.5-1.5B	SuperNI full	15	0.946	0.505	0.716	4.325	0.632	0.122	0.333	0.227
Qwen2.5-7B	SuperNI a5	5	0.922	0.467	0.707	4.792	0.531	0.133	0.341	0.249
Llama-3.1-8B	P3 a5	5	0.899	0.418	0.619	3.062	0.539	0.121	0.336	0.211
Qwen2.5-1.5B	P3 a5	5	0.765	0.476	0.723	4.162	0.440	0.218	0.352	0.228
Qwen2.5-1.5B	P3 full	15	0.775	0.514	0.709	4.364	0.491	0.237	0.342	0.239
Qwen2.5-7B	P3 a5	5	0.799	0.431	0.696	4.284	0.494	0.190	0.350	0.259

Table 10: NYUv2 IFlip sensitivity under the primary protocol.

Sensitivity axis	Setting	IFlip
Threshold	$\tau = 10^{-6}$	157/315 (49.8%)
Threshold	$\tau = 10^{-5}$	156/312 (50.0%)
Threshold	$\tau = 10^{-4}$	153/304 (50.3%)
Threshold	$\tau = 10^{-3}$	124/243 (51.0%)
Drop-largest condition	seed 102, balanced_random_3	48.8%

Table 11: **Snapshot-equivalent states can require different next-task decisions.** TaskRev is the fraction of response-task contexts whose optimal next task changes between matched historical states; PairRev and MacroRev report pair- and response-task-aggregated reversal rates. Task nADG measures the normalized decision gap induced by choosing the action preferred by the other historical state. NearTie reports reversals attributable to near-tied alternatives. Results aggregate seven seeds and 14 matched pairs per configuration.

Task world	Backbone	MacroRev	TaskRev	PairRev	Task nADG	NearTie
P3	Llama-3.1-8B	0.357	0.495	0.500	0.1066	0.000
SUPERNI	Llama-3.1-8B	0.214	0.505	0.533	0.0733	0.000
P3	Qwen2.5-1.5B	0.429	0.714	0.733	0.1203	0.005
SUPERNI	Qwen2.5-1.5B	0.429	0.605	0.633	0.1096	0.000
P3	Qwen2.5-7B	0.429	0.629	0.600	0.0999	0.000
SUPERNI	Qwen2.5-7B	0.357	0.633	0.667	0.1118	0.000

matched numerical floor obtained by replaying the update with the intervention removed. We run 7 intervention seeds, giving 35 conditions and 315 ordered training–response interactions per setting.

Effect sizes are far above the floor for every task. Under the primary protocol the per-task separations are 8.3×10^4 (segmentation), 6.9×10^4 (depth), and 2.2×10^4 (normal); the pooled median is $3536\times$ with a minimum of $58.4\times$, and every one of the 315 interactions exceeds $10\times$ the floor. The median fixed-normalized effect magnitude is 0.201.

The sign reversal is robust. Table 10 reports IFlip under threshold variation and after removing the single largest-effect condition. The reversal rate stays near one half throughout, and the shared-burn-in, eval-128, and independent base-checkpoint variants (Table 1) confirm that the effect is not tied to a particular base-training run or evaluation panel.

Table 12: **Snapshot-conditioned baselines do not eliminate the decision gap.** Original reports nADG over the single-task actions. Each expanded column adds the corresponding weighted update as one candidate action and reports mean nADG with a seed-bootstrap 95% confidence interval in brackets. Bold indicates a lower mean than the original action set; all expanded intervals exclude zero.

Task world	Backbone	h	Original	DB-MTL expanded	PiKE expanded
P3	Llama-3.1-8B	1	0.0574	$0.0589\,[0.0491, 0.0682]$	$0.0589\,[0.0492, 0.0693]$
P3	Qwen2.5-1.5B	1	0.0929	$\mathbf{0.0889}\,[0.0740, 0.1053]$	$0.0974\,[0.0829, 0.1155]$
P3	Qwen2.5-7B	1	0.0735	$0.0747\,[0.0636, 0.0896]$	$0.0781\,[0.0703, 0.0884]$
SUPERNI	Llama-3.1-8B	1	0.0451	$\mathbf{0.0414}\,[0.0265, 0.0573]$	$0.0486\,[0.0327, 0.0635]$
SUPERNI	Qwen2.5-1.5B	1	0.0748	$0.0759\,[0.0590, 0.0929]$	$0.0781\,[0.0630, 0.0931]$
SUPERNI	Qwen2.5-7B	1	0.0786	$\mathbf{0.0682}\,[0.0587, 0.0777]$	$\mathbf{0.0772}\,[0.0692, 0.0842]$
P3	Llama-3.1-8B	20	0.0368	$0.0508\,[0.0347, 0.0716]$	$0.0370\,[0.0314, 0.0418]$
P3	Qwen2.5-1.5B	20	0.0486	$0.0545\,[0.0410, 0.0717]$	$0.0499\,[0.0393, 0.0604]$
P3	Qwen2.5-7B	20	0.0372	$0.0403\,[0.0293, 0.0530]$	$0.0387\,[0.0312, 0.0452]$
SUPERNI	Llama-3.1-8B	20	0.0240	$\mathbf{0.0237}\,[0.0194, 0.0284]$	$0.0242\,[0.0210, 0.0275]$
SUPERNI	Qwen2.5-1.5B	20	0.0343	$\mathbf{0.0340}\,[0.0232, 0.0449]$	$0.0348\,[0.0230, 0.0459]$
SUPERNI	Qwen2.5-7B	20	0.0370	$0.0391\,[0.0294, 0.0493]$	$0.0374\,[0.0290, 0.0480]$

C.5 SNAPSHOT-CONDITIONED POLICY BASELINES

DB-MTL and PiKE produce weighted multi-task updates rather than a single-task choice. For each exact-alias pair, we compute their weights from the common snapshot and add the resulting update as one candidate shared action. Because the model parameters are identical, each method prescribes the same action for both states. We evaluate $h \in \{1, 20\}$ on the six task configurations, using seven seeds and 210 contexts per configuration.

Table 12 compares the original nADG with the expanded-action nADG. Every expanded 95% confidence interval remains above zero. DB-MTL lowers the mean in 5/12 settings and PiKE in 1/12, but neither eliminates the aliasing gap.