THE ANSWER IS NOT THE STATE: TRAJECTORY-LEVEL CORRIGIBILITY IN DIFFUSION LANGUAGE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion language models commit tokens across a shared response canvas, making unfinished outputs appear globally revisable. We show that the answer is not the full decision state. Remasking it can leave visible rationale that preserves trajectory history and reconstructs a prior decision despite current evidence. Matched answer and path resets isolate the retained rationale state. Transplanting a coherent rationale between paired trajectories changes the rebuilt decision, whereas shuffling the same tokens weakens the effect. On source-disjoint multi-hop questions, path reset makes decisions in Dream, LLaDA, and iLLaDA responsive to evidence identity, including controlled corrupt evidence. Answer reset retains path dependence, and comparisons with absent evidence rule out a generic regeneration benefit. In the two-model natural development analysis, direct leave-one-out attribution improves repair prioritization, but even the best partial reset remains far below path reset. Final-answer regeneration alone does not reveal whether a masked-dLLM trajectory remains revisable.

1 INTRODUCTION

Diffusion language models turn generation order into a policy. Rather than emit tokens in a fixed left-to-right sequence, they repeatedly predict across a shared masked canvas. The model proposes token values, and a commitment policy decides which positions become visible state (Austin et al., 2023; Sahoo et al., 2024; Nie et al., 2025; Ye et al., 2025). This flexibility enables arbitrary-order and blockwise generation, semantic planning, and early answer commitment (Arriola et al., 2025; Israel et al., 2026; Li et al., 2026a), but it also determines when generated text begins to condition later predictions. If a retrieved passage, tool result, or verification signal arrives after part of the canvas has committed, can the model still revise its decision? An unfinished canvas is not necessarily a revisable decision state.

Under monotone decoding, committed positions remain visible to every later denoising step. We call a trajectory *corrigible* when the evidence currently present can still control its decision after such commitments have formed. Across Dream and LLaDA, however, keeping the answer masked does not guarantee that later evidence can redirect the decision: other visible commitments can already constrain what the answer becomes. Answer openness is necessary for monotone revision, but it does not describe the full decision state. Prior work shows that intermediate tokens can steer denoising trajectories (Li et al., 2026b; Yamabe & Sakuma, 2026; Jeung et al., 2026) (see Appendix A); we ask which retained visible content prevents newly arrived evidence from redirecting the decision.

The answer is not the state. Committed rationale tokens remain on the canvas and condition later predictions; here, *state* refers to these visible tokens, not hidden activations. *Answer reset* remasks the decision while retaining the rationale. *Path reset* remasks both regions while leaving the prompt, current evidence, and reconstruction policy unchanged. Their contrast isolates what the retained rationale contributes to the rebuilt decision. We next remask the answer and transplant a coherent rationale while holding the recipient prompt and evidence fixed. A shuffled version of the same rationale tokens tests whether their sequence organization matters. Figure 1 summarizes these interventions.

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

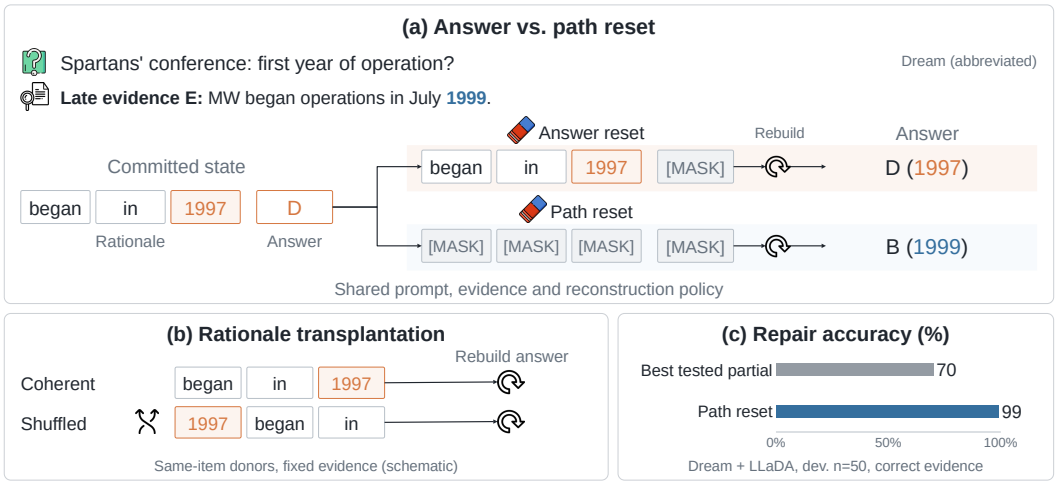

Figure 1: **Visible rationale is causal trajectory state.** (a) In a Dream example, answer reset reconstructs the prior decision despite late correct evidence, whereas path reset changes it. (b) A separate same-item transplantation compares coherent donor rationale with a shuffle of the same tokens under fixed recipient evidence. (c) The best tested partial reset remains below the full-path endpoint. Case text is abbreviated and token strips are schematic.

Across Dream, LLaDA, and iLLaDA on source-disjoint HotpotQA questions, the gain from path reset over answer reset is 49.6 points larger with correct evidence than with matched absent evidence. Together, the correct, corrupt, and absent conditions show that path reset makes the rebuilt decision follow the evidence currently present, including a controlled corruption, rather than benefiting from additional regeneration alone. Rationale transplantation tests the content of that retained state under a separate intervention. With the recipient prompt and evidence fixed, a coherent opposite rationale moves the rebuilt decision 28.3 points further toward the donor-associated answer than a shuffled version of the same rationale tokens.

Identifying rationale as decision state does not tell us whether that state is compact. On the two-model, 50-item natural development set, attribution-guided resets outperform five count-matched random baselines. That advantage does not close the endpoint gap: the best tested partial reset reaches 70% accuracy, compared with 99% after path reset. This attribution–repair gap shows that local influence can guide token choice even when the unreopened rationale still supports reconstruction of the prior trajectory.

Contributions. We make three contributions:

1. We formulate *trajectory-level corrigibility* as whether current evidence can control a decision after the canvas has partially committed, and show that an open answer does not by itself make the trajectory revisable.
2. We identify committed rationale as persistent, transferable visible state: it retains arrival history under answer reset, whereas path reset transfers control to current evidence, including controlled corrupt evidence, across three checkpoints. Coherent same-item transplants further shift rebuilt decisions beyond shuffled-token controls.
3. We establish an attribution–repair gap in a two-model natural development analysis: repair-aligned scores improve token prioritization, but no tested partial reset approaches the full-path causal endpoint.

2 Method

We identify which visible commitments determine whether late evidence can revise a decision. Our comparisons begin from paired trajectories with the same pre-arrival state. We use resets to locate the retained state, transplantation to test its content, and directional attribution to probe local repair.

2.1 Token Commitment Trajectories

Fix a prompt and context c. A dLLM generates a length-L response canvas

$$x^{(t)} \in (\mathcal{V} \cup \{[\texttt{MASK}]\})^L, \qquad t = 0, \ldots, T, \tag{1}$$

where $\mathcal{V}$ is the token vocabulary and $x^{(0)}$ is initially masked. We partition the generated positions into three disjoint regions: an externally controlled evidence region E, a free-rationale region R, and a decision region D. The decision region may contain multiple positions in the formalism; our experiments use a constrained one-token decision.

At step t, the model produces a position-wise token distribution $p_{\theta,j}(\cdot \mid c, x^{(t)})$. A commitment policy π selects a subset U_t of currently masked endogenous positions and writes model-proposed values into them. Changing π can therefore change the trajectory even when p_θ, c, and $x^{(0)}$ are fixed.

We define the endogenous commitment set

$$C_t = \left\{ j \in R \cup D : x_j^{(t)} \neq [\texttt{MASK}] \right\}. \tag{2}$$

Evidence positions are excluded because their values are supplied by the intervention rather than committed by the model. The first commitment time of endogenous position j is $\kappa_j = \min\{t : j \in C_t\}$. Under the base monotone sampler, $C_t \subseteq C_{t+1}$: once written, a token remains part of the canvas. The reset interventions below deliberately break this monotonicity by returning selected positions to the masked state.

2.2 Evidence Arrival and Paired Trajectories

We represent evidence identity by $e \in \{e^+, e^-, e^0\}$. Here e^+ supports the correct decision, e^- is a controlled corruption supporting a designated wrong proposal, and e^0 is a matched absent-evidence message. The absent condition is an explicit input rather than the absence of an intervention, which preserves the same evidence slot and arrival event across conditions.

For an arrival time τ, the evidence region is protected from endogenous commitment and remains masked before arrival. At step τ, we write the tokenized evidence before the model forward pass and keep it fixed thereafter:

$$x_E^{(t)} = \begin{cases} [\texttt{MASK}], & t < \tau, \\ \mathrm{tok}(e), & t \geq \tau. \end{cases} \tag{3}$$

Thus evidence identity can affect neither token values nor commitment choices before τ.

Our comparisons use paired trajectories. An experimental unit i fixes the model, item, decoding seed, commitment policy, and initial canvas. Trajectories that vary only in evidence identity are therefore identical for all $t < \tau$. We use γ for a later state-intervention time, with $0 < \tau < \gamma < T$. Evidence identity and arrival time are thus controlled changes to a shared pre-arrival trajectory.

2.3 State Interventions and Corrigibility

We intervene on the state present at step γ through a reset scope $s \in \mathcal{S} = \{\varnothing, D, R \cup D\}$. For any scope s, define

$$[\mathcal{R}_s(x)]_j = \begin{cases} [\texttt{MASK}], & j \in s, \\ x_j, & j \notin s. \end{cases} \tag{4}$$

The three scopes correspond to no reset, answer reset, and path reset. The intervention never alters E, the fixed prompt, or the decision labels. After either nonempty reset, both arms use the same reasoning-first reconstruction policy and regenerate the answer. Answer reset keeps the committed rationale, whereas path reset removes it, so their contrast isolates the effect of retaining the path. The no-reset trajectory retains the standard commitment policy and serves as an operational decoder baseline.

Let $O_i(e; \tau, \gamma, s)$ denote the terminal rationale and decision canvas, and $Y_i(e; \tau, \gamma, s)$ its extracted decision. For correct decision $y_i^\star$ and designated wrong proposal $\widetilde{y}_i$, define

$$Z_i(e; \tau, \gamma, s) = \mathbf{1}[Y_i(e; \tau, \gamma, s) = y_i^\star],$$
$$F_i(e; \tau, \gamma, s) = \mathbf{1}[Y_i(e; \tau, \gamma, s) = \widetilde{y}_i].$$

These denote accuracy and proposal following. Corrigibility is the extent to which current evidence controls these outcomes after endogenous commitments have formed.

Path reset regenerates more of the canvas than answer reset, so an accuracy gain need not come from evidence. We isolate evidence-conditioned revision by subtracting the same path-versus-answer contrast under absent evidence. With τ and γ fixed, we suppress them in the notation and define

$$\Delta_{\mathrm{corr}} = \mathbb{E}_i\left[\big(Z_i(e^+, R \cup D) - Z_i(e^+, D)\big) - \big(Z_i(e^0, R \cup D) - Z_i(e^0, D)\big)\right]. \tag{5}$$

A positive Δ_{corr} means that removing R improves correct-evidence revision beyond any generic effect of regenerating more tokens.

Correctness alone cannot distinguish evidence responsiveness from a reset that favors a fixed label. We therefore swap correct for controlled corrupt evidence and measure two complementary interactions:

$$\Delta_{\mathrm{id}}^{\mathrm{acc}} = \mathbb{E}_i\left[\big(Z_i(e^+, R \cup D) - Z_i(e^-, R \cup D)\big) - \big(Z_i(e^+, D) - Z_i(e^-, D)\big)\right], \tag{6}$$

$$\Delta_{\mathrm{id}}^{\mathrm{follow}} = \mathbb{E}_i\left[\big(F_i(e^-, R \cup D) - F_i(e^+, R \cup D)\big) - \big(F_i(e^-, D) - F_i(e^+, D)\big)\right]. \tag{7}$$

A positive interaction means that path reset makes the final decision more responsive to which evidence is currently present, rather than imposing a fixed direction of change.

Arrival-time pairs test whether reset removes trajectory history. For the same evidence, we compare decisions generated with $\tau_{\mathrm{early}} < \tau_{\mathrm{late}} < \gamma$. Because path reset masks every endogenous position while preserving the evidence, its post-intervention states satisfy

$$\mathcal{R}_{R \cup D}\left(x_i^{(\gamma;\tau_{\mathrm{early}},e)}\right) = \mathcal{R}_{R \cup D}\left(x_i^{(\gamma;\tau_{\mathrm{late}},e)}\right). \tag{8}$$

Under a shared deterministic decoder, or shared post-reset randomness, this equality implies identical terminal canvases O_i. Answer reset does not imply this equivalence because it retains the rationale generated under each arrival history.

2.4 Rationale Transplantation

We test whether rationale content carries the prior decision state by transplanting an exact intermediate rationale between paired trajectories. Within each item, the recipient model, prompt, evidence, intervention time, and post-intervention decoding remain fixed; only the donor rationale changes. The decision is remasked before reconstruction. Donor trajectories receive early correct, corrupt, or absent evidence and are captured immediately before step γ. Formally, for recipient state x_i^r and donor state x_i^d,

$$\big[\mathcal{T}(x_i^r, x_i^d)\big]_j = \begin{cases} x_{ij}^d, & j \in R, \\ [\texttt{MASK}], & j \in D, \\ x_{ij}^r, & j \notin R \cup D. \end{cases} \tag{9}$$

Donor states remain as token IDs in memory; decoded text is never retokenized into the recipient canvas.

For correct and corrupt recipients, we compare five rationale conditions: a self-evidence donor, the opposite-evidence donor, an absent-evidence donor, a shuffled opposite donor, and a fully masked rationale. Shuffling permutes committed donor tokens across the same committed positions, preserving the token multiset, filled-position count, prompt, and evidence.

Let $Z_i^{r,q}$ and $F_i^{r,q}$ denote accuracy and proposal following for recipient evidence $r \in \{+, -\}$ and rationale condition q. For comparator $q \in \{\mathrm{opp}, \mathrm{shuf}\}$, define the donor-consistent contrasts

$$\begin{aligned} D_{\mathrm{acc}}(q) &= \frac{1}{2}\,\mathbb{E}_i\Big[Z_i^{+,\mathrm{self}} - Z_i^{+,q} + Z_i^{-,q} - Z_i^{-,\mathrm{self}}\Big], \\ D_{\mathrm{follow}}(q) &= \frac{1}{2}\,\mathbb{E}_i\Big[F_i^{+,q} - F_i^{+,\mathrm{self}} + F_i^{-,\mathrm{self}} - F_i^{-,q}\Big]. \end{aligned} \tag{10}$$

Writing $\mathbf{D}(q) = (D_{\mathrm{acc}}(q), D_{\mathrm{follow}}(q))$, the identity effect is $\mathrm{T1} = \mathbf{D}(\mathrm{opp})$, and the sequence-organization effect is $\mathrm{T2} = \mathbf{D}(\mathrm{opp}) - \mathbf{D}(\mathrm{shuf})$. The fully masked condition is paired with a separately executed path-reset run under the same transplantation configuration; their agreement is the T4 implementation check.

2.5 Directional Attribution and Repair Sufficiency

To test whether persistent rationale admits a compact repair, we rank its committed positions at step γ from two directions. Let $A_i^\gamma = \{j \in R : x_{ij}^{(\gamma)} \neq [\texttt{MASK}]\}$ be the candidate positions. The decision-masking score measures how removing the decision changes the likelihood of an observed rationale token,

$$a_{ij}^{D\to R} = \left|\log p_{\theta,j}\left(x_{ij}^{(\gamma)} \mid c, x_i^{(\gamma)}\right) - \log p_{\theta,j}\left(x_{ij}^{(\gamma)} \mid c, \mathcal{R}_D(x_i^{(\gamma)})\right)\right|. \tag{11}$$

The rationale token remains visible in both evaluations, so $a_{ij}^{D\to R}$ measures observed-token coupling. It is not a masked-token conditional or a terminal causal effect.

To evaluate the masked intervention used for repair, let d_i be the committed pre-reset decision and $x_i^{-D} = \mathcal{R}_D(x_i^{(\gamma)})$. The direct rationale-to-decision score measures the effect of removing rationale token j on the current decision distribution,

$$a_{ij}^{R\to D} = \left|\log p_\theta(d_i \mid c, x_i^{-D}) - \log p_\theta\big(d_i \mid c, \mathcal{R}_{\{j\}}(x_i^{-D})\big)\right|. \tag{12}$$

The probability is computed over the same constrained decision labels used during generation. For either score, fixed-budget repair resets the top 25%, 50%, or 75% of committed candidates together with D; adaptive mass-80 resets the shortest ranked prefix containing 80% of measured attribution.

Each direct repair set is compared with five independently seeded random subsets drawn from the same candidates and matched exactly in size for each trajectory. Answer reset, low confidence, recency, the original score, and path reset provide additional reference points. We call a ranking *selective* if it beats count-matched random subsets, and a partial reset *sufficient* if it approaches the path-reset endpoint. The two criteria answer different questions.

3 Experiments

An open answer does not guarantee that a dLLM will use newly arrived evidence. We localize the remaining decision state with matched resets, test how rationale sequence organization affects it through transplantation, and ask whether token-level attribution can repair it compactly.

3.1 Experimental Setup

Models and tasks. We evaluate Dream-v0-Instruct-7B, LLaDA-8B-Instruct, and iLLaDA-8B-Instruct (Ye et al., 2025; Nie et al., 2025; 2026). The controlled timing experiments use depth-8 arithmetic chains, and the natural experiments use HotpotQA bridge questions (Yang et al., 2018). Correct evidence contains annotated supporting facts, corrupt evidence substitutes the answer entity to support a designated wrong option, and absent evidence is a matched explicit message. The final 150-question holdout shares no source IDs with the 50-item development set.

Generation and outcomes. Unless stated otherwise, generation uses 32 denoising steps at temperature zero. Tokens follow natural confidence commitment before the intervention and reasoning-first commitment after a reset; evidence arrives at step 20 and reset occurs at step 24. Primary experiments constrain the decision to one A/B/C/D token, removing answer parsing and within-answer factorization from the comparison. We report decision accuracy, wrong-proposal following, and paired early–late disagreement. A separate free-form stress test reports token F1 and parsing integrity. Path reset serves as a causal endpoint, not a deployment policy (see Appendix B).

Inference. Every primary comparison pairs the same model, item, and seed. Binary paired outcomes use exact McNemar tests. For pooled mean effects, an item-clustered bootstrap keeps every model, seed, evidence condition, and intervention for a sampled item together; the evaluated checkpoints and stochastic seeds are fixed. Appendix C documents dataset construction, frozen decision rules, and integrity checks.

3.2 Answer Openness Does Not Determine Evidence Use

We sweep evidence arrival over steps $2, 4, 8, 12, 16$, and 20, tracking answer openness and correct-evidence gain under standard confidence commitment. In Dream, the open-decision fraction falls

from 78% to 16% and correct-evidence gain from +36 points to zero (Figure 2). In LLaDA, only 22% remain open at step 2 and correct evidence produces no positive curve. These incompatible curves rule out a shared answer-token deadline.

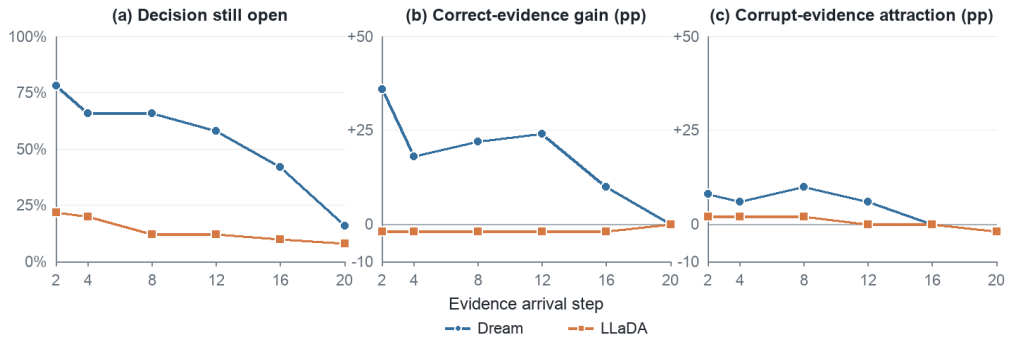

Figure 2: **Dream's correct-evidence benefit contracts as its decisions commit.** LLaDA commits most decisions before the earliest intervention and shows no corresponding correct-evidence curve. The panels report the fraction of decisions open at arrival, correct-over-absent accuracy gain, and corrupt-over-correct attraction to the proposed wrong answer.

All 395 decisions committed before evidence arrival remain invariant across correct, corrupt, and absent evidence. Among the 205 decisions that remain open, correct evidence raises pooled accuracy by 24.4 points and corrupt evidence raises wrong-proposal following by 7.8 points. The correct-evidence gain is +33.7 points in Dream and −11.9 in LLaDA. Revision requires an open answer under monotone sampling, and evidence use also depends on state beyond the decision token.

3.3 Committed Rationale Retains Arrival History

On the controlled arithmetic task, we compare two resets from the same natural-confidence trajectory to locate the state beyond the answer. At step 24, we reopen either the answer alone or the answer together with the free rationale; both arms then use the same reasoning-first reconstruction policy. At late evidence arrival, path reset increases correct-versus-corrupt accuracy control by 31 points and wrong-proposal-following control by 26 points relative to answer reset. Because the decision is reopened in both arms, the additional control comes from removing the rationale. Complete model-level cells appear in Appendix D.2.

We then compare early and late evidence arrivals to test whether the rebuilt decision retains the arrival history. Under answer reset, decisions disagree in 26.7% of controlled pairs. Under path reset, the decision and complete rationale agree in all 300/300 pairs: masking $R \cup D$ exposes identical prompt and evidence to the same decoder.

Table 1: **Early and late evidence arrivals yield identical decisions after path reset.** Late-arrival cells report pooled accuracy/wrong-proposal following (A/F, %) over Dream and LLaDA; the final column reports early–late decision disagreement across evidence identities.

Reset	Correct A/F	Corrupt A/F	Early–late disagree (%)
Answer	21/48	22/46	26.7
Path	47/33	17/57	0.0

Whether the retained rationale helps depends on when evidence arrives. In Dream, path reset lowers accuracy from 64% to 38% when correct evidence arrives early, but raises it from 26% to the same 38% when that evidence arrives late. The shared endpoint shows that reset removes trajectory history rather than improving accuracy by construction.

3.4 Path Reset Transfers Control to Current Evidence

We compare no reset, answer reset, and path reset in Dream, LLaDA, and iLLaDA on 150 held-out HotpotQA multi-hop questions with no source-ID overlap with the development set. Correct evidence uses the original annotated supporting sentences, while corrupt evidence applies a controlled answer-entity substitution. Matched absent evidence provides the reference for separating evidence use from additional regeneration.

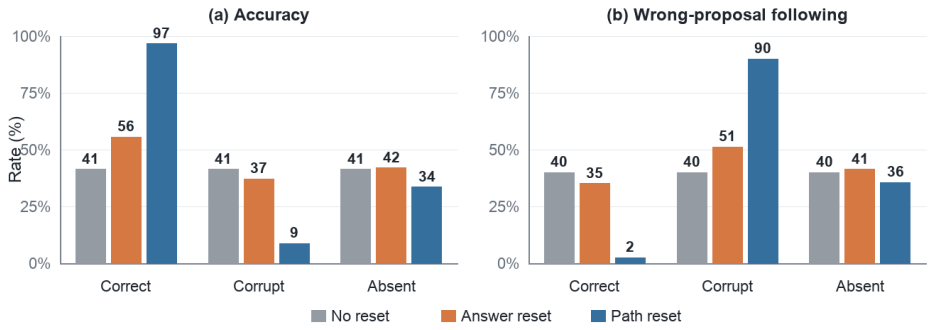

Figure 3: **After path reset, correct and corrupt evidence drive decisions in opposite directions.** Pooled over Dream, LLaDA, and iLLaDA on the source-disjoint holdout, correct-evidence accuracy and following of the controlled corruption both rise, while absent-evidence accuracy does not improve. No-reset decisions remain evidence-invariant because they commit before evidence arrives.

With correct evidence, path reset reaches 96.7% accuracy, compared with 55.6% for answer reset and 41.3% for the standard decoder (Figure 3). Under matched absent evidence, its accuracy instead falls to 33.6%. Path reset's advantage over answer reset is 49.6 points larger with correct evidence than with absent evidence (95% interval $[44.0, 54.9]$; Equation 5). The absent-evidence comparison shows that additional regeneration alone cannot explain the gain.

Replacing correct evidence with a controlled corruption changes path-reset accuracy from 96.7% to 8.7% and wrong-proposal following from 2.4% to 90.0%. Relative to answer reset, the corresponding accuracy and following interactions are $+69.6$ and $+71.6$ points (Equations 6 and 7), with positive effects in Dream, LLaDA, and iLLaDA (Table 2). The same intervention that enables correction also makes the controlled corruption effective.

Table 2: **Correct-specific and evidence-identity effects replicate across all three checkpoints.** Entries are percentage-point effects with 95% bootstrap intervals; the pooled row resamples source items as clusters.

Model	Correct-specific benefit	Accuracy control	Following control
Dream	$+41.3$ $[30.7, 52.0]$	$+60.0$ $[51.3, 68.7]$	$+61.3$ $[52.7, 70.0]$
LLaDA	$+51.3$ $[42.0, 60.7]$	$+76.0$ $[68.0, 83.3]$	$+75.3$ $[67.3, 82.7]$
iLLaDA	$+56.0$ $[47.3, 64.0]$	$+72.7$ $[64.7, 80.0]$	$+78.0$ $[70.7, 84.7]$
Pooled	$\mathbf{+49.6}$ $\mathbf{[44.0, 54.9]}$	$\mathbf{+69.6}$ $\mathbf{[64.4, 74.4]}$	$\mathbf{+71.6}$ $\mathbf{[66.4, 76.4]}$

Answer reset also allows evidence to affect the decision. Relative to no reset, correct-evidence accuracy rises 14.2 points and corrupt-evidence following rises 11.1 points, while absent-evidence accuracy changes by $+0.7$ points. Path reset has the larger effect, with positive path-over-answer effects across all four answer types (Appendix D.3).

In a supplementary holdout analysis, answer reset reproduces the pre-reset label more often than path reset under both correct evidence (71.8% versus 43.3%) and controlled corrupt evidence (75.8% versus 46.0%; Appendix D.3).

On the two-model, 50-item natural development factorial, answer-reset decisions disagree between early and late evidence in 21.7% of pairs. After path reset, both the decision and complete rationale

agree in every pair. On natural questions, removing $R \cup D$ changes which evidence controls the decision and erases the effect of when that evidence arrived.

3.5 Rationale Sequence Organization Controls the Rebuilt Decision

Changing only the rationale donor changes the rebuilt decision. Here, both donor and recipient trajectories receive evidence at step 2. At step 24, we replace the recipient's rationale with exact token IDs from a donor that received correct, corrupt, or absent evidence. The recipient retains its own correct or corrupt evidence. We remask the decision and hold the reconstruction policy and seed fixed. A shuffled-opposite donor preserves committed positions and the token multiset while disrupting sequence organization.

As an implementation check, the fully masked donor matches a separately executed path-reset run in all 900 paired decisions, with evidence arriving at step 2 and reset at step 24 in both runs.

The rebuilt decision moves with donor identity. Replacing a self donor with the opposite donor lowers accuracy from 70.4% to 54.0% for correct recipients and raises it from 29.1% to 48.2% for corrupt recipients. Pooling these paired directions gives a T1 effect of $+17.8$ points for accuracy and $+18.9$ points for wrong-proposal following (Equation 10). Both effects are positive in all three models (Table 3).

Table 3: **Coherent donor rationale controls rebuilt decisions under fixed recipient evidence.** T1 is the donor-consistent self-to-opposite effect; T2 compares the coherent opposite donor with its shuffled-token control. Entries are percentage-point effects with 95% bootstrap intervals; the pooled row resamples source items as clusters.

Model	T1 Acc.	T1 Follow	T2 Acc.	T2 Follow
Dream	$+18.0$ $[11.7, 24.3]$	$+18.7$ $[12.0, 25.3]$	$+26.0$ $[20.7, 31.7]$	$+26.0$ $[20.7, 31.7]$
LLaDA	$+20.7$ $[14.7, 26.7]$	$+20.7$ $[14.7, 26.7]$	$+38.7$ $[33.7, 44.0]$	$+38.0$ $[32.7, 43.3]$
iLLaDA	$+14.7$ $[9.7, 20.0]$	$+17.3$ $[12.3, 22.7]$	$+20.3$ $[15.0, 25.7]$	$+19.0$ $[13.7, 24.3]$
Pooled	$\mathbf{+17.8}$ $\mathbf{[14.4, 21.1]}$	$\mathbf{+18.9}$ $\mathbf{[15.4, 22.4]}$	$\mathbf{+28.3}$ $\mathbf{[25.0, 31.8]}$	$\mathbf{+27.7}$ $\mathbf{[24.3, 31.1]}$

Shuffling the opposite donor shifts decisions back toward the recipient evidence. For correct recipients, accuracy rises from 54.0% to 86.2%, closer to 96.9% with a fully masked rationale. For corrupt recipients, it falls from 48.2% to 23.8%, toward the masked endpoint of 9.1%. The pooled coherent-versus-shuffled effects are $+28.3$ points for donor-consistent accuracy and $+27.7$ points for following (T2; Table 3). Because shuffling preserves the occupied positions and token multiset, this contrast identifies a role for sequence organization.

The absent-evidence donor lies between the correct- and corrupt-donor endpoints. The donor and shuffle contrasts show that organized rationale influences reconstruction within same-item trajectories under fixed recipient evidence (Appendix D.4).

3.6 Direct Leave-One-Out Attribution Improves Prioritization, Not Sufficiency

Can a small subset of rationale tokens provide a sufficient repair? On the two-model, 50-item natural development set, the decision-masking score in Equation 11 places 97.3% of measured coupling in its top 25% of tokens, and its mass-80 policy selects 5.5 tokens on average. Neither fixed-budget selection nor the adaptive mass-80 policy outperforms its count-matched random control.

The direct leave-one-out score in Equation 12 prioritizes useful reset locations. Relative to the mean of five count-matched random subsets, its top-50 and mass-80 sets improve accuracy by $+16.8$ points $[9.0, 25.2]$ and $+18.2$ points $[11.0, 26.0]$, with positive effects in both models. The two scores select different tokens: their rankings have a mean Spearman correlation of 0.115 and a top-quarter overlap of 30.0%.

The gap to path reset remains large. Under correct evidence, accuracy rises from 46% with answer reset to 49%, 66%, and 70% with direct top-25, top-50, and top-75 resets. Direct mass-80 reaches 65% while reopening 30.7 rationale tokens on average. Path reset reaches 99%, and no tested partial set comes within five points of that endpoint (Figure 4; Appendix D.5).

These partial resets also make controlled corrupt evidence more influential. Switching from correct to corrupt evidence raises wrong-proposal following from 30% to 75% for direct top-50 and from 30% to 74% for mass-80.

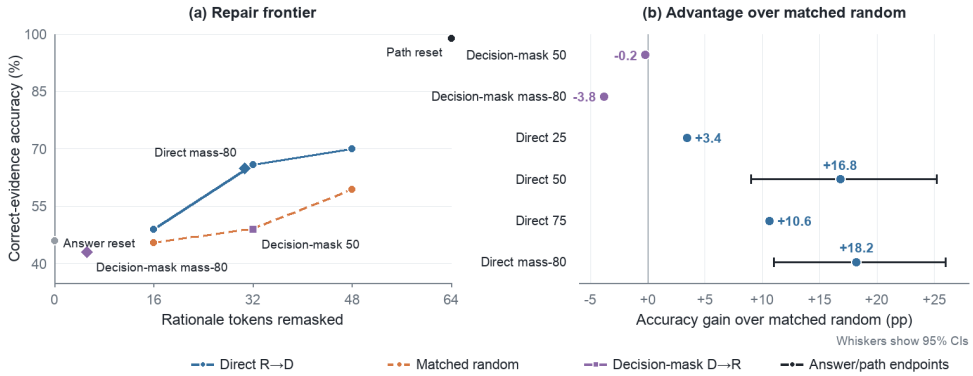

Figure 4: **Direct attribution improves token choice while tested partial resets remain below the full-path endpoint.** The direct ranking outperforms the mean of five count-matched random subsets at top-50 and mass-80; the decision-masking ranking does not. Path reset remains the 99% endpoint.

The direct score and the repair outcome probe different stages of generation. Each leave-one-out probe masks one rationale token while keeping the others visible and measures the change in the pre-reset decision's probability. Repair instead masks a set of tokens and runs the reconstruction process to completion. Individual scores therefore do not determine the joint effect of a reset set. The advantage over count-matched random repair establishes useful prioritization; the remaining gap to path reset shows that the retained rationale still limits revision under these tested policies.

3.7 Robustness Across Decoding and Decision Interfaces

The path-level mechanism persists under temperature-one decoding in Dream and LLaDA (Table 4).

Table 4: **Evidence-control effects persist under temperature-one decoding.** Entries are percentage points with 95% bootstrap intervals over five fixed seeds; the pooled row resamples source items as clusters.

Model	Δ_{corr}	$\Delta_{\text{id}}^{\text{acc}}$	$\Delta_{\text{id}}^{\text{follow}}$
Dream	$+11.6\ [4.4, 18.8]$	$+12.4\ [4.4, 20.4]$	$+14.4\ [7.2, 21.6]$
LLaDA	$+16.0\ [6.8, 25.2]$	$+23.2\ [14.8, 32.0]$	$+26.0\ [18.4, 33.6]$
Pooled	$\mathbf{+13.8\ [8.2, 19.4]}$	$\mathbf{+17.8\ [10.8, 24.6]}$	$\mathbf{+20.2\ [14.4, 26.2]}$

Four of five seeds reproduce the Δ_{corr} direction, and all five reproduce both evidence-identity directions; every pooled interval excludes zero. The effects also persist at reset steps 22/24/26/28 and under a proportional 64-step schedule (Appendix D.6).

The free-form stress test probes decision spans rather than A/B/C/D tokens. Across three models, it yields a correct-specific token-F1 effect of $+26.0$ points $[22.9, 29.2]$. Per-cell parsing ranges from 1.1% to 88.6%, below the 95% threshold, so this experiment supports span-level movement but not exact multi-token replication (Appendix D.7).

4 Conclusion

Diffusion language models can reopen an answer even as retained rationale reconstructs the prior decision. Reset and transplantation identify that rationale as visible trajectory state. In the two-model natural development analysis, direct attribution improves token prioritization without yielding a compact sufficient repair. Alignment evaluations for masked dLLMs should track commitment order and the evidence that remains able to control the trajectory.

Reproducibility Statement

The main paper defines the interventions and estimands needed to reproduce each claim. Appendix C reports the data construction, state layout, schedules, controls, pairing, and statistical tests; Appendix D contains complete result tables and failed gates. The anonymous supplementary artifact releases executable runners and analyses, frozen protocols, deterministic dataset artifacts, path-sanitized raw trajectories, reference summaries, environment metadata, and a checksum-verified entry point that regenerates every reported aggregate, data figure, generated table, and qualitative case report.

Ethics Statement

This work uses public model checkpoints and benchmark data and does not involve human subjects or private user data. Corrupt evidence is a controlled answer-entity substitution on benign question-answering examples, not naturally occurring misinformation. The results expose a dual-use property: reopening trajectory state can transfer control to correct evidence or to an untrusted corruption. Any deployed revision mechanism therefore requires an independent evidence-trust and provenance policy. Released data derivatives remain subject to the terms of their source datasets and models. Icons in Figure 1 were designed by Freepik (Magnific), from Flaticon.

AI Use Statement

Generative AI tools were used to support research ideation, hypothesis refinement, experimental-design feedback, code development and review, dataset and artifact preparation, result analysis and interpretation, literature discovery and summarization, and drafting and editing the manuscript. We also used independent language-model reviews as non-authoritative feedback on clarity, scope, and claim calibration. All AI-assisted work was reviewed by the authors; experiments were executed with the reported code and checkpoints, numerical claims were verified against frozen raw artifacts and analysis outputs, and citations and scientific conclusions were checked before inclusion. Generative AI was not used to fabricate or replace experimental measurements. The authors take responsibility for the final content of the paper and all accompanying artifacts.

References

Marianne Arriola, Aaron Gokaslan, Justin T. Chiu, Zhihan Yang, Zhixuan Qi, Jiaqi Han, Subham Sekhar Sahoo, and Volodymyr Kuleshov. Block diffusion: Interpolating between autoregressive and diffusion language models, 2025. URL https://arxiv.org/abs/2503.09573.

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces, 2023. URL https://arxiv.org/abs/2107.03006.

Shansan Gong, Shivam Agarwal, Yizhe Zhang, Jiacheng Ye, Lin Zheng, Mukai Li, Chenxin An, Peilin Zhao, Wei Bi, Jiawei Han, Hao Peng, and Lingpeng Kong. Scaling diffusion language models via adaptation from autoregressive models, 2025. URL https://arxiv.org/abs/2410.17891.

Zhibin Gou, Zhihong Shao, Yeyun Gong, Yelong Shen, Yujiu Yang, Nan Duan, and Weizhu Chen. Critic: Large language models can self-correct with tool-interactive critiquing, 2024. URL https://arxiv.org/abs/2305.11738.

Xiaochuang Han, Sachin Kumar, and Yulia Tsvetkov. Ssd-lm: Semi-autoregressive simplex-based diffusion language model for text generation and modular control, 2023. URL https://arxiv.org/abs/2210.17432.

Jie Huang, Xinyun Chen, Swaroop Mishra, Huaixiu Steven Zheng, Adams Wei Yu, Xinying Song, and Denny Zhou. Large language models cannot self-correct reasoning yet, 2024. URL https://arxiv.org/abs/2310.01798.

Zemin Huang, Yuhang Wang, Zhiyang Chen, and Guo-Jun Qi. Don't settle too early: Self-reflective remasking for diffusion language models, 2025. URL https://arxiv.org/abs/2509.23653.

Daniel Israel, Tian Jin, Ellie Cheng, Guy Van den Broeck, Aditya Grover, Suvinay Subramanian, and Michael Carbin. Planned diffusion, 2026. URL https://arxiv.org/abs/2510.18087.

Wonje Jeung, Sangyeon Yoon, Yoonjun Cho, Dongjae Jeon, Sangwoo Shin, Hyesoo Hong, and Albert No. A2d: Any-order, any-step safety alignment for diffusion language models, 2026. URL https://arxiv.org/abs/2509.23286.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions, 2020. URL https://arxiv.org/abs/1703.04730.

Tamera Lanham, Anna Chen, Ansh Radhakrishnan, Benoit Steiner, Carson Denison, Danny Hernandez, Dustin Li, Esin Durmus, Evan Hubinger, Jackson Kernion, et al. Measuring faithfulness in chain-of-thought reasoning, 2023. URL https://arxiv.org/abs/2307.13702.

Pengxiang Li, Yefan Zhou, Dilxat Muhtar, Lu Yin, Shilin Yan, Li Shen, Soroush Vosoughi, and Shiwei Liu. Diffusion language models know the answer before decoding, 2026a. URL https://arxiv.org/abs/2508.19982.

Xiang Lisa Li, John Thickstun, Ishaan Gulrajani, Percy Liang, and Tatsunori B. Hashimoto. Diffusion-lm improves controllable text generation, 2022. URL https://arxiv.org/abs/2205.14217.

Zherui Li, Zheng Nie, Zhenhong Zhou, Yue Liu, Yitong Zhang, Yu Cheng, Qingsong Wen, Kun Wang, Yufei Guo, and Jiaheng Zhang. Diffuguard: How intrinsic safety is lost and found in diffusion large language models, 2026b. URL https://arxiv.org/abs/2509.24296.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution, 2024. URL https://arxiv.org/abs/2310.16834.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. Self-refine: Iterative refinement with self-feedback, 2023. URL https://arxiv.org/abs/2303.17651.

Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. Locating and editing factual associations in gpt, 2023. URL https://arxiv.org/abs/2202.05262.

Shen Nie, Fengqi Zhu, Zebin You, Xiaolu Zhang, Jingyang Ou, Jun Hu, Jun Zhou, Yankai Lin, Ji-Rong Wen, and Chongxuan Li. Large language diffusion models, 2025. URL https://arxiv.org/abs/2502.09992.

Shen Nie, Qiyang Min, Shaoxuan Xu, Zihao Huang, Yuxuan Song, Yong Shan, Yankai Lin, Wayne Xin Zhao, Chongxuan Li, and Ji-Rong Wen. Improved large language diffusion models, 2026. URL https://arxiv.org/abs/2606.25331.

Debjit Paul, Robert West, Antoine Bosselut, and Boi Faltings. Making reasoning matter: Measuring and improving faithfulness of chain-of-thought reasoning, 2024. URL https://arxiv.org/abs/2402.13950.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models, 2024. URL https://arxiv.org/abs/2406.07524.

Miles Turpin, Julian Michael, Ethan Perez, and Samuel R. Bowman. Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting, 2023. URL https://arxiv.org/abs/2305.04388.

Shojiro Yamabe and Jun Sakuma. Toward safer diffusion language models: Discovery and mitigation of priming vulnerability, 2026. URL https://arxiv.org/abs/2510.00565.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. Hotpotqa: A dataset for diverse, explainable multi-hop question answering, 2018. URL https://arxiv.org/abs/1809.09600.

Jiacheng Ye, Zhihui Xie, Lin Zheng, Jiahui Gao, Zirui Wu, Xin Jiang, Zhenguo Li, and Lingpeng Kong. Dream 7b: Diffusion large language models, 2025. URL https://arxiv.org/abs/2508.15487.

Kevin Zhai, Sabbir Mollah, Zhenyi Wang, and Mubarak Shah. Core: Context-robust remasking for diffusion language models, 2026. URL https://arxiv.org/abs/2602.04096.

Fred Zhang and Neel Nanda. Towards best practices of activation patching in language models: Metrics and methods, 2024. URL https://arxiv.org/abs/2309.16042.

A Related Work

A.1 Generation Order and Early Commitment

Text diffusion includes continuous embedding denoising (Li et al., 2022) as well as discrete corruption processes, simplex diffusion, score-entropy modeling, and masked diffusion objectives (Austin et al., 2023; Han et al., 2023; Lou et al., 2024; Sahoo et al., 2024). Scaling and hybridization then produced adapted 7B diffusion models, blockwise diffusion, and instruction-tuned masked dLLMs (Gong et al., 2025; Arriola et al., 2025; Nie et al., 2025; Ye et al., 2025; Nie et al., 2026). These models replace a fixed left-to-right factorization with repeated prediction over bidirectional or blockwise canvases, making commitment order an inference variable. Planned Diffusion plans semantically independent chunks before parallel denoising, while Prophet exploits early answer convergence (Israel et al., 2026; Li et al., 2026a). This literature establishes generation order as consequential for quality, latency, and controllability. We study its alignment consequence: whether a decision remains responsive to evidence that arrives after endogenous rationale or answer commitments have formed.

A.2 Safety Along the Denoising Trajectory

Revision depends on both feedback and whether previously generated content can be revisited. Autoregressive work alternates generation with model-produced critique, tool-grounded verification, or unaided self-correction, with external feedback often determining whether refinement succeeds (Madaan et al., 2023; Gou et al., 2024; Huang et al., 2024). In masked dLLMs, monotone unmasking instead freezes early tokens; RemeDi learns an explicit remasking policy, while CoRe probes context brittleness for training-free revision (Huang et al., 2025; Zhai et al., 2026). Diffusion-specific safety work demonstrates that intermediate affirmative tokens, refusal signals, and safe/harmful token polarity can steer the final denoising path (Yamabe & Sakuma, 2026; Jeung et al., 2026; Li et al., 2026b). We study a complementary observable: evidence-conditioned corrigibility in controlled reasoning and natural question answering. Correct, corrupt, and absent evidence separate current-evidence control from generic regeneration, while answer and path resets separate decision-token openness from rationale-state openness.

A.3 Attribution and State Repair

Rationale faithfulness asks whether generated reasoning causally mediates a prediction rather than merely explaining it after the fact. Biasing-feature tests, direct chain-of-thought interventions, and causal mediation reveal substantial task- and model-dependent variation in how answers use stated reasoning (Turpin et al., 2023; Lanham et al., 2023; Paul et al., 2024). A parallel interpretability literature distinguishes attribution from behavioral intervention: influence functions trace training examples, causal tracing supports factual model editing, and activation-patching conclusions depend on corruption and metric choices (Koh & Liang, 2020; Meng et al., 2023; Zhang & Nanda, 2024). Our design applies that distinction to visible tokens within one denoising trajectory. We transplant exact rationale states, compare backward decision-masking with repair-aligned rationale-to-decision scoring, and remask selected tokens against cardinality- and compute-matched controls. The tested direct score improves local prioritization, but neither local objective identifies a compact sufficient intervention.

B Limitations and Broader Discussion

The evidence supports a trajectory-level mechanism within fully bidirectional masked diffusion language models. Dream, LLaDA, and iLLaDA provide three checkpoints with distinct model code and tokenizers, but they remain members of the same broad generative family; other diffusion parameterizations and larger model scales may organize persistent state differently. The controlled task isolates multi-step arithmetic, and the primary natural task converts extractive HotpotQA bridge questions into one-token multiple-choice decisions. A free-form 16-token stress test produces positive token-F1 effects but fails its frozen parsing-integrity gate, so it does not establish exact multi-token decision replication. Correct HotpotQA evidence contains original supporting facts, whereas corrupt evidence is a controlled answer-entity substitution rather than naturally occurring misinformation. Extending the mechanism to open-ended generation, non-extractive evidence,

multimodal contexts, and interactive evidence streams requires corresponding outcome and state-equivalence tests. Autoregressive models can also exhibit anchoring and revision failures; the diffusion-specific object isolated here is sampler-controlled commitment order within one shared response canvas.

Path reset is a causal intervention on generated state, not a ready deployment policy. It removes all free-rationale positions, can discard useful computation when evidence arrived early, and incurs a reconstruction cost whose acceptability depends on the application. The tested direct rationale-to-decision score outperforms matched random selection, whereas the decision-masking score does not. Nevertheless, no tested partial policy approaches the path-reset endpoint. This rules out compact sufficiency for the tested local objectives, not methods based on evidence compatibility, joint token interactions, learned state representations, or adaptive reset timing. Temperature-one generation across five seeds preserves the mechanism in Dream and LLaDA, but does not cover other samplers, temperatures, or stochastic schedules.

Restoring evidence control is intrinsically dual-use. Path reset makes correct evidence decisive, but the same intervention makes controlled corrupt evidence decisive; partial reopening also increases corrupt-evidence following even when it fails to repair correct-evidence behavior. A deployed revision mechanism therefore needs an independent account of evidence trustworthiness, provenance, and conflict, together with a policy for deciding when existing state is stale. This separates corrigibility from correctness: the former concerns whether current evidence can control a committed trajectory, while the latter depends on whether that evidence deserves control. More broadly, final-output evaluation is incomplete for diffusion language models because the sampler determines when generated text becomes persistent state. Alignment evaluations should measure commitment order, evidence arrival, retained trajectory history, and the ability to revise state under controlled interventions.

C Experimental Details

C.1 Confirmatory Experiment Inventory

The confirmatory chain contains 30,426 output trajectories, excluding software smoke runs, donor-prefix state construction, direct-attribution probe states, and earlier exploratory studies. Table 5 assigns one inferential role to each frozen stage. The sequence measures commitment timing, identifies and transplants retained state, tests evidence control on disjoint natural questions, separates attribution direction from repair sufficiency, and varies decoder and decision interfaces.

Table 5: Confirmatory experiment inventory. Counts include new trajectories in each frozen stage.

Stage	Models	Data	Primary role	Trajectories
Evidence arrival	Dream, LLaDA	Arithmetic-50	Answer commitment window	1,800
State reset	Dream, LLaDA	Arithmetic-50	Answer versus path causality	1,200
Natural pilot	Dream, LLaDA	HotpotQA-50	Natural-task mechanism	1,200
Backward attribution repair	Dream, LLaDA	HotpotQA-50	Coupling-score sufficiency	2,000
Third checkpoint	iLLaDA	HotpotQA-50	Model breadth	300
Disjoint holdout	Dream, LLaDA, iLLaDA	HotpotQA-150	Main factorial replication	4,050
Decoder schedule robustness	Dream, LLaDA	HotpotQA-50	Budget and reset timing	1,600
Rationale transplant	Dream, LLaDA, iLLaDA	HotpotQA-150	Organized content identity	4,500
Direct attribution	Dream, LLaDA	HotpotQA-50	Repair-aligned prioritization	6,000
Multi-token stress	Dream, LLaDA, iLLaDA	HotpotQA-88	Decision-interface scope	2,376
Stochastic replication	Dream, LLaDA	HotpotQA-50	Temperature/seeds	4,500
Transplant anchors	Dream, LLaDA, iLLaDA	HotpotQA-150	Runner integrity	900
Total	3 checkpoints	Controlled + natural		30,426

C.2 Models, Tasks, and Decision Interface

We evaluate the instruction-tuned Dream-v0-Instruct-7B, LLaDA-8B-Instruct, and iLLaDA-8B-Instruct checkpoints with their frozen tokenizers (Ye et al., 2025; Nie et al., 2025; 2026). Dream and LLaDA appear throughout the controlled and natural development experiments; iLLaDA is included in the third-checkpoint pilot, holdout, transplant, and multi-token replications. Primary generation uses greedy decoding at temperature zero. The stochastic replication fixes temperature one and seeds 0–4 in Dream and LLaDA.

The controlled task contains 50 depth-8 arithmetic chains. Each item presents four labeled options and a proposed wrong answer, together with a protected evidence slot that can reveal a correct checkpoint, a corrupt checkpoint supporting the proposed answer, or an explicit absent-evidence message. The natural task converts HotpotQA bridge questions (Yang et al., 2018) into four-way decisions with type-matched distractors. Primary experiments constrain the final decision to one A/B/C/D token, removing answer parsing, answer length, and within-answer token order from the measured outcome. The separate multi-token stress test instead generates the selected option text in an unconstrained 16-token decision region.

C.3 Prompt Semantics and State Layout

Every one-token item presents the proposed option as an untrusted user belief and instructs the model to verify it independently from the question, choices, and any evidence that later enters the protected slot. The question, option order, proposal, field labels, and response length are fixed within each paired experimental unit. The assistant canvas explicitly separates externally supplied evidence E, model-generated free rationale R, and the constrained decision D. This layout makes the causal scopes concrete: evidence arrival changes only E, answer reset changes only D, and path reset changes $R \cup D$. Exact prompt strings and chat-template handling are released with the code.

C.4 Natural-Task Construction and Integrity

We retain HotpotQA distractor-development bridge questions with at least two supporting titles. We exclude yes/no questions, questions that contain the gold answer, and examples whose annotated supporting facts do not contain the answer verbatim. Four high-precision answer buckets—person, date/year, place, and number—provide type-matched distractors. Options are shuffled deterministically before model inference.

Correct evidence contains the original annotated supporting sentences and titles. Corrupt evidence replaces each standalone occurrence of the gold answer in those sentences and titles with the designated type-matched wrong option; the remaining evidence and question are unchanged. This condition is a controlled counterfactual entity substitution. Absent evidence explicitly states that no verified supporting passage is available.

The 50-item development artifact has person/date/place/number composition 13/13/12/12. The 150-item holdout excludes every development source ID, has zero source overlap, and has composition 38/38/37/37. No item is selected, removed, or replaced using model output. Exact artifact fingerprints and source-ID audits are released with the data.

The multi-token artifact excludes all 200 source IDs in the development and holdout artifacts. Gold and designated-wrong option texts must each tokenize to 2–8 tokens under all three frozen tokenizers. The number bucket supplies 22 viable examples after this metadata-only audit, so the preregistered fallback freezes the largest balanced subset: 88 items with 22 person, date, place, and number questions each. No model output enters construction.

C.5 Canvas, Evidence Arrival, and Reset Schedule

Controlled arithmetic uses 32 denoising steps, 96 generated reasoning positions, a protected 32-token evidence region, and a one-token decision. The evidence-arrival sweep reveals the protected region at steps 2, 4, 8, 12, 16, or 20. The state-reset factorial uses early arrival at step 2 or late arrival at step 20 and resets at step 24.

Natural experiments reserve 160 evidence tokens and 64 free-rationale tokens. The slot size was frozen before inference after tokenizer audits found maximum development evidence lengths of 153 tokens for Dream and 157 for LLaDA. On the holdout, maximum lengths are 154, 154, and 159 for Dream, LLaDA, and iLLaDA. Every evidence text therefore fits the common 160-token slot, and every condition retains the same 64-token free-rationale budget.

Before reset, endogenous positions follow the standard confidence commitment policy. Answer reset masks the decision token and retains the free rationale; path reset masks the decision and all free-rationale positions. Both reset arms then use the same reasoning-first reconstruction schedule. Evidence, prompt tokens, and answer labels remain fixed. The no-reset holdout arm continues the

standard confidence decoder and provides an operational baseline rather than the matched reset-scope comparison.

Evidence injection and reset occur before the model forward pass at their indexed step. Before arrival, E is excluded from endogenous commitment. The base decoder ranks selectable masked positions by their maximum non-mask token probability under a shared mask schedule. After either reset, rationale positions take priority over D, with confidence breaking ties within the active region; the final step commits every remaining mask. Decision logits are restricted to the four label tokens in the primary experiments. These rules are identical across evidence identities and reset scopes, so the matched comparison changes retained state rather than reconstruction policy.

C.6 Rationale-Transplant Protocol

Transplant donors use the 150-item holdout, natural-confidence commitment, early evidence at step 2, and correct, corrupt, or absent evidence. We capture the exact step-24 token canvas in memory. Recipients retain correct or corrupt evidence, their fixed prompt, decision position, and post-intervention seed. We replace the 64 free-rationale positions with one of five states: self-evidence donor, opposite-evidence donor, absent-evidence donor, shuffled opposite donor, or masks. The decision is masked in every condition and reconstruction uses the same reasoning-first policy.

The shuffled control randomly permutes committed opposite-donor token IDs over the same committed positions, preserving the token multiset, filled-position count, and masked-position pattern. Shuffle seeds are deterministic functions of model, item, and generation seed and are stored with the permutation. Donor decoded strings are never retokenized. The complete design contains 3 models $\times$ 150 items $\times$ 2 recipient-evidence conditions $\times$ 5 rationale conditions = 4,500 recipient trajectories. A separate standard-runner audit executes 900 path-reset anchors with evidence arriving at step 2 and reset at step 24, matching the transplantation configuration for both recipient-evidence conditions.

C.7 Directional Attribution and Repair Policies

Partial-repair experiments use late evidence at step 20 and reset at step 24 on the 50-item natural artifact. Every policy also resets the decision token. The original score masks the decision and measures the change in observed rationale-token likelihood. Its fixed-budget policies reset the top 25%, 50%, or 75% of the 64 committed rationale positions; mass-80 resets the shortest prefix containing 80% of measured backward coupling. Cardinality-matched random, low-confidence 50%, and recent-token 50% provide controls.

The direct score first masks the decision, records the constrained log probability of its committed pre-reset label, and then masks each committed rationale token in turn. Its absolute leave-one-out log-probability change is the rationale-to-decision score. Direct policies use top-25/50/75 and mass-80 sets. Five independently seeded random subsets are drawn for every fixed quota and five additional random subsets match the trajectory-specific direct mass-80 count. Original top-50/mass-80, low-confidence, recent-token, answer-reset, and path-reset policies serve as anchors. This yields 30 repair modes for each model, item, and evidence condition, or 6,000 terminal repair trajectories. Random controls use the same committed candidates and exact reset counts as the corresponding attribution policy.

Direct scoring evaluates one decision-masked baseline and one leave-one-out state per committed rationale candidate, up to 65 model-evaluated states per unit. We report these attribution probes separately from the base 32-step decoding budget. The direct-versus-random tests therefore identify ranking selectivity at matched intervention size; they are not claims of lower inference cost.

C.8 Stochastic and Multi-Token Protocols

The stochastic replication fixes temperature one and seeds 0–4 for Dream and LLaDA on the 50-item natural artifact. Evidence is correct, corrupt, or absent; interventions are no reset, answer reset, or path reset. The resulting 4,500 trajectories share the same step-indexed random-generator construction within each model/item/seed. Statistical intervals resample items and retain all five seeds in each sampled cluster.

The multi-token stress test uses the 88-item disjoint artifact in all three models. The prompt requests the complete selected option text rather than its label, and the model fills an unconstrained 16-token decision region. Parsing first tests normalized exact match, then a unique normalized option substring; every other output is invalid and retained as incorrect. Token F1 against each option is always reported. Evidence conditions and reset interventions form a complete 3×3 factorial, yielding 2,376 trajectories. The parser, 95% integrity threshold, and treatment of invalid outputs were frozen before inference.

C.9 Outcomes, Pairing, and Statistical Tests

The primary outcomes are exact decision accuracy and following of the designated wrong proposal. Arrival-history analyses additionally compare final decisions and complete decoded rationale-plus-decision canvases. Transplant analyses use symmetric donor-consistent accuracy and following contrasts; direct analyses average each item's five matched-random outcomes before computing the paired effect. Multi-token analyses additionally report gold token F1 and valid-parse rate.

Every primary contrast is paired by model, item, and seed. Binary paired outcomes use exact McNemar tests. Mean differences and differences-in-differences use 10,000 bootstrap samples; model-specific intervals resample items. For the main holdout and follow-up pooled effects, source items are clusters: every evaluated model, evidence condition, intervention, and—under stochastic decoding—all five seeds for an item move together. These intervals generalize over source items while conditioning on the tested checkpoints and fixed seeds; model-level rows are also reported separately. Protocols, thresholds, parsers, random seeds, and directional gates are frozen before each confirmatory run. Every formal result file is complete; the independent transplant audit pairs all 900 masked-transplant decisions with their separately executed path-reset anchors without a mismatch. Failed H1, S1, M1–M3, DA3, and MT4 gates remain in the evidence ledger and constrain the manuscript claims.

C.10 Reproducibility Artifact

All model forwards use BF16 and one GPU per process; the experiments ran on NVIDIA A100 40GB and H800 GPUs. The anonymous artifact releases the runners, frozen protocols, constructed datasets, path-sanitized trajectories, and reference summaries. Its reproduction entry point verifies raw checksums and regenerates every reported aggregate, data figure, generated appendix table, and frozen-predicate case report. Exact prompts, sampler pseudocode, seed construction, checkpoint revisions, software versions, batching, memory use, and resume behavior are documented with the code rather than repeated here.

D Additional Experimental Results

D.1 Complete Evidence-Arrival Results

Table 6 reports every cell in the six-point controlled sweep. Dream's correct-evidence gain contracts as its open-decision fraction falls, while LLaDA closes the decision channel earlier and does not exhibit a positive correct-evidence curve. The preregistered shared cross-model answer-deadline hypothesis is rejected. The commitment boundary itself is exact: all 395 decisions committed before evidence arrival remain invariant across correct, corrupt, and absent evidence.

Table 6: Complete evidence-arrival sweep. All entries are percentages; C/K/0 denote correct, corrupt, and absent evidence.

Model	Step	Open	Acc. C	Acc. K	Acc. 0	Follow C	Follow K	Follow 0
Dream	2	78	62	34	26	24	32	32
Dream	4	66	44	32	26	26	32	26
Dream	8	66	50	28	28	26	36	24
Dream	12	58	46	38	22	24	30	26
Dream	16	42	34	24	24	26	26	26
Dream	20	16	20	20	20	34	32	32
LLaDA	2	22	12	10	14	70	72	68
LLaDA	4	20	14	12	16	70	72	68
LLaDA	8	12	16	14	18	68	70	66
LLaDA	12	12	16	16	18	68	68	66
LLaDA	16	10	14	14	16	68	68	68
LLaDA	20	8	16	18	16	68	66	68

D.2 Complete State-Reset Results

Table 7 gives the full controlled arithmetic factorial. The result separates accuracy improvement from state control. In Dream, removing a rationale built from early correct evidence lowers accuracy from 64% to 38%; under late evidence, path reset raises accuracy from 26% to 38%. Across the late correct/corrupt comparison, path reset produces positive pooled accuracy and following control interactions even though unconditional improvement is not shared by every model and arrival time.

Table 7: Complete controlled state-reset factorial. Each evidence cell reports accuracy/wrong-proposal following in percent.

Model	Arrival	Reset	Correct	Corrupt	Absent
Dream	2	Answer	64/20	32/30	26/28
Dream	2	Path	38/36	14/42	20/44
Dream	20	Answer	26/28	26/26	26/26
Dream	20	Path	38/36	14/42	20/44
LLaDA	2	Answer	14/70	10/74	14/68
LLaDA	2	Path	56/30	20/72	22/38
LLaDA	20	Answer	16/68	18/66	16/68
LLaDA	20	Path	56/30	20/72	22/38
Pooled	2	Answer	39/45	21/52	20/48
Pooled	2	Path	47/33	17/57	21/41
Pooled	20	Answer	21/48	22/46	21/47
Pooled	20	Path	47/33	17/57	21/41

Under path reset, early- and late-arrival outputs agree in 300/300 decisions and 300/300 complete rationale-plus-decision canvases. Under answer reset, decisions disagree in 42.7% of Dream pairs, 10.7% of LLaDA pairs, and 26.7% pooled. Removing the path therefore erases arrival history exactly; reopening the answer retains it.

D.3 Natural-Task and Holdout Results

The complete 50-item natural factorial appears in Table 8. At late arrival, path reset reaches 100% correct-evidence accuracy in both Dream and LLaDA, while corrupt-evidence following reaches 92% and 94%. Absent-evidence accuracy remains 24% and 28%, so reconstruction is not a generic accuracy gain. Answer-reset early/late disagreement is 21.7% pooled; path-reset decisions and complete rationales remain identical in every pair.

Table 8: Complete HotpotQA-50 state-reset factorial. Each evidence cell reports accuracy/wrong-proposal following in percent.

Model	Arrival	Reset	Correct	Corrupt	Absent
Dream	2	Answer	82/14	24/74	38/48
Dream	2	Path	100/0	8/92	24/38
Dream	20	Answer	58/32	36/54	38/46
Dream	20	Path	100/0	8/92	24/38
LLaDA	2	Answer	56/40	24/72	30/64
LLaDA	2	Path	100/0	6/94	28/62
LLaDA	20	Answer	36/62	26/70	28/64
LLaDA	20	Path	100/0	6/94	28/62
Pooled	2	Answer	69/27	24/73	34/56
Pooled	2	Path	100/0	7/93	26/50
Pooled	20	Answer	47/47	31/62	33/55
Pooled	20	Path	100/0	7/93	26/50

Table 9 reports all 27 raw cells from the source-disjoint three-model holdout, together with pooled rates. Every no-reset decision is committed before step-20 evidence, making the no-reset output exactly evidence-invariant. Answer reset creates partial evidence responsiveness. Path reset exposes the largest correct/corrupt separation in every checkpoint while lowering pooled absent-evidence accuracy relative to no reset.

The effect is present across all preregistered answer types (Table 10). Dates produce the largest correct-specific and evidence-control effects, but the source-item bootstrap interval remains positive for person, date, place, and number questions.

Reconstruction of the pre-reset answer. In a supplementary post-hoc analysis, we compare each rebuilt answer with its recorded pre-reset label on the frozen holdout. Both reset arms begin with the same label in all 1,350 paired model–item–evidence units. Table 11 reports the fraction that reproduce this label, whether correct or incorrect. This directly measures reconstruction of an earlier answer; the early–late comparisons above instead measure sensitivity to arrival history. Intervals for the paired answer-minus-path difference resample the 150 source items, keeping all model and evidence cells together. This analysis reuses existing trajectories.

With correct or controlled corrupt evidence, all three models more often reproduce the prior label after answer reset. Under absent evidence, iLLaDA reverses this ordering: answer reset reproduces the label in 57.3% of trajectories, compared with 94.0% after path reset.

Table 9: Complete three-model HotpotQA holdout cells. Each evidence cell reports accuracy/wrong-proposal following in percent.

Model	Intervention	Correct	Corrupt	Absent
Dream	No reset	53.3/36.7	53.3/36.7	53.3/36.7
Dream	Answer	64.7/28.7	40.7/51.3	47.3/40
Dream	Path	97.3/2	13.3/86	38.7/38
LLaDA	No reset	38/60.7	38/60.7	38/60.7
LLaDA	Answer	48.7/47.3	36.7/58.7	39.3/50.7
LLaDA	Path	95.3/4	7.3/90.7	34.7/44
iLLaDA	No reset	32.7/22.7	32.7/22.7	32.7/22.7
iLLaDA	Answer	53.3/29.3	34/43.3	39.3/33.3
iLLaDA	Path	97.3/1.3	5.3/93.3	27.3/24.7
Pooled	No reset	41.3/40	41.3/40	41.3/40
Pooled	Answer	55.6/35.1	37.1/51.1	42/41.3
Pooled	Path	96.7/2.4	8.7/90	33.6/35.6

Table 10: Holdout effects by answer type. Effects are percentage points with 95% bootstrap intervals; pooled intervals resample source items as clusters and n counts source items.

Type	n	Correct-specific	Accuracy control	Following control
Date	38	+66.7 [57.0, 76.3]	+85.1 [77.2, 92.1]	+86.0 [78.1, 93.0]
Number	37	+49.5 [37.8, 60.4]	+72.1 [62.2, 81.1]	+73.9 [64.9, 82.0]
Person	38	+39.5 [28.1, 50.9]	+58.8 [47.4, 69.3]	+61.4 [49.1, 72.8]
Place	37	+42.3 [33.3, 51.4]	+62.2 [52.3, 71.2]	+64.9 [55.9, 73.0]

Table 11: Reconstruction of the pre-reset answer on the three-model holdout. Entries report matching labels/total trajectories (%). The paired answer-minus-path gap is in percentage points with 95% source-item bootstrap intervals. Each model contributes 150 items; pooled rows combine 450 paired units.

Model	Evidence	Answer same/n (%)	Path same/n (%)	Gap [95% CI]
Dream	Correct	125/150 (83.3%)	79/150 (52.7%)	$+30.7[22.7, 38.7]$
Dream	Corrupt	122/150 (81.3%)	71/150 (47.3%)	$+34.0[25.3, 42.7]$
Dream	Absent	126/150 (84.0%)	81/150 (54.0%)	$+30.0[20.0, 39.3]$
LLaDA	Correct	126/150 (84.0%)	63/150 (42.0%)	$+42.0[34.0, 50.7]$
LLaDA	Corrupt	138/150 (92.0%)	97/150 (64.7%)	$+27.3[18.7, 36.0]$
LLaDA	Absent	129/150 (86.0%)	81/150 (54.0%)	$+32.0[22.0, 42.0]$
iLLaDA	Correct	72/150 (48.0%)	53/150 (35.3%)	$+12.7[4.0, 21.3]$
iLLaDA	Corrupt	81/150 (54.0%)	39/150 (26.0%)	$+28.0[18.7, 37.3]$
iLLaDA	Absent	86/150 (57.3%)	141/150 (94.0%)	$-36.7[-46.0, -27.3]$
Pooled	Correct	323/450 (71.8%)	195/450 (43.3%)	$+28.4[23.6, 33.3]$
Pooled	Corrupt	341/450 (75.8%)	207/450 (46.0%)	$+29.8[24.4, 35.6]$
Pooled	Absent	341/450 (75.8%)	303/450 (67.3%)	$+8.4[2.4, 14.4]$

D.4 Rationale Transplantation and Anchor Integrity

Rationale identity controls reconstructed decisions in every checkpoint. Table 12 reports the four preregistered directional contrasts. Dream, LLaDA, and iLLaDA each have positive intervals for

self-to-opposite donor-consistent accuracy and following (T1) and for coherent-opposite versus shuffled accuracy and following (T2). The pooled effects are $+17.8/+18.9$ points for T1 and $+28.3/+27.7$ points for T2.

Table 12: Rationale-transplant directional effects. T1 is the donor-consistent self-to-opposite effect; T2 compares coherent opposite donors with shuffled-token controls. Entries are percentage points with 95% bootstrap intervals; pooled intervals resample source items as clusters.

Model	T1 Acc.	T1 Follow	T2 Acc.	T2 Follow
Dream	$+18.0$ [11.7, 24.3]	$+18.7$ [12.0, 25.3]	$+26.0$ [20.7, 31.7]	$+26.0$ [20.7, 31.7]
LLaDA	$+20.7$ [14.7, 26.7]	$+20.7$ [14.7, 26.7]	$+38.7$ [33.7, 44.0]	$+38.0$ [32.7, 43.3]
iLLaDA	$+14.7$ [9.7, 20.0]	$+17.3$ [12.3, 22.7]	$+20.3$ [15.0, 25.7]	$+19.0$ [13.7, 24.3]
Pooled	$+17.8$ [14.4, 21.1]	$+18.9$ [15.4, 22.4]	$+28.3$ [25.0, 31.8]	$+27.7$ [24.3, 31.1]

Table 13: Pooled terminal rates for all transplant cells. Accuracy and wrong-following are percentages; every cell contains 450 model–item units.

Recipient / rationale	n	Accuracy	Wrong follow
Correct / Absent Donor	450	61.8	28.7
Correct / Masked	450	96.9	2.4
Correct / Opposite	450	54.0	39.8
Correct / Self	450	70.4	22.0
Correct / Shuffled Opposite	450	86.2	9.1
Corrupt / Absent Donor	450	35.1	52.4
Corrupt / Masked	450	9.1	89.1
Corrupt / Opposite	450	48.2	43.6
Corrupt / Self	450	29.1	63.6
Corrupt / Shuffled Opposite	450	23.8	68.2

The raw pooled rates in Table 13 show both recipient directions. A corrupt donor reduces correct-recipient accuracy from 70.4% to 54.0%, while a correct donor raises corrupt-recipient accuracy from 29.1% to 48.2%. The coherent opposite donor exceeds its shuffled-token control while preserving committed positions and token multiset, isolating an effect of sequence organization. Absent-donor accuracy lies between the correct- and corrupt-donor endpoints, satisfying T3.

The masked transplant is compared with a separately executed path-reset anchor for every model, item, and recipient evidence condition. Both runs receive evidence at step 2 and reset at step 24 under the transplantation configuration. All 900/900 decisions agree: 300/300 in Dream, LLaDA, and iLLaDA, with no unmatched cells. T4 is an implementation check over the full transplantation experiment. These anchors are distinct runs from the step-20 holdout cells in Table 9.

D.5 Directional Attribution and Partial Repair

Tables 14–16 report every partial-repair policy for correct and corrupt evidence. Attribution mass is highly concentrated in both checkpoints, yet attribution-guided selection does not define a superior repair frontier. At the 50% budget, it exceeds matched random accuracy by 6 points in Dream and trails by 8 points in LLaDA; pooled accuracy is 53% versus 54%. The adaptive mass-80 policy ties its random baseline in Dream, trails by 8 points in LLaDA, and reaches 45% versus 49% pooled. These model directions preserve the failed preregistered localization gates rather than pooling them away.

Table 14: Selective-repair results for Dream. Reset counts show correct/corrupt values when they differ; attribution mass and outcome rates are percentages; A/F denotes accuracy/following.

Policy	Rationale reset	Probes	NFE	Attr. mass	Correct A/F	Corrupt A/F
Answer reset	0.0	0	32	–	58/32	36/54
Path reset	64.0	0	32	–	100/0	8/92
Attribution 25	16.0	2	34	98.5	56/34	32/58
Attribution 50	32.0	2	34	99.8	64/28	28/64
Attribution 75	48.0	2	34	100.0	76/20	24/72
Random 25	16.0	2	34	23.3	54/34	32/54
Random 50	32.0	2	34	48.3	58/34	40/52
Random 75	48.0	2	34	78.1	76/16	28/66
Low confidence 50	32.0	1	33	–	64/30	28/64
Recent 50	32.0	0	32	–	72/26	28/70
Attribution mass-80	4.9/5.1	2	34	83.8	58/32	36/54
Random mass-80	4.9/5.1	2	34	6.8	58/32	34/56

Table 15: Selective-repair results for LLaDA. Reset counts show correct/corrupt values when they differ; attribution mass and outcome rates are percentages; A/F denotes accuracy/following.

Policy	Rationale reset	Probes	NFE	Attr. mass	Correct A/F	Corrupt A/F
Answer reset	0.0	0	32	–	36/62	26/70
Path reset	64.0	0	32	–	100/0	6/94
Attribution 25	16.0	2	34	96.1	34/64	24/74
Attribution 50	32.0	2	34	99.2	42/54	22/74
Attribution 75	48.0	2	34	99.9	58/38	16/80
Random 25	16.0	2	34	25.6	42/54	24/72
Random 50	32.0	2	34	48.7	50/48	24/74
Random 75	48.0	2	34	74.7	60/38	22/76
Low confidence 50	32.0	1	33	–	42/54	22/76
Recent 50	32.0	0	32	–	46/50	22/76
Attribution mass-80	6.1/5.8	2	34	84.7	32/64	26/70
Random mass-80	6.1/5.8	2	34	9.2	40/58	26/70

Table 16: Selective-repair results for Pooled. Reset counts show correct/corrupt values when they differ; attribution mass and outcome rates are percentages; A/F denotes accuracy/following.

Policy	Rationale reset	Probes	NFE	Attr. mass	Correct A/F	Corrupt A/F
Answer reset	0.0	0	32	–	47/47	31/62
Path reset	64.0	0	32	–	100/0	7/93
Attribution 25	16.0	2	34	97.3	45/49	28/66
Attribution 50	32.0	2	34	99.5	53/41	25/69
Attribution 75	48.0	2	34	100.0	67/29	20/76
Random 25	16.0	2	34	24.5	48/44	28/63
Random 50	32.0	2	34	48.5	54/41	32/63
Random 75	48.0	2	34	76.4	68/27	25/71
Low confidence 50	32.0	1	33	–	53/42	25/70
Recent 50	32.0	0	32	–	59/38	25/73
Attribution mass-80	5.5/5.5	2	34	84.3	45/48	31/62
Random mass-80	5.5/5.5	2	34	8.0	49/45	30/63

Partial reset remains dual-use. Under attribution-guided 50% reset, corrupt evidence increases wrong-proposal following by 36 points in Dream, 20 points in LLaDA, and 28 points pooled. Under mass-80 reset, the corresponding effects are 22, 6, and 14 points. Evidence responsiveness can therefore increase without producing a sufficient repair set.

Direct rationale-to-decision attribution reverses the failed selectivity result without producing a compact endpoint. Relative to the mean of five matched random subsets, direct top-50 improves

correct-evidence accuracy by $+16.8$ points with interval $[9.0, 25.2]$; direct mass-80 improves it by $+18.2$ points with interval $[11.0, 26.0]$. Both model-specific point estimates are nonnegative. The direct and decision-masking rankings differ substantially: mean Spearman correlation is 0.115, top-quarter overlap is 30.0%, and top-half overlap is 55.0%.

Table 17: Complete pooled direct-attribution repair results over Dream and LLaDA. Reset is the mean number of rationale tokens reopened under correct evidence. Outcomes are percentages. Each random row is retained separately rather than averaged away.

Policy	Reset	Correct acc.	Correct follow	Corrupt acc.	Corrupt follow
Direct 25%	16.0	49.0	44.0	27.0	70.0
Direct 50%	32.0	66.0	30.0	22.0	75.0
Direct 75%	48.0	70.0	25.0	20.0	76.0
Direct mass-80	30.7	65.0	30.0	22.0	74.0
Random 25%, seed 0	16.0	44.0	51.0	31.0	63.0
Random 25%, seed 1	16.0	45.0	49.0	29.0	64.0
Random 25%, seed 2	16.0	45.0	49.0	29.0	65.0
Random 25%, seed 3	16.0	48.0	46.0	28.0	67.0
Random 25%, seed 4	16.0	46.0	49.0	29.0	66.0
Random 50%, seed 0	32.0	51.0	43.0	29.0	65.0
Random 50%, seed 1	32.0	47.0	47.0	31.0	65.0
Random 50%, seed 2	32.0	51.0	44.0	28.0	67.0
Random 50%, seed 3	32.0	46.0	48.0	28.0	64.0
Random 50%, seed 4	32.0	51.0	45.0	26.0	70.0
Random 75%, seed 0	48.0	56.0	37.0	22.0	72.0
Random 75%, seed 1	48.0	60.0	35.0	24.0	69.0
Random 75%, seed 2	48.0	60.0	36.0	23.0	72.0
Random 75%, seed 3	48.0	60.0	36.0	24.0	68.0
Random 75%, seed 4	48.0	61.0	34.0	24.0	70.0
Random mass-count, seed 0	30.7	44.0	51.0	29.0	65.0
Random mass-count, seed 1	30.7	48.0	47.0	26.0	69.0
Random mass-count, seed 2	30.7	49.0	45.0	29.0	65.0
Random mass-count, seed 3	30.7	47.0	46.0	27.0	66.0
Random mass-count, seed 4	30.7	46.0	48.0	25.0	68.0
Decision-mask 50%	32.0	49.0	45.0	26.0	69.0
Decision-mask mass-80	5.3	43.0	51.0	29.0	65.0
Low confidence 50%	32.0	54.0	41.0	28.0	66.0
Recent 50%	32.0	55.0	42.0	23.0	75.0
Answer reset	0.0	46.0	49.0	30.0	64.0
Path reset	64.0	99.0	0.0	7.0	93.0

Table 17 retains all 30 policies, including each random seed. Direct accuracy rises from 49% to 66% to 70% over the 25/50/75 budgets, while path reset reaches 99%. Direct mass-80 resets 30.7 tokens on average and reaches 65%. DA1 and DA2 pass, but no direct partial policy comes within five points of path reset, so DA3 fails. Under corrupt evidence, direct top-50 and mass-80 wrong-proposal following rises to 75% and 74%, satisfying the dual-use DA4 gate.

D.6 Decoder and Timing Robustness

Tables 18–20 contain the complete reset-time and denoising-budget comparison. Reset steps 22, 24, 26, and 28 preserve positive path-minus-answer evidence-control interactions in both models. With only four reconstruction steps after reset 28, pooled accuracy and following control remain $+76$ and $+77$ points. Doubling the decoder to 64 steps with proportional evidence/reset timing yields $+88$ and $+89$ points pooled.

Table 18: Decoder robustness for Dream. Configuration is total/evidence/reset steps; cells report accuracy/following in percent and controls are percentage-point interactions with 95% intervals.

Config.	Post	Answer C	Answer K	Path C	Path K	Acc. control	Follow control
32/20/22	10	58/32	36/54	96/2	6/94	+68.0 [54.0, 80.0]	+70.0 [56.0, 82.0]
32/20/24	8	58/32	36/54	100/0	8/92	+70.0 [56.0, 82.0]	+70.0 [56.0, 82.0]
32/20/26	6	58/32	36/54	100/0	8/92	+70.0 [56.0, 82.0]	+70.0 [56.0, 82.0]
32/20/28	4	58/32	36/54	100/0	8/92	+70.0 [56.0, 82.0]	+70.0 [56.0, 82.0]
64/40/48	16	34/50	32/52	96/2	4/96	+90.0 [80.0, 98.0]	+92.0 [84.0, 98.0]

Table 19: Decoder robustness for LLaDA. Configuration is total/evidence/reset steps; cells report accuracy/following in percent and controls are percentage-point interactions with 95% intervals.

Config.	Post	Answer C	Answer K	Path C	Path K	Acc. control	Follow control
32/20/22	10	36/62	26/70	98/2	6/94	+82.0 [70.0, 94.0]	+84.0 [72.0, 96.0]
32/20/24	8	36/62	26/70	100/0	6/94	+84.0 [72.0, 94.0]	+86.0 [74.0, 96.0]
32/20/26	6	36/62	26/70	100/0	8/92	+82.0 [70.0, 94.0]	+84.0 [72.0, 94.0]
32/20/28	4	36/62	26/70	96/4	4/96	+82.0 [70.0, 92.0]	+84.0 [72.0, 94.0]
64/40/48	16	30/68	26/72	98/2	8/92	+86.0 [76.0, 94.0]	+86.0 [76.0, 94.0]

Table 20: Decoder robustness for Pooled. Configuration is total/evidence/reset steps; cells report accuracy/following in percent and controls are percentage-point interactions with 95% intervals.

Config.	Post	Answer C	Answer K	Path C	Path K	Acc. control	Follow control
32/20/22	10	47/47	31/62	97/2	6/94	+75.0 [64.0, 85.0]	+77.0 [67.0, 86.0]
32/20/24	8	47/47	31/62	100/0	7/93	+77.0 [68.0, 86.0]	+78.0 [69.0, 87.0]
32/20/26	6	47/47	31/62	100/0	8/92	+76.0 [66.0, 86.0]	+77.0 [67.0, 86.0]
32/20/28	4	47/47	31/62	98/2	6/94	+76.0 [66.0, 85.0]	+77.0 [68.0, 86.0]
64/40/48	16	32/59	29/62	97/2	6/94	+88.0 [80.0, 95.0]	+89.0 [81.0, 95.0]

Temperature-one sampling preserves the mechanism across five frozen seeds. Table 21 reports item-clustered model and pooled effects. The pooled correct-specific effect is $+13.8$ points with interval $[8.2, 19.4]$; accuracy and following evidence-control interactions are $+17.8$ and $+20.2$ points with intervals $[10.8, 24.6]$ and $[14.4, 26.2]$. Both models have positive point estimates and intervals for all three effects.

Table 21: Temperature-1 item-clustered effects. Entries are percentage points with item-bootstrap 95% intervals; all five seeds remain inside each sampled item cluster.

Model	O1 correct-specific	O2 accuracy	O2 following
Dream	$+11.6$ $[4.4, 18.8]$	$+12.4$ $[4.4, 20.4]$	$+14.4$ $[7.2, 21.6]$
LLaDA	$+16.0$ $[6.8, 25.2]$	$+23.2$ $[14.8, 32.0]$	$+26.0$ $[18.4, 33.6]$
Pooled	$+13.8$ $[8.2, 19.4]$	$+17.8$ $[10.8, 24.6]$	$+20.2$ $[14.4, 26.2]$

Table 22: Pooled temperature-1 effects by frozen sampling seed. Effects are percentage points over 100 Dream/LLaDA item units. Four of five O1 values and all O2 values have the predicted sign.

Seed	n	O1	O2 accuracy	O2 following
0	100	-3.0	+1.0	+6.0
1	100	+23.0	+27.0	+25.0
2	100	+21.0	+20.0	+21.0
3	100	+18.0	+22.0	+27.0
4	100	+10.0	+19.0	+22.0

Four of five seed-level O1 effects and all five O2 accuracy and following effects have the predicted sign (Table 22). Under absent evidence, path-reset accuracy exceeds no reset by 4.8 points, within the preregistered five-point bound. SR1–SR4 all pass.

D.7 Multi-Token Decision Stress Test

The free-form decision stress test produces evidence-conditioned span changes across all three models. The pooled correct-specific token-F1 effect is $+26.0$ points with interval $[22.9, 29.2]$ (Table 23); every model has a positive token-F1 interval. Exact O1 and the exact accuracy/following O2 interactions also have positive intervals under the frozen parser, while absent-evidence path accuracy is lower than no reset. MT1–MT3 therefore have the preregistered directions.

Table 23: Free-form multi-token stress-test effects and parsing integrity. Effects are percentage points with 95% bootstrap intervals; pooled intervals resample source items as clusters. Exact effects are shown for completeness but are not licensed as confirmatory because every model violates the frozen 95% per-cell parsing gate.

Model	Exact O1	Token-F1 O1	O2 acc.	O2 follow	Parse range
Dream	$+46.6$ $[34.1, 58.0]$	$+22.4$ $[16.5, 28.1]$	$+50.0$ $[37.5, 61.4]$	$+40.9$ $[29.5, 52.3]$	18.2–79.5
LLaDA	$+59.1$ $[45.5, 71.6]$	$+34.1$ $[28.3, 39.7]$	$+65.9$ $[54.5, 76.1]$	$+60.2$ $[50.0, 70.5]$	54.5–85.2
iLLaDA	$+42.0$ $[30.7, 53.4]$	$+21.6$ $[17.6, 25.9]$	$+44.3$ $[33.0, 55.7]$	$+28.4$ $[14.8, 40.9]$	1.1–88.6
Pooled	$+49.2$ $[41.7, 56.4]$	$+26.0$ $[22.9, 29.2]$	$+53.4$ $[46.2, 60.2]$	$+43.2$ $[35.6, 50.8]$	24.6–82.6

Parsing integrity fails decisively. Per-cell valid-parse rates range from 18.2–79.5% in Dream, 54.5–85.2% in LLaDA, and 1.1–88.6% in iLLaDA, below the 95% requirement in every model. MT4 fails. We retain every invalid output as incorrect and report the exact effects for completeness, but use this experiment only as token-F1 span evidence rather than an exact multi-token decision replication.

D.8 Additional Tasks and Trajectory Cases

Earlier controlled experiments test when commitment order matters. Across 4,800 arithmetic and state-tracking trajectories, neutral-controlled wrong-suggestion attraction grows with arithmetic depth: the per-doubling slopes are 0.158 for Dream, 0.114 for LLaDA, and 0.136 pooled, with all bootstrap intervals excluding zero. The same gate does not hold for four-box state tracking; its pooled slope is 0.018 with interval $[-0.023, 0.060]$. Commitment order therefore matters when the model has usable intermediate computation, rather than as a universal function of nominal task depth.

Four qualitative cases are selected by frozen behavioral predicates and then by the lowest matching source ID. They illustrate answer-level correction, corrupt-evidence control after path reset, three-way evidence identity control in iLLaDA, and the failure of a high-mass sparse repair set. Table 24 reports their decision-level outcomes.

Table 24: Frozen-predicate trajectory cases. Letters are final constrained decisions; parentheses mark correct (C), wrong-proposal following (F), or neither (N).

Case	Model/evidence	No reset	Partial intervention	Path reset
Answer correction	Dream/correct	C (F)	Answer: B (C)	B (C)
Corrupt control	LLaDA/corrupt	D (C)	Answer: D (C)	A (F)
Evidence identity	iLLaDA/path	–	Correct/Corrupt/Absent	B (C) / D (F) / A (N)
Sparse repair	Dream/correct	–	Answer/Attr./Random/Recent: C (F)	B (C)