CONTROL ENTERS THROUGH THE MEMORY: SELF-CONDITIONING AS A CONTROL INTERFACE IN CONTINUOUS LANGUAGE FLOWS

Anonymous authors
Paper under double-blind review

ABSTRACT

Continuous language flows keep a real-valued state until the final decode, which makes the state a natural target for inference-time guidance. On the released sampler of Embedded Language Flows (ELF), however, a state push moves the state's decode toward the target while the emitted text lags. We show that the final decode reads exactly the model's clean prediction, its *belief*, which a state push does not directly optimize. The evidence points to stale self-conditioning: the denoiser's previous-prediction memory lags the pushed state, and the belief it conditions lags with it. Holding the self-conditioning scale fixed and editing only the memory's content restores control on three checkpoints, more than any of four control edits of the memory. The self-conditioning input is therefore a control interface, and *memory-port gradient writing* writes the reward gradient into the memory at every step. On three public ELF checkpoints, inside a pre-registered quality gate, the write alone beats state guidance at scale .25 on held-out keyword coverage at equal cost. Added to state guidance, it beats every state-guidance scale inside the gate at under twice the cost and also beats state guidance on two further reward families. Controllability in continuous language flows depends not only on what guidance optimizes but on where it enters.

1 INTRODUCTION

Text diffusion and flow models that operate in a continuous embedding space keep a real-valued state for the whole trajectory and produce discrete tokens once, at the end (Li et al., 2022; Dieleman et al., 2022; Gulrajani & Hashimoto, 2023; Hu et al., 2026). This makes the state look like a convenient target for training-free inference-time guidance, as in image diffusion, where guidance differentiates through the denoiser (Chung et al., 2023; Bansal et al., 2023) or resamples particles by reward (Singhal et al., 2025). On the released sampler of Embedded Language Flows (ELF) (Hu et al., 2026), however, a reward gradient applied to the state drives the state's own decode toward the target while the emitted text lags behind.

What quantity, then, must guidance move to change the final text? Three quantities live inside a self-conditioned sampler: the **state** z_k the sampler advances, the **belief** $\hat{x}_k$, which is the network's clean prediction of the finished sequence, and the **self-conditioning memory** m_k, the previous belief fed back as a second input (Chen et al., 2023; Strudel et al., 2022; Dieleman et al., 2022; Lovelace et al., 2023). They are three distinct targets for control, and they do not behave alike (Figure 1).

The state's own decode is not what the sampler emits. On the released implementation the last transition lands exactly on the belief, so the terminal reward is the indicator of a belief event at the last intermediate state, which we call *acceptance*. The state's decode, the *preview*, tracks that event only loosely. Without guidance the two agree on almost every row. A state push separates them (Figure 2), and that separation is the gap a practitioner sees.

The evidence points to stale self-conditioning as what limits the transfer from a state push to the belief: the previous-prediction memory the network reads lags the pushed state, and the belief it conditions lags with it. That pathway is known to misbehave with no guidance present (Yang et al., 2026; Liu et al., 2024; Zhang et al., 2026; Li & Xu, 2026); what is new here is what it costs a practitioner trying to steer the model. The obvious knob, the self-conditioning scale, is an input token of the denoiser, so

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.

turning it down changes the function and not only its input. We therefore hold the scale fixed and edit only what the memory *contains*, mixing it toward the memory-free belief of the current, post-push state. This restores guided coverage on all three checkpoints at much lower perplexity than removing the memory, and more than any direction-matched control.

The self-conditioning input is therefore a control interface in its own right. *Memory-port gradient writing* writes the decoder's reward gradient, taken at the memory input, into the memory at every sampler step. With no state push and at the wall-clock cost of state guidance, the write alone beats state guidance at scale .25 on held-out keyword coverage on all three checkpoints, inside a pre-registered quality gate. Added to state guidance, it beats every state-guidance scale inside the gate at under twice the cost and also beats state guidance on two further reward families. A random-direction write on the same schedule has no statistically resolved effect. Where guidance enters thus becomes a design choice; memory-port writing opens the memory port on a frozen network without retraining.

Contributions.

(1) **Interface.** On the released ELF sampler the terminal decode reads the belief exactly. We measure preview and acceptance along every trajectory on three checkpoints and show that a state push separates them (§4).
(2) **Mechanism.** At fixed self-conditioning scale, aligning the memory with the pushed trajectory restores control on three checkpoints and beats every direction-matched control, which on two of them show no statistically resolved recovery. This supports stale previous-prediction memory as the bottleneck (§5).
(3) **Method.** Memory-port gradient writing uses the self-conditioning input as a reward-guidance channel. Alone it beats state guidance at scale .25 at equal cost on three checkpoints; added to state guidance it beats every gated state-guidance scale and also wins on two further reward families, all inside a pre-registered quality gate (§6).

2 Related work

Guidance in continuous text diffusion. Guidance on the noisy state is inherited from image diffusion, first with a classifier trained on noisy inputs (Dhariwal & Nichol, 2021), then training-free with an off-the-shelf loss on the clean estimate (Chung et al., 2023; Bansal et al., 2023; Yu et al., 2023; He et al., 2024; Ye et al., 2024); sequential Monte Carlo resamples particles drawn from gradient-guided proposals (Wu et al., 2023), and search over the sampling noise, scored by a verifier, steers without any gradient (Ma et al., 2025). For a pixel-space model no decoder sits between state and output. In continuous text diffusion (Gong et al., 2023; Lin et al., 2023), Li et al. (2022) guide the state with a fluency regularizer, while Han et al. (2023) and Tae et al. (2025) guide the clean prediction, through the predicted logits and the predicted simplex. Kim et al. (2026) report classifier-guidance baselines that raised the classifier's score without raising constraint satisfaction. We hold the reward and decoder fixed, vary only what is decoded, and show which of the two events the sampler emits.

Self-conditioning. Self-conditioning feeds the model's previous clean prediction back as a second input (Chen et al., 2023); TESS (Karimi Mahabadi et al., 2024) uses it at every step, RIN (Jabri et al., 2023) feeds back latents, and Lovelace et al. (2023), Dieleman et al. (2022) and Strudel et al. (2022) ablate it for unguided sample quality and diversity only. Several works find the loop harmful without guidance. Yang et al. (2026) trace an entropy collapse at high step counts to it; Liu et al. (2024) find that a self-conditioned denoiser relies far more on the previous prediction than on the noised latent, the asymmetry our mechanism turns on, and retrain to mitigate it. On ELF, Zhang et al. (2026) trace repetition to one direction of the fed-back prediction and subtract it at every step, and Li & Xu (2026) damp the components of the self-conditioning input that the state already carries. These edits are fixed or label-free and target sample quality, not what the loop costs an external control signal. We measure that on a frozen network and write a target-dependent reward gradient into the self-conditioning input.

Control beyond continuous flows. For autoregressive language models, PPLM (Dathathri et al., 2020) shifts the key–value history, the only path by which future tokens depend on the past, along an attribute-model gradient taken through the network; other methods reweight the next-token

distribution with a discriminator (Yang & Klein, 2021; Krause et al., 2021), enforce lexical constraints inside beam search (Lu et al., 2021), or run Langevin dynamics on a continuous relaxation of the sequence (Qin et al., 2022; Kumar et al., 2022). Discrete diffusion language models (Austin et al., 2021; Lou et al., 2024; Sahoo et al., 2024; Nie et al., 2025) are guided through their transition rates or denoising distributions (Nisonoff et al., 2025; Schiff et al., 2025; Rojas et al., 2026). A self-conditioned flow offers another point of entry, the memory of its own previous prediction. Writing to it is the analogue of PPLM's history edit, with the decoder-only reward gradient in place of a gradient through the network.

ELF. Hu et al. (2026) model language with flow matching in a continuous embedding space, decoding to tokens only at the final step with a shared-weight network, and report better generation quality than discrete and continuous diffusion language models with fewer sampling steps. Self-conditioning is its conditioning signal for classifier-free guidance, trained in with the scale as an input; reward guidance is not studied. Dilip et al. (2026) guide ELF toward sample quality with probes on its hidden states, without a reward. On the same model, Du & Ma (2026) analyse the clean prediction's decoder margin along the trajectory on public ELF checkpoints, ablate self-conditioning with no external push, and report sentiment-directed repairs that edit the belief at the final state or at one intermediate step. We instead edit the self-conditioning input, at every step, and decode state and belief separately along the trajectory.

3 Background and Setup

3.1 Model and Sampler

ELF (Hu et al., 2026) embeds a token sequence with a frozen T5 encoder (Raffel et al., 2020) and trains one network on the linear interpolant $z_t = t\,x + (1-t)\,\sigma\epsilon$, $\epsilon \sim \mathcal{N}(0, I)$ (Lipman et al., 2023; Liu et al., 2023), regressing the clean embedding x rather than the velocity (Li & He, 2025); $\sigma = 2$ in the released configuration. Tokens come from running the same network once at $t = 1$ in a learned decode mode followed by an unembedding; we write D for this terminal-mode decoder, so a state z decodes to $\mathrm{text}(D(z))$. Self-conditioning feeds the previous clean prediction back as a second input m (Chen et al., 2023; Strudel et al., 2022). ELF distils classifier-free guidance (Ho & Salimans, 2022) on that input into a single pass with a scale token ω, giving one network $X_\omega(z, t, m)$ (Appendix A): $\omega = 1$ is plain self-conditioning and $\omega > 1$ extrapolates away from the memory-free prediction; the released default is $\omega = 3$. We use three public checkpoints, ELF-B, ELF-M and ELF-L (105M, 342M and 652M parameters).

Sampler. Transitions $k = 0, \ldots, N-1$ run on a logit-normal grid $0 = t_0 < \cdots < t_N = 1$ ($N = 32$ unless stated; observed $t_{N-1} \in [0.41, 0.70]$). With churn (Karras et al., 2022) $\gamma = 1.5$, $h_k = t_{k+1} - t_k$, $\lambda_k = 1 - \gamma h_k$, $\tilde{t}_k = \lambda_k t_k$, churn noise $\epsilon_k \sim \mathcal{N}(0, I)$ and $z_0 = \sigma\epsilon$,

$$\tilde{z}_k = \lambda_k z_k + (1-\lambda_k)\,\sigma\epsilon_k, \qquad \hat{x}_k = X_\omega(\tilde{z}_k, \tilde{t}_k, m_k),$$
$$z_{k+1} = \tilde{z}_k + \frac{t_{k+1} - \tilde{t}_k}{1 - \tilde{t}_k}\,(\hat{x}_k - \tilde{z}_k), \qquad m_{k+1} = \hat{x}_k, \tag{1}$$

except that the final transition is a churn-free Euler step from t_{N-1} to 1, which the released sampler forces whatever the interior method. Guided arms apply $z_{k+1} \leftarrow z_{k+1} + s_z h_k g_k$ for $k < N-1$, with $g_k = \nabla_z S / \mathrm{RMS}(\nabla_z S)$ for the reward proxy S of §3.2 and push scale s_z. All statements concern this released implementation.

3.2 Guidance Interfaces

Three quantities live inside (1). The **state** z_k is the interpolant the sampler advances, and is what conventional latent-side guidance pushes. The **belief** $\hat{x}_k$ is the network's clean prediction. The **self-conditioning memory** m_k is the previous belief, fed back as the network's second input; we call it *stale* when it lags the current, post-push state. Because ELF trains the denoising branch with a null memory half of the time (Appendix A), $B(z, t) := X_\omega(z, t, 0)$ is a configuration the network is trained on, and we call it the *memory-free belief* of a state.

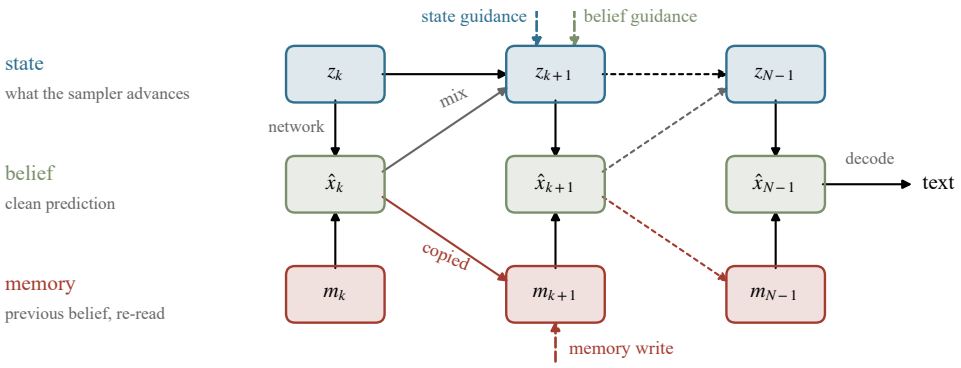

Figure 1: **The three quantities of the self-conditioned sampler and its three control interfaces.** State guidance and belief guidance push the state; the write of §6 edits the memory. The text is the decode of the last belief (Proposition 1).

With R the indicator that a text meets a target (§3.3), two events are recorded along every trajectory:

$$A_k := \{\, z : R(\text{text}(D(z))) = 1 \,\}, \qquad A_k^\star := \{\, z : R(\text{text}(D(X_\omega(z, t_k, m_k)))) = 1 \,\}. \quad (2)$$

We call $z_k \in A_k$ *preview*, the region a state-side proxy pushes toward, and $z_k \in A_k^\star$ *acceptance*. The acceptance probe reads the belief at the pair before re-noising, (z_k, t_k), which coincides with the sampler's own call only at the churn-free terminal step.

Guidance. The differentiable proxy S_{proxy} is a soft maximum over start positions of each target's mean token log-probability under D, averaged over targets (Appendix A), and the interfaces differ only in where it is evaluated. State guidance, the baseline this paper measures against, uses the *state proxy* $S(z) = S_{\text{proxy}}(D(z))$; belief guidance uses the *belief proxy* $S(z) = S_{\text{proxy}}(D(X_\omega(z, t_k, m_k)))$, with the gradient taken through the denoiser; memory-port writing evaluates $S_{\text{proxy}} \circ D$ at the memory input (§6).

3.3 Rewards, quality gate and protocol

Rewards. The main reward is keyword coverage. Each sample is drawn for a set of one or two target keywords or phrases; a target is met ($R = 1$) when it occurs as whole words in the text before the first EOS, and coverage is the rate at which targets are met. Two further families are never used for selection: *positional*, where a target counts only if it starts before token 128, and *exclusion*, where each target is replaced by a frequent word that must be absent from the whole text.

Quality gate. The method and control arms of §6 are held to a four-part gate registered in advance, each part scored against the unguided arm on the same rows as the coverage numbers: GPT-2-small (Radford et al., 2019) perplexity at most 2 and 4-gram repetition (Welleck et al., 2020) at most .01 above the unguided arm's, no resolved rise in duplicated, very short or extra sentences, and no resolved collapse of token entropy relative to OpenWebText (Gokaslan & Cohen, 2019) (Appendix A).

Protocol. Arms are paired on shared noise tapes, with 95 % intervals from a paired bootstrap (Efron & Tibshirani, 1993) that resamples whole blocks of rows. The method evaluation uses a separate held-out target split; the dev/test partition and the cell chosen on dev were fixed before any test row was generated. Appendix G gives counts, seeds and bootstrap blocks.

4 Preview and acceptance along the trajectory

4.1 The terminal identity

Proposition 1. *Consider the released implementation* (1) *and assume that (i) the final transition is the forced churn-free Euler step of §3; (ii) the floor of* 0.05 *that the released code puts on the velocity*

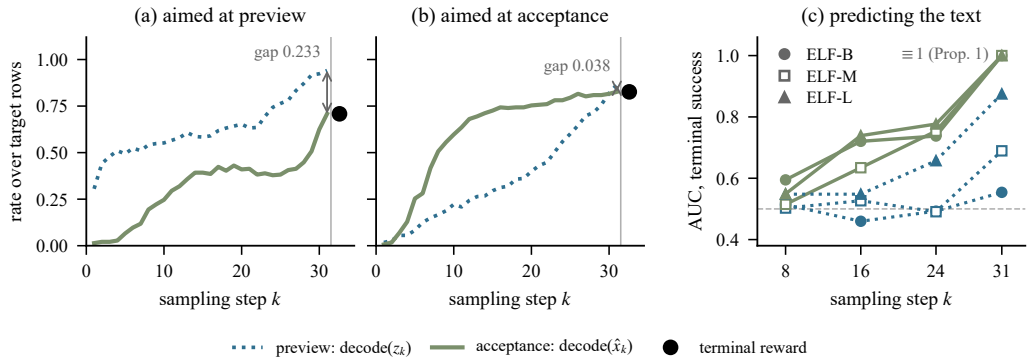

Figure 2: **A state push opens a preview–acceptance gap; aiming the push at acceptance nearly closes it, and acceptance predicts the text better.** (a, b) ELF-B at push scale .25 under state guidance (a) and belief guidance (b). (c) AUC of each event, read after step k, for terminal success under the push of (a), on all three checkpoints; dashed, 0.5.

denominator $1 - \tilde{t}_k$ *is inactive at the final step; and (iii) no guidance is applied after the final update. Then the trajectory lands on the belief evaluated at the pair* before *re-noising,*

$$z_N = X_\omega(z_{N-1}, t_{N-1}, m_{N-1}), \qquad \textit{so} \qquad \mathbf{1}\{\textit{terminal reward}\} = \mathbf{1}\{z_{N-1} \in A^\star_{N-1}\}. \tag{3}$$

Proof. By (i), $z_N = z_{N-1} + \frac{1-t_{N-1}}{1-t_{N-1}} \left(X_\omega(z_{N-1}, t_{N-1}, m_{N-1}) - z_{N-1}\right)$; by (ii) the ratio is exactly 1, so the z_{N-1} terms cancel; by (iii) nothing is added afterwards. $A^\star_{N-1}$ is defined at the same pair, which gives the second equality. □

Assumptions (i)–(iii) hold in the released code; (i) fails under the ELF paper's pseudocode, which also churns the final transition (Appendix A). The identity holds to float32 rounding on all three checkpoints (Appendix B.1), so a state-side proxy scores a decode the sampler never emits.

4.2 Preview and acceptance under the state push

Figure 2 shows the two probes along the trajectory on ELF-B at scale .25. At the last intermediate step the state push reaches preview 0.941 against acceptance 0.708, and the terminal reward equals the acceptance endpoint; aimed at acceptance, the same push leaves a gap of 0.038 and a terminal reward of 0.826. Under the push, 50 of 192 target rows end inside the preview region but outside the acceptance region; without guidance the two agree on 190 of 192, so the push creates the split. Figure 2(b) suggests an obvious remedy: aim the push at acceptance, which is belief guidance. It is not a reliable one. In the held-out evaluation of §6.2 it beats state guidance on ELF-B, but on ELF-M it covers less than state guidance and its perplexity fails the quality gate (Table 1). Skipping the final step and emitting the preview is no remedy either: it raises coverage over state guidance but fails the quality gate on perplexity on every checkpoint and covers less than the combined arm (Appendix E.3).

Figure 2(c) asks which event predicts the text. The preview stays near chance for most of the trajectory, informative only near the end, while acceptance predicts the terminal outcome from early on and more strongly at every step; the preview never overtakes it. A row that reaches acceptance by mid-trajectory almost always keeps it to the end. The event worth steering is therefore acceptance.

5 Self-conditioning and control transfer

Why does the belief lag a pushed state? The question matters because a push reaches the emitted text only through the belief (Proposition 1). We answer in three steps: weakening self-conditioning restores control; at a fixed self-conditioning scale, editing only what the memory contains restores it as well; and control edits of the memory in four other directions recover less.

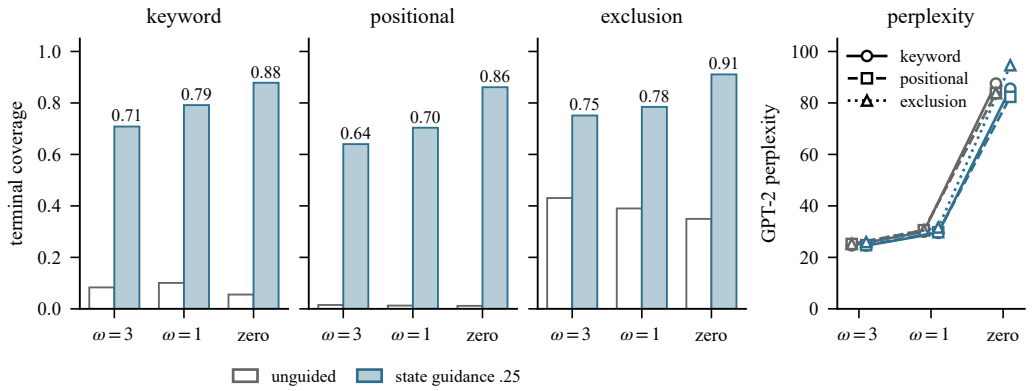

Figure 3: **The memory ladder on three reward families, and its fluency cost.** ELF-B, state guidance at .25. Left three panels: guided and unguided terminal coverage at $\omega = 3$ (default), $\omega = 1$ and zero memory. Right: GPT-2-small perplexity, which rises with the memory setting with or without guidance. Intervals in Table 2.

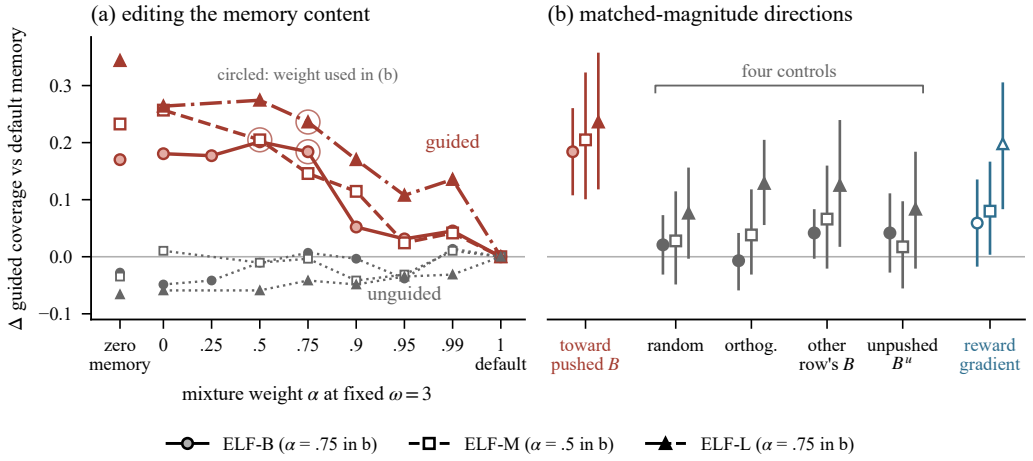

Figure 4: **At fixed ω, what matters is the memory's content.** State guidance at .25, keyword family, $\omega = 3$; change in guided coverage against default memory ($\alpha = 1$) on the same rows. (a) Against the mixture weight α of (4), guided and unguided; zero memory at the left, and circled the weights used in (b). (b) The mixture against the four direction controls and the state proxy's reward gradient, which is not a control; paired 95 % intervals.

5.1 The memory ladder

Weakening self-conditioning restores control. Under an identical state push, guided keyword coverage exceeds that of the default $\omega = 3$ on all three checkpoints, by $+0.061$ to $+0.083$ at $\omega = 1$ and by $+0.170$ to $+0.274$ with zero memory, every interval excluding zero, while unguided coverage does not rise (Table 2); the positional and exclusion families, run on ELF-B, show the same rise (Figure 3).

Removing the memory is not a usable remedy. At zero memory perplexity rises by a factor of 3.5 to 4.9, with or without guidance (Figure 3, right).

5.2 Memory content at fixed self-conditioning scale

Because ω is an input of the network, the $\omega = 1$ rung changes the *function* X_ω, not only the memory it is fed. To isolate the memory's content from the network it is fed to, we hold ω at its default 3 and

replace the memory before every transition $k \geq 1$ by

$$m_k^{(\alpha)} = m_k + (1-\alpha)\,(B_k - m_k), \qquad B_k = B(z_k, t_k), \tag{4}$$

with B the memory-free belief of the *current, post-push* state. Because B_k is evaluated after the push, any $\alpha < 1$ lets the memory track the pushed chain; $\alpha = 1$ is the default sampler.

Control returns at fixed ω: on all three checkpoints guided coverage is above the default at every $\alpha < 1$ while unguided coverage does not rise (Figure 4a), and on ELF-B $\alpha = 0$ recovers about as much as removing the memory. It does so without the fluency cost of removing the memory: on ELF-B the $\alpha = .75$ arm matches zero memory's coverage at about a third of its perplexity (Table 3). The shape differs by model, a step on ELF-B and a graded rise on ELF-M and ELF-L. Since ω, and with it the network, is unchanged, the gain comes from what the memory contains, not from weakening self-conditioning.

5.3 Direction-matched memory controls

The mixture moves the memory toward the belief of the pushed state. It could help because of that direction, or because any edit of the memory would. We separate the two by moving the memory by the same amount in other directions.

Before every transition, each arm moves the memory by the mixture's own per-row amount, in one of six directions: (a) toward the pushed belief, which is the mixture itself; (b) random; (c) random, orthogonal to the memory gap $B_k - m_k$; (d) another sequence's belief; (e) the belief B_k^u of a paired shadow chain on the same row and noise tape that receives no push and no memory edit, mixed in at the same weight α; and (f) the state proxy's reward gradient, which is *not* a null control. Directions (b)–(e) are the four controls (Appendix C.3).

At the mixture weight circled in Figure 4a, the mixture is separated from all four controls on all three checkpoints (Figure 4b). On ELF-B and ELF-M the controls themselves have no resolved effect; on ELF-L they recover part of the gain, but less than the mixture.

Control (e) is the decisive one. A fresh belief of the *unpushed* trajectory is what the default memory already approximates; fed explicitly, it recovers less than the mixture at the circled weight and at a second one on every checkpoint, so what helps is tracking the pushed trajectory, not freshness, although (e), matched by mixture weight rather than displacement, does not exclude every freshness account. Direction (f) is not inert, but the method writes a different gradient, the decoder's at the memory input. Among the directions tested, alignment with the currently pushed trajectory beats generic displacement and a fresh unpushed belief.

6 Memory-port gradient writing

Section 5 shows that, with the network fixed, what the memory contains decides how much of a push reaches the belief. The memory is therefore a control input in its own right, one that can be written directly, without pushing the state. The write uses this port but not the mixture tested there: it writes a reward gradient rather than the belief of the pushed state, and is not claimed to repair staleness. We ask two questions on held-out rows: whether the write alone controls the text at the cost of state guidance, and whether adding it to state guidance improves further. We also compare against belief guidance, an existing clean-prediction baseline (Han et al., 2023; Tae et al., 2025; Kim et al., 2026).

6.1 The write and its design choices

State guidance pushes the state along $\nabla_z S$. The memory is itself a clean prediction, in the space the decoder D reads, so the reward can be differentiated where the memory lives: we write, at every sampler step $k = 1, \ldots, N-1$, before the transition,

$$m_k \leftarrow m_k + \eta\,(1 - t_k)\,\frac{\hat{g}_k}{\mathrm{RMS}(\hat{g}_k)}, \qquad \hat{g}_k \;=\; \nabla_u S_{\mathrm{proxy}}(D(u))\big|_{u=m_k}, \tag{5}$$

while the base sampler applies its state push $s_z h_k g_k$ as usual, at $s_z = 0$ (the write alone) or $s_z = .25$ (the write on top of state guidance). We call the write **memory-port gradient writing**. The two settings share every other choice.

Table 1: **Held-out evaluation.** Test half, 384 rows per arm per checkpoint, shared rows and noise tapes; guidance at $s_z = .25$, the write alone at $s_z = 0$; method selected on keyword only. Δ: paired contrast against state guidance [95 % interval]; other contrasts in Tables 12 and 9. PPL and cost (wall-clock, relative to state guidance): keyword pass. †: outside the quality gate. Bold: best inside the gate. ELF-L, on which nothing was selected, lacks belief guidance (its denoiser backward exceeds 24 GB).

		keyword				other rewards	
	arm	coverage	Δ	PPL	cost	positional	exclusion
ELF-B	unguided	0.048	—	25.14	0.22×	0.003	0.441
	state guidance	0.595	—	25.95	1.00×	0.530	0.763
	belief guidance	0.750	—	25.92	1.49×	0.768	0.941†
	write alone	0.905	+0.310 [+0.245, +0.375]	26.08	0.98×	0.859	0.789
	combined	**0.951**	+0.355 [+0.284, +0.422]	26.17	1.76×	**0.914**	**0.905**
ELF-M	unguided	0.070	—	27.01	0.25×	0.004	0.339
	state guidance	0.617	—	27.27	1.00×	0.591	0.729
	belief guidance	0.391	—	29.33†	1.62×	0.327	0.810
	write alone	0.762	+0.145 [+0.089, +0.204]	27.46	0.99×	0.634	0.556
	combined	**0.883**	+0.266 [+0.217, +0.314]	28.05	1.74×	**0.846**	**0.835**
ELF-L	unguided	0.060	—	30.34	0.26×	0.007	0.344
	state guidance	0.609	—	31.28	1.00×	0.514	0.729
	write alone	0.783	+0.173 [+0.107, +0.241]	31.04	0.99×	0.686	0.607
	combined	**0.882**	+0.272 [+0.211, +0.333]	31.76	1.73×	**0.854**	**0.810**

Three choices complete the method (Appendix D). *Gradient:* we write the decoder's reward gradient at the memory, which needs no denoiser pass, not the exact pullback through the denoiser; at the tested setting the pullback is no better held out, and written alone it covers less, fails the quality gate and costs 1.5–1.6× state guidance against 1.0× for the write alone. *Schedule:* we write at every step, tapered by $1 - t_k$; writing at every third step keeps most of the gain at about 1.3× the cost of state guidance, against 1.7× at every step. *Magnitude:* $\eta = .1$, selected on the dev half, with nothing tuned per checkpoint. The write alone costs what state guidance costs, since its decoder backward replaces the push's; in the combined arm it is additional. Where the proxy saturates and its gradient vanishes, the write is skipped (the saturation rule), for a quarter to a third of the combined arm's writes; the gradient is still computed, so the reported costs include the skipped writes (Appendix E.2).

6.2 Held-out evaluation

Table 1 and Figure 5(a–c) answer both questions on identical rows: the write alone improves on state guidance inside the gate on all three checkpoints, and the combined arm improves further (Table 12). Against belief guidance the combined arm wins inside the gate on ELF-B; on ELF-M belief guidance leaves the gate. On all three checkpoints every guided arm, state guidance included, exceeds best-of-N sampling that selects with the reward itself, at its own matched wall-clock budget (Table 7; operation-count sensitivity in Appendix E.1).

Figure 5(d–f) shows where the gain comes from. State guidance moves the preview ahead of acceptance; the write alone raises acceptance while the preview stays low, so it moves the event the sampler emits rather than the state's own decode; the combined arm raises both and ends highest. Appendix F shows generated text.

The method also transfers to the positional and exclusion rewards, on which it was not selected. With the write and scale unchanged, the combined arm beats both state guidance and the write alone on both families and all three checkpoints, all three arms inside the gate (Tables 1 and 9). The write alone beats state guidance on positional ELF-B and ELF-L but not in the other four cases, where the write's gain comes as an addition to state guidance.

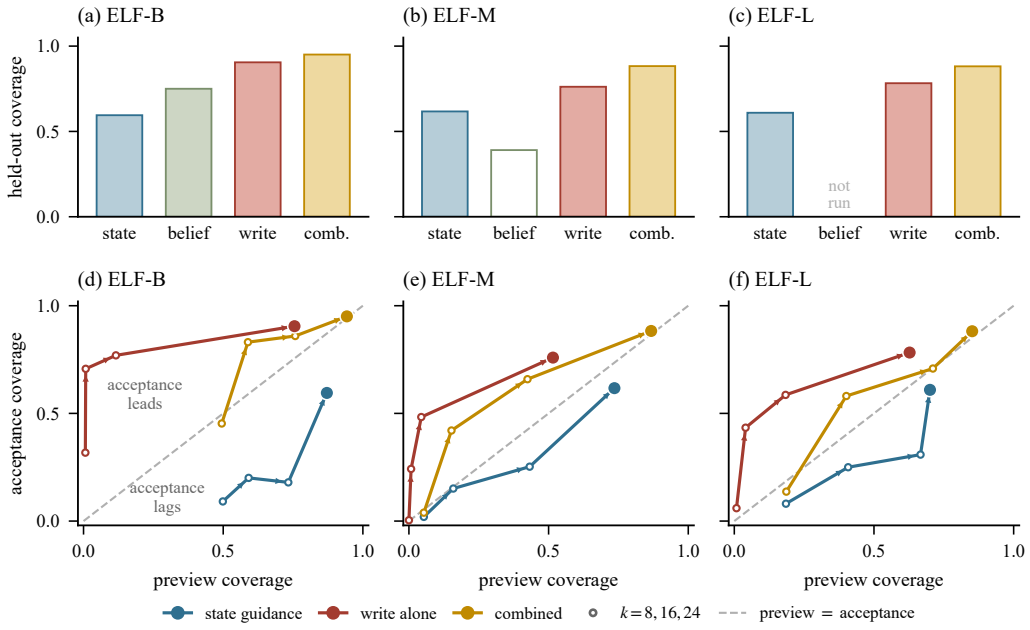

Figure 5: **What the write buys and where it acts.** Held-out rows of Table 1. (a–c) Keyword coverage of state and belief guidance at .25, the write alone and the combined arm; hollow: outside the quality gate. (d–f) Each path follows one arm through steps $k = 8, 16, 24, 31$: keyword coverage of the decode of the state (preview, horizontal) against that of the belief (acceptance, vertical).

6.3 Controls

The gains of §6.2 could have simpler causes: the write's schedule, any edit of the memory, or an under-tuned baseline. Three controls on the test rows address each in turn (Appendix E.5).

Schedule: the write's $(1 - t_k)$ taper, applied to state or belief guidance, lowers their coverage on every checkpoint tested, so the taper alone does not reproduce the gain. *Direction:* a random direction written on top of state guidance, with the write's schedule, magnitude and saturation rule and never skipping a step, has no resolved effect on any checkpoint, and the combined arm beats it on all three; which property of the gradient helps is left open. *State scale:* against every state scale inside the gate on the test rows, not only the .25 of Table 1, the combined arm wins on every checkpoint; the write alone wins on ELF-B and is not separated from .35 or .50 on ELF-M and ELF-L.

7 Conclusion

On the released ELF sampler the emitted text is the decode of the belief, so a state push steers the text only as far as the belief follows it. The evidence points to stale self-conditioning as what holds the belief back: at a fixed self-conditioning scale, aligning the memory with the pushed trajectory restores control, and direction-matched edits recover less, within the limit that control (e) is matched by mixture weight rather than displacement. The self-conditioning input therefore doubles as a control port. Memory-port writing alone beats state guidance at scale .25 on held-out keyword coverage at equal cost; added to state guidance, it beats every gated state-guidance scale on keyword coverage, and state guidance on all three reward families, at under twice the cost. Where guidance enters a self-conditioned sampler is thus a design choice.

These conclusions have a limited scope. All evidence comes from one model family: the memory ladder, the mixture and its separation from matched controls hold on three checkpoints. The rewards are lexical, and which property of the gradient makes the write useful remains open.

AI USE STATEMENT

In this work, we used generative AI tools in implementing the experiments and analysing the results, as well as for literature search, code, figures and polishing. No synthetic data sets were generated. All AI-assisted work was reviewed by the authors, including checks of cited statements against the papers and of reported numbers against the experimental records. We take responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

Code for the sampler, the memory interventions, memory-port gradient writing, the quality gate and the paired bootstrap, together with the run configurations of the held-out evaluation and the ELF-L memory experiments, is available at https://anonymous.4open.science/r/SelfConditioning-939E. All experiments use the three public ELF checkpoints with fixed seeds. Appendix A specifies the sampler, rewards, proxy and quality gate, Appendix D.1 the dev/test split and the selection rule, and Appendix G the checkpoints, seeds, sample counts and timings.

REFERENCES

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems*, pp. 17981–17993, 2021.

Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 843–852, 2023.

Ting Chen, Ruixiang Zhang, and Geoffrey Hinton. Analog bits: Generating discrete data using diffusion models with self-conditioning. In *International Conference on Learning Representations*, 2023.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. Plug and play language models: A simple approach to controlled text generation. In *International Conference on Learning Representations*, 2020.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, pp. 8780–8794, 2021.

Sander Dieleman, Laurent Sartran, Arman Roshannai, Nikolay Savinov, Yaroslav Ganin, Pierre H. Richemond, Arnaud Doucet, Robin Strudel, Chris Dyer, Conor Durkan, Curtis Hawthorne, Rémi Leblond, Will Grathwohl, and Jonas Adler. Continuous diffusion for categorical data. *arXiv preprint arXiv:2211.15089*, 2022.

Rohit Dilip, Tianrong Chen, Yuyang Wang, David Van Valen, Joshua Susskind, and Miguel Angel Bautista. How to guide your language flow. *arXiv preprint arXiv:2609.19356*, 2026.

Zhicheng Du and Lan Ma. Continuous language diffusion as a decoder-interface problem. *arXiv preprint arXiv:2606.08810*, 2026.

Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, New York, 1993.

Aaron Gokaslan and Vanya Cohen. OpenWebText corpus. https://doi.org/10.5281/zenodo.3834942, 2019.

Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu, and Lingpeng Kong. DiffuSeq: Sequence to sequence text generation with diffusion models. In *International Conference on Learning Representations*, 2023.

Ishaan Gulrajani and Tatsunori B. Hashimoto. Likelihood-based diffusion language models. In *Advances in Neural Information Processing Systems*, pp. 16693–16715, 2023.

Xiaochuang Han, Sachin Kumar, and Yulia Tsvetkov. SSD-LM: Semi-autoregressive simplex-based diffusion language model for text generation and modular control. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 11575–11596, 2023.

Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, J. Zico Kolter, Ruslan Salakhutdinov, and Stefano Ermon. Manifold preserving guided diffusion. In *International Conference on Learning Representations*, 2024.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Keya Hu, Linlu Qiu, Yiyang Lu, Hanhong Zhao, Tianhong Li, Yoon Kim, Jacob Andreas, and Kaiming He. ELF: Embedded language flows. *arXiv preprint arXiv:2605.10938*, 2026.

Allan Jabri, David J. Fleet, and Ting Chen. Scalable adaptive computation for iterative generation. In *International Conference on Machine Learning*, pp. 14569–14589, 2023.

Rabeeh Karimi Mahabadi, Hamish Ivison, Jaesung Tae, James Henderson, Iz Beltagy, Matthew E. Peters, and Arman Cohan. TESS: Text-to-text self-conditioned simplex diffusion. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2347–2361, 2024.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, pp. 26565–26577, 2022.

Jinwoo Kim, Taylor Berg-Kirkpatrick, and Loris D'Antoni. Continuous diffusion models can obey formal syntax. *arXiv preprint arXiv:2602.12468*, 2026.

Ben Krause, Akhilesh Deepak Gotmare, Bryan McCann, Nitish Shirish Keskar, Shafiq Joty, Richard Socher, and Nazneen Fatema Rajani. GeDi: Generative discriminator guided sequence generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pp. 4929–4952, 2021.

Taku Kudo and John Richardson. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 66–71, 2018.

Sachin Kumar, Biswajit Paria, and Yulia Tsvetkov. Gradient-based constrained sampling from language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 2251–2277, 2022.

Bocheng Li and Linli Xu. Improving few-step language flows with untied self-conditioning. *arXiv preprint arXiv:2608.22244*, 2026.

Tianhong Li and Kaiming He. Back to basics: Let denoising generative models denoise. *arXiv preprint arXiv:2511.13720*, 2025.

Xiang Lisa Li, John Thickstun, Ishaan Gulrajani, Percy Liang, and Tatsunori B. Hashimoto. Diffusion-LM improves controllable text generation. In *Advances in Neural Information Processing Systems*, pp. 4328–4343, 2022.

Zhenghao Lin, Yeyun Gong, Yelong Shen, Tong Wu, Zhihao Fan, Chen Lin, Nan Duan, and Weizhu Chen. Text generation with diffusion language models: A pre-training approach with continuous paragraph denoise. In *International Conference on Machine Learning*, pp. 21051–21064, 2023.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023.

Yuxuan Liu, Tianchi Yang, Shaohan Huang, Zihan Zhang, Haizhen Huang, Furu Wei, Weiwei Deng, Feng Sun, and Qi Zhang. Text diffusion with reinforced conditioning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 14069–14077, 2024.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution. In *International Conference on Machine Learning*, pp. 32819–32848, 2024.

Justin Lovelace, Varsha Kishore, Chao Wan, Eliot Shekhtman, and Kilian Q. Weinberger. Latent diffusion for language generation. In *Advances in Neural Information Processing Systems*, pp. 56998–57025, 2023.

Ximing Lu, Peter West, Rowan Zellers, Ronan Le Bras, Chandra Bhagavatula, and Yejin Choi. NeuroLogic decoding: (un)supervised neural text generation with predicate logic constraints. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4288–4299, 2021.

Nanye Ma, Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, and Saining Xie. Scaling inference time compute for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2523–2534, 2025.

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. In *Advances in Neural Information Processing Systems*, pp. 56354–56392, 2025.

Hunter Nisonoff, Junhao Xiong, Stephan Allenspach, and Jennifer Listgarten. Unlocking guidance for discrete state-space diffusion and flow models. In *International Conference on Learning Representations*, 2025.

Lianhui Qin, Sean Welleck, Daniel Khashabi, and Yejin Choi. COLD decoding: Energy-based constrained text generation with Langevin dynamics. In *Advances in Neural Information Processing Systems*, pp. 9538–9551, 2022.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. Technical report, OpenAI, 2019.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020.

Kevin Rojas, Ye He, Chieh-Hsin Lai, Yuhta Takida, Yuki Mitsufuji, and Molei Tao. Improving classifier-free guidance in masked diffusion: Low-dim theoretical insights with high-dim impact. In *International Conference on Learning Representations*, 2026.

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. In *Advances in Neural Information Processing Systems*, pp. 130136–130184, 2024.

Yair Schiff, Subham Sekhar Sahoo, Hao Phung, Guanghan Wang, Sam Boshar, Hugo Dalla-Torre, Bernardo P. de Almeida, Alexander Rush, Thomas Pierrot, and Volodymyr Kuleshov. Simple guidance mechanisms for discrete diffusion models. In *International Conference on Learning Representations*, 2025.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. In *International Conference on Machine Learning*, pp. 55810–55827, 2025.

Robin Strudel, Corentin Tallec, Florent Altché, Yilun Du, Yaroslav Ganin, Arthur Mensch, Will Grathwohl, Nikolay Savinov, Sander Dieleman, Laurent Sifre, and Rémi Leblond. Self-conditioned embedding diffusion for text generation. *arXiv preprint arXiv:2211.04236*, 2022.

Jaesung Tae, Hamish Ivison, Sachin Kumar, and Arman Cohan. TESS 2: A large-scale generalist diffusion language model. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 21171–21188, 2025.

Sean Welleck, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. Neural text generation with unlikelihood training. In *International Conference on Learning Representations*, 2020.

Luhuan Wu, Brian L. Trippe, Christian A. Naesseth, David M. Blei, and John P. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *Advances in Neural Information Processing Systems*, pp. 31372–31403, 2023.

Kevin Yang and Dan Klein. FUDGE: Controlled text generation with future discriminators. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 3511–3535, 2021.

Zhihan Yang, Wei Guo, Shuibai Zhang, Subham Sekhar Sahoo, Yongxin Chen, Arash Vahdat, Morteza Mardani, and John Thickstun. Continuous diffusion scales competitively with discrete diffusion for language. *arXiv preprint arXiv:2605.18530*, 2026.

Haotian Ye, Haowei Lin, Jiaqi Han, Minkai Xu, Sheng Liu, Yitao Liang, Jianzhu Ma, James Zou, and Stefano Ermon. TFG: Unified training-free guidance for diffusion models. In *Advances in Neural Information Processing Systems*, pp. 22370–22417, 2024.

Jiwen Yu, Yinhuai Wang, Chen Zhao, Bernard Ghanem, and Jian Zhang. FreeDoM: Training-free energy-guided conditional diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 23174–23184, 2023.

Shuai Zhang, Zijie Chen, Hongliang He, Lun Du, and Zhenzhong Lan. Low perplexity is repetition: A one-dimensional self-conditioning attractor in continuous diffusion LMs. *arXiv preprint arXiv:2607.00588*, 2026.

A Model, sampler and reward details

Self-conditioning guidance. Training regresses the velocity implied by the network's clean prediction, $v_\theta(z_t \mid t, m, \omega) = (X_\omega(z_t, t, m) - z_t)/(1-t)$, onto

$$v_{\text{target}} = x - \sigma\epsilon + \Big(1 - \tfrac{1}{\omega}\Big)\big(v_\theta(z_t \mid t, m, \omega) - v_\theta(z_t \mid t, 0, \omega)\big), \tag{6}$$

whose fixed point is the guided velocity $\omega\, v(z_t \mid m) + (1-\omega)\, v(z_t \mid 0)$, with $v(z_t \mid \cdot)$ the velocity that unguided training would learn for that memory input. The encoder is a frozen T5-small mapping a sequence of length L to $x \in \mathbb{R}^{L\times 512}$; the memory is concatenated with the state, and in the denoising branch it is the prediction of a first pass with probability one half and the null memory 0 otherwise, while the decoding branch always receives 0; the scale token ranges over $\omega \in [0.5, 5]$, so self-conditioning strength is an inference-time knob that needs no retraining. Self-conditioning throughout means this previous-prediction form; a current-state bootstrap has no staleness to remove.

Released sampler. Equation (1) is the *released implementation* of the churn sampler, in which the velocity is evaluated at the re-noised pair $(\tilde{z}_k, \tilde{t}_k)$ and the step is taken from $\tilde{z}_k$. The pseudocode in the ELF paper's sampler appendix instead takes the step from the original (z_k, t_k), while still evaluating the velocity at the re-noised pair, and carries no endpoint exception, so under it the final transition is churned like any other. The two maps coincide only at $\gamma = 0$.

The acceptance probe. At intermediate steps the probe of (2) is a diagnostic read at the state the trajectory actually occupies, (z_k, t_k), not a replay of the sampler's internal call at $(\tilde{z}_k, \tilde{t}_k)$; the two coincide at the churn-free terminal step, which is what (3) uses.

Reward and proxies. The hard reward of §3.3 matches keywords or phrases after NFKC normalisation and casefolding, at ASCII word boundaries, over three target sets in §4–§5 and eight per dev or test half in §6; the targets enter the proxy as SentencePiece (Kudo & Richardson, 2018) token sequences. Let target i of a set of $n \in \{1, 2\}$ be the token sequence $v^i_1, \ldots, v^i_{L_i}$ of its lower-cased form with a leading word-start marker, and $\log p_j(v)$ the decoder's log-probability of token v at position j of the length-T sequence. The proxy is the mean over targets of a soft maximum over start positions s of the target's mean token log-probability,

$$S_{\text{proxy}} = \frac{1}{n}\sum_{i=1}^{n} \tau \log \frac{1}{W_i} \sum_{s=0}^{W_i - 1} \exp\Big(\frac{1}{\tau L_i}\sum_{j=1}^{L_i} \log p_{s+j-1}(v^i_j)\Big), \tag{7}$$

with $W_i = T - L_i + 1$ and $\tau = .25$. The positional family restricts the starts to $s < 128$, $W_i = \min(128, T - L_i + 1)$; the exclusion family uses $-S_{\text{proxy}}$ over all starts.

Quality gate. Each criterion compares an arm with the unguided arm of the same pass on the same rows; an arm is inside the gate if no criterion fires. Intervals are paired block-bootstrap 95 % intervals over the sampler's four-sample batches. *Perplexity:* mean per-row GPT-2-small perplexity above unguided $+2$, or below unguided -2 while another criterion fires. *Repetition:* mean 4-gram repetition, the fraction of repeated 4-grams among a sample's whitespace-delimited words (seq-rep-4 of Welleck et al., 2020), above unguided $+.01$. *Structure:* a rise of at least .02 in the fraction of duplicated sentences or of sentences shorter than five words, or of at least 20 % in the number of sentences, with the interval excluding zero. *Entropy:* per-sample unigram entropy, against OpenWebText documents truncated to the same length, either below 4 nats, or with a gap below OpenWebText at least $1.5\times$ the unguided gap and an interval excluding zero; this criterion fires only if it fires under both the GPT-2 and the T5 tokenizer. Where a result uses the perplexity criterion alone (unguided mean $+2$) — the dev grid of (5), the ELF-B and ELF-M write-count frontier and the controls session's dev selection rule — the caption or text says so.

B Preview, acceptance and the displacement recursion

B.1 Terminal identity and acceptance

The terminal state equals the belief up to float32 rounding on every recorded row: the maximum absolute difference is 7.2×10^{-7} on ELF-B (the rows of Figure 2), 7.2×10^{-7} on ELF-M (3,456

rows) and 9.5×10^{-7} on ELF-L (all 5,760 rows of the memory experiments, none bitwise equal). Acceptance is also nearly absorbing: $\Pr[\text{terminal} \mid \text{accept}_k = 1] \geq 0.9375$ at every probed step on ELF-B and ELF-L, and from $k = 16$ on ELF-M. In Figure 2(c) the acceptance AUC at $k = 31$ is 1 by Proposition 1; the ELF-M arm replicates the ELF-B targets on new noise tapes, and the ELF-L arm reruns it on the same rows and tapes. The preview–acceptance split is not an artefact of a strong push: at the weaker state scale .10, 103 of 192 target rows end inside the preview region and outside the acceptance region, against $50/192$ at .25.

B.2 The displacement recursion along the chain

Lemma 1 (Displacement recursion)**.** *Run (1) twice on the same noise tape (ϵ_k) and the same grid; on the guided chain the push $s_z h_k g_k$ is added to z_{k+1} after transition k, for $k \leq N-2$. Let z_k denote the state entering transition k (on the guided chain, after the push of transition $k-1$), $u_k = z_k^g - z_k^u$ its displacement, u_{k+1}^- the displacement immediately after transition k and before that transition's push, so that $u_{k+1} = u_{k+1}^- + s_z h_k g_k$, and $\hat{x}_k^g, \hat{x}_k^u$ the two beliefs. Assume $\gamma h_k < 1$ for $k \leq N-2$. Then*

$$u_{k+1}^- = (1 - b_k)\,\lambda_k\,u_k + b_k\left(\hat{x}_k^g - \hat{x}_k^u\right), \qquad b_k = \frac{t_{k+1} - \tilde{t}_k}{1 - \tilde{t}_k}, \tag{8}$$

with $0 < (1-b_k)\lambda_k < 1$ for $k \leq N-2$, and under the assumptions of Proposition 1, $u_N = u_N^- = \hat{x}_{N-1}^g - \hat{x}_{N-1}^u$.

Proof. The re-noising step of (1) is affine in z_k with the same ϵ_k on both chains, so $\tilde{z}_k^g - \tilde{z}_k^u = \lambda_k u_k$; the update step is affine in $(\tilde{z}_k, \hat{x}_k)$ with coefficients $1 - b_k$ and b_k, and the push is added only after it; subtracting the two copies gives (8). For $k \leq N-2$, $\gamma h_k < 1$ gives $\lambda_k \in (0,1)$, and $\tilde{t}_k < t_{k+1} < 1$ gives $b_k \in (0,1)$. At the final transition, Proposition 1(i)–(iii) give a pair that is not re-noised, $b_{N-1} = 1$ and no push, so the first term vanishes, $u_N = u_N^-$, and the last claim is (3) applied to both chains. □

Corollary 1. *If the belief is locked, $\hat{x}_k^g = \hat{x}_k^u$ for every k, then $u_N = 0$: the final transition erases the direct residual of every push, whatever preterminal displacement the pushes maintained.*

The first term of (8) decays geometrically, and the second is the only route by which a push reaches the next state other than through its own decay. The final step thus discards whatever state displacement the pushes left behind: any effect of the pushes on z_N is mediated entirely by the belief's response at the displaced state and its evolved memory, and a locked belief (Corollary 1) buys preview and nothing terminal.

Numerical check. On 192 fresh rows per setting on ELF-B and ELF-M, with state guidance at .25 and an unguided chain on the same noise tape under each memory rule ($\omega = 3$, $\omega = 1$, the $\alpha = .75$ mixture of §5.2, zero memory), the residual of (8) is at most 2.0×10^{-5} of $\|u_{k+1}^-\|$ at $k \leq 30$ and at most 1.6×10^{-7} at the terminal transition, and $\gamma h_k < 0.362$ on every schedule drawn.

C Memory experiments

The ELF-L rows of Tables 2, 3 and 4 and of Figure 4 come from the ELF-M arms rerun unchanged (arm definitions, seeds, rows, noise tapes and bootstrap), with nothing chosen on ELF-L: its circled weight, $\alpha = .75$, is set by a rule fixed on ELF-B and ELF-M. Both ladder rungs, the return of control at fixed ω and the mixture's separation from every control replicate on ELF-L and are stated on three checkpoints; on ELF-L some controls recover part of the gain and some unguided contrasts decline (Appendices C.1–C.3), so the controls' lack of a resolved effect is stated on ELF-B and ELF-M only.

C.1 The memory ladder

Weakening self-conditioning raises guided coverage monotonically on all three checkpoints (Table 2), and the count of ELF-B target rows that end inside the preview region but outside the acceptance region falls $50 \to 35 \to 6$ down the ladder. On the first three ELF-M seeds alone the intermediate rung has the same sign with an interval that includes zero ($+0.069$ $[-0.007, +0.153]$); the six further

Table 2: The self-conditioning ladder under state guidance at scale .25, keyword family, $N = 32$; ω is the distilled self-conditioning guidance scale of (6), and zero memory feeds $m = 0$ at $\omega = 3$. Coverage is terminal target coverage; PPL is mean GPT-2 perplexity (unguided / guided), scored with GPT-2-small on all three checkpoints and GPT-2-Large on ELF-B and ELF-M. Contrasts are against the default $\omega = 3$, paired block bootstrap with 95 % intervals; ELF-B has 144 samples per arm, ELF-M and ELF-L 432 over nine seeds (Appendix G); the ELF-M GPT-2-Large row is scored on three seeds (144 texts). Preview and acceptance are measured at the last intermediate step on the ELF-B arms. The positional and exclusion blocks are ELF-B (Appendix C.4; nine seeds, 432 samples per arm, stratified cluster bootstrap over seeds); contrast: the zero-memory contrast in the decision form of Appendix C.4, raw or guided minus unguided (g−u).

model	quantity	$\omega = 3$ (default)	$\omega = 1$	zero memory
ELF-B	unguided coverage	0.083	0.101	0.056
	guided coverage	0.708	0.792	0.878
	guided contrast	—	$+0.083$ $[+0.021, +0.142]$	$+0.170$ $[+0.076, +0.257]$
	unguided contrast	—	$+0.017$ $[-0.024, +0.063]$	-0.028 $[-0.080, +0.024]$
	PPL (GPT-2-small)	24.6 / 24.6	30.0 / 29.5	87.5 / 85.6
	PPL (GPT-2-Large)	25.2 / 25.1	31.0 / 30.3	83.7 / 81.5
	preview / acceptance	0.941 / 0.708	0.944 / 0.792	0.885 / 0.878
ELF-M	unguided coverage	0.098	0.101	0.080
	guided coverage	0.701	0.763	0.907
	guided contrast	—	$+0.061$ $[+0.015, +0.109]$	$+0.206$ $[+0.148, +0.269]$
	unguided contrast	—	$+0.002$ $[-0.027, +0.031]$	-0.019 $[-0.051, +0.014]$
	PPL (GPT-2-small)	27.4 / 27.8	36.8 / 38.1	98.9 / 107.2
	PPL (GPT-2-Large)	28.3 / 28.7	38.1 / 39.4	102.1 / 109.6
ELF-L	unguided coverage	0.110	0.102	0.069
	guided coverage	0.603	0.685	0.877
	guided contrast	—	$+0.082$ $[+0.027, +0.140]$	$+0.274$ $[+0.216, +0.333]$
	unguided contrast	—	-0.008 $[-0.030, +0.013]$	-0.041 $[-0.072, -0.012]$
	PPL (GPT-2-small)	30.5 / 31.8	39.1 / 40.5	150.0 / 154.7
positional	unguided coverage	0.015	0.013	0.012
	guided coverage	0.640	0.704	0.861
	contrast (raw)	—	—	$+0.221$ $[+0.169, +0.271]$
	PPL (GPT-2-small)	25.2 / 24.7	30.6 / 29.8	83.8 / 82.4
exclusion	unguided coverage	0.431	0.390	0.350
	guided coverage	0.751	0.785	0.911
	contrast (g−u)	—	—	$+0.241$ $[+0.190, +0.289]$
	PPL (GPT-2-small)	25.2 / 25.9	30.6 / 31.6	83.8 / 94.6

seeds were fixed before their rows existed and are pooled. On ELF-L the unguided zero-rung contrast is a resolved decline, not a rise.

Chain versus final step. On ELF-B, ablating the memory only at the final transition recovers little: $+0.007$ $[-0.007, +0.024]$ for $\omega = 1$ at the last step and $+0.021$ $[+0.003, +0.045]$ for a memory replaced by the memory-free belief of the pushed state, 1 and 4 of the 50 target rows lost to the gap. The memory map is applied $N - 1$ times, and the gain builds up along the chain, not at the final read.

Fluency. Guided and unguided perplexities differ far less than perplexity differs across memory settings; the memory setting alone moves perplexity, by a factor of about 3.5 on ELF-B, 3.6–3.9 on ELF-M and 4.9 on ELF-L at the zero rung, and GPT-2-Large gives the same ordering on ELF-B and ELF-M. Zero-memory texts contain non-words and code-switching.

C.2 Memory mixture

At fixed $\omega = 3$ the guided series is not monotone cell by cell (ELF-B 0.889, 0.885, 0.910, 0.892 at $\alpha = 0$, .25, .5, .75), although every $\alpha < 1$ is above the default and the low rungs are well above the high ones (Table 3). The same pattern appears at $\omega = 1$ with about half the gain ($+0.090$ against

$+0.184$ at $\alpha = .75$). Down the mixture rungs perplexity climbs gradually rather than jumping, and at $\alpha = .75$ guided perplexity is 3.0 to 4.3 times lower than at zero memory on all three checkpoints. On ELF-L the $\alpha = 0$ contrast is $+0.264$ $[+0.156, +0.375]$ ($+0.323$ $[+0.198, +0.458]$ guided minus unguided), and its unguided counterpart is a resolved decline, -0.059 $[-0.108, -0.014]$, not a rise.

Table 3: Memory mixture at fixed ω (keyword family, state guidance .25; 144 rows per cell; $\alpha = 1$ and zero memory are the ladder arms; GPT-2 perplexity, mean). ELF-M and ELF-L rows: $\omega = 3$ only, on three seeds; Table 2 pools nine seeds for their ladder arms.

model, ω	arm	$\alpha = 0$	.25	.5	.75	.9	.95	.99	1	zero memory
ELF-B, 3	guided cov.	0.889	0.885	0.910	0.892	0.760	0.740	0.753	0.708	0.878
ELF-B, 3	unguided cov.	0.035	0.042	0.073	0.090	0.080	0.045	0.097	0.083	0.056
ELF-B, 3	PPL guided	44.6	39.2	34.1	28.1	25.9	25.0	24.5	24.6	85.6
ELF-B, 3	PPL unguided	45.0	39.5	33.6	28.1	25.6	25.3	24.5	24.6	87.5
ELF-B, 1	guided cov.	0.885	0.882	0.896	0.882	—	—	—	0.792	—
ELF-B, 1	unguided cov.	0.045	0.024	0.059	0.069	—	—	—	0.101	—
ELF-B, 1	PPL guided	50.5	46.7	40.9	35.4	—	—	—	29.5	—
ELF-B, 1	PPL unguided	50.3	46.2	41.2	35.3	—	—	—	30.0	—
ELF-M, 3	guided cov.	0.931	—	0.878	0.819	0.788	0.698	0.715	0.674	0.906
ELF-M, 3	unguided cov.	0.108	—	0.087	0.094	0.056	0.066	0.108	0.097	0.062
ELF-M, 3	PPL guided	51.5	—	38.8	32.5	29.4	28.4	28.1	27.9	110.2
ELF-M, 3	PPL unguided	49.3	—	37.1	31.4	28.9	28.0	27.6	27.4	102.5
ELF-L, 3	guided cov.	0.802	—	0.812	0.774	0.708	0.646	0.674	0.538	0.882
ELF-L, 3	unguided cov.	0.080	—	0.080	0.097	0.090	0.104	0.108	0.139	0.073
ELF-L, 3	PPL guided	62.0	—	44.0	37.0	33.1	32.2	31.2	31.2	160.1
ELF-L, 3	PPL unguided	59.6	—	42.5	35.1	31.8	30.8	30.2	29.9	156.2

C.3 Direction-matched controls

The six directions are those of §5.3. Direction (a) is the mixture itself in exact arithmetic and reproduces the α arms token for token on $144/144$ rows per arm and model; the random directions (b) and (c) are seeded per row and step. Directions (b), (c), (d) and (f) are applied at the mixture's own per-row magnitude $\delta_k = (1 - \alpha)\,\mathrm{RMS}(B_k - m_k)$, about 0.05 state-RMS units at $\alpha = .75$ on ELF-B. The matching is exact per row within an arm but not identical across arms, since each arm recomputes $\mathrm{RMS}(B_k - m_k)$ on a trajectory its own intervention has already moved; the controls end up displacing slightly more than the mixture. Direction (e) is α-matched rather than δ-matched: its magnitude is recomputed from the shadow trajectory's own gap $\mathrm{RMS}(B^u_k - m_k)$ at the same α. Controls (b)–(d) keep perplexity close to their unguided baselines while the mixture's is higher (Table 4), so the mixture's extra perplexity goes with movement toward the current belief rather than with movement alone.

At the second weight ($\alpha = .5$ on ELF-B and ELF-L, .75 on ELF-M) the mixture exceeds control (e) by $+0.236$ $[+0.146, +0.323]$, $+0.115$ $[+0.035, +0.198]$, $+0.170$ $[+0.090, +0.250]$. On ELF-L at $\alpha = .5$ all four controls exclude zero ($+0.101$ to $+0.125$) against the mixture's $+0.274$ $[+0.167, +0.389]$, with every separation from the mixture excluding zero ($+0.149$ to $+0.174$).

The limits of control (e). Three matching gaps remain. (i) Being α-matched, its realized displacement lies inside the range spanned by the mixture's displacements at its two weights on 19 of 31 transitions but exceeds the larger of them on the final 9. Across that band the mixture's own coverage barely moves, so (e)'s lack of a resolved effect on ELF-B and ELF-M is not due to too small a displacement; and equal displacement of the memory would not in any case mean equal effect on the belief. (ii) It exists only in the state-pushed form, so the freshness contrast has no unguided column. (iii) It advances a shadow chain, one extra sampler transition and one memory-free forward per step; both chains index a precomputed tape, so this is a procedure difference, not noise coupling. Two text properties are also unmatched: the changed-token fraction against (e) (0.678 against 0.503, paired difference $+0.175$ $[+0.163, +0.187]$) and embedding dispersion against (d). Neither is established as a rival explanation — (e) raises dispersion without recovering control — but neither is excluded.

Table 4: Direction-matched memory perturbations at the mixture weight circled in Figure 4a. D is coverage minus the default-memory arm of the same column; paired block bootstrap, 144 rows. Directions (b)–(e) are the four controls; control (e) advances a shadow chain and has guided cells only. Row (f), the reward gradient, is not a control; its bold cells are read in the text.

direction	D guided	mixture − dir.	D unguided	PPL g / u
ELF-B, $\alpha = .75$				
(a) pushed belief	$+0.184$ $[+.108, +.260]$	—	$+0.007$ $[-.031, +.049]$	28.1 / 28.1
(b) random	$+0.021$ $[-.031, +.073]$	$+0.163$ $[+.087, +.243]$	-0.024 $[-.052, +.000]$	24.6 / 24.4
(c) orthogonal	-0.007 $[-.059, +.042]$	$+0.191$ $[+.111, +.281]$	-0.014 $[-.052, +.021]$	24.2 / 24.5
(d) other sequence	$+0.042$ $[-.003, +.083]$	$+0.142$ $[+.066, +.222]$	-0.014 $[-.052, +.024]$	24.5 / 24.5
(e) unguided belief	$+0.042$ $[-.028, +.111]$	$+0.142$ $[+.073, +.219]$	—	27.8 / —
(f) reward gradient	$+0.059$ $[-.017, +.135]$	$+0.125$ $[+.059, +.198]$	$+\mathbf{0.868}$ $[+.799, +.931]$	24.5 / 24.9
ELF-M, $\alpha = .5$				
(a) pushed belief	$+0.205$ $[+.101, +.323]$	—	-0.010 $[-.052, +.028]$	38.8 / 37.1
(b) random	$+0.028$ $[-.049, +.115]$	$+0.177$ $[+.083, +.264]$	-0.021 $[-.073, +.017]$	27.5 / 27.2
(c) orthogonal	$+0.038$ $[-.031, +.118]$	$+0.167$ $[+.080, +.257]$	-0.021 $[-.066, +.017]$	27.9 / 27.4
(d) other sequence	$+0.066$ $[-.021, +.160]$	$+0.139$ $[+.049, +.229]$	$+0.010$ $[-.031, +.056]$	27.8 / 27.8
(e) unguided belief	$+0.017$ $[-.056, +.097]$	$+0.188$ $[+.101, +.285]$	—	33.0 / —
(f) reward gradient	$+0.080$ $[+.003, +.167]$	$+0.125$ $[+.038, +.219]$	$+\mathbf{0.559}$ $[+.465, +.649]$	28.9 / 28.6
ELF-L, $\alpha = .75$				
(a) pushed belief	$+0.236$ $[+.118, +.358]$	—	-0.042 $[-.097, +.007]$	37.0 / 35.1
(b) random	$+0.076$ $[-.003, +.156]$	$+0.160$ $[+.069, +.260]$	-0.014 $[-.038, +.010]$	31.1 / 30.3
(c) orthogonal	$+0.128$ $[+.056, +.205]$	$+0.108$ $[+.017, +.201]$	-0.031 $[-.062, +.000]$	31.1 / 30.1
(d) other sequence	$+0.125$ $[+.017, +.240]$	$+0.111$ $[+.038, +.184]$	-0.021 $[-.056, +.010]$	32.6 / 30.8
(e) unguided belief	$+0.083$ $[-.021, +.184]$	$+0.153$ $[+.049, +.260]$	—	35.7 / —
(f) reward gradient	$+0.198$ $[+.083, +.306]$	$+0.038$ $[-.066, +.139]$	$+0.260$ $[+.184, +.340]$	32.4 / 31.0

The state-gradient memory write. Row (f) of Table 4 writes the state proxy's normalised reward gradient into the memory at every transition at the mixture's per-row magnitude, $m_k \leftarrow m_k + \delta_k\, g_k$. Its D-unguided cells ($+0.868$ on ELF-B, $+0.559$ on ELF-M, with no state push) show that a reward direction written into the memory moves the text on its own. They are not evidence for (5): the direction differs, since (5) writes the decoder's gradient at the memory input and (f) the state proxy's gradient, and nothing about this write was used to select the method. Under the state push the mixture's margin over (f) holds at the circled weight on ELF-B and ELF-M, but not at the second weight on ELF-M ($\alpha = .75$; -0.014 $[-0.108, +0.076]$) or on ELF-L.

C.4 Positional and exclusion families

The positional and exclusion families of §3.3 were run on ELF-B on the target sets of the main suite; the exclusion lexicon is *like*, *time*, *people* and *way*. Nine seeds (three, then six added by a rule on interval width alone), stratified cluster bootstrap over seeds with 2000 resamples; the decision form (raw contrast, or guided-minus-unguided) was fixed per family from the unguided contrast before any guided estimate was read, and the unguided arm is the stored unguided texts re-scored under each family's reward. Figure 3 draws the three ladders side by side, with the guided perplexities, and Table 2 gives the coverages and the contrasts.

The zero-memory contrast is positive under every form and family. On exclusion the *unguided* contrast is negative (zero memory is worse without guidance, -0.081 $[-0.117, -0.045]$), which is why its decision form is guided-minus-unguided; under the raw form the zero-memory contrast is still $+0.160$ $[+0.124, +0.194]$, but the $\omega = 1$ contrast becomes $+0.034$ $[-0.001, +0.069]$, so the intermediate rung is form-sensitive on this family while the zero rung is not.

D MEMORY-PORT GRADIENT WRITING: SELECTION AND DESIGN

D.1 SELECTION ON THE DEV HALF

The held-out design uses a difficulty-matched retained-target split: 24 candidate targets probed unguided, the 4 hardest and 4 easiest dropped, the retained 16 split 8 dev / 8 test before any grid ran. Dev values are selection diagnostics, not held-out estimates.

The direction, schedule and magnitude of (5). A grid on the dev half, at 192 rows per cell, compared thirteen arms per checkpoint at $s_z = .25$ with a write at every step: five $\hat{g}$ write arms (constant magnitude $\eta \in \{.05, .1, .2\}$ and $(1-t_k)$-tapered $\eta \in \{.1, .2\}$), one arm writing the pullback $\nabla_m S = \nabla_m S_{\text{proxy}}(D(X_\omega(z_k, t_k, m_k)))$, the belief reward differentiated through the denoiser to the memory input, at $\eta = .05$, four memory-mixture arms ($\alpha = 0$ and $\alpha = .75$, each with and without the tapered $\hat{g}$ write), and unguided, state-guidance and belief-guidance references. The gate is this grid's own unguided mean PPL $+2$ (26.66 on ELF-B, 29.32 on ELF-M). All four mixture arms fall outside the gate on both checkpoints; on ELF-M, the constant and tapered $\hat{g}$ arms at $\eta = .2$ also fall outside it.

The gated argmax was the tapered $\hat{g}$ write at $\eta = .1$ on both checkpoints (dev coverage 0.948 and 0.896, against 0.570 and 0.495 for state guidance); out of sample its gain over state guidance shrank from $+0.378$ to $+0.355$ on ELF-B and from $+0.401$ to $+0.266$ on ELF-M, the winner's-curse optimism expected after an argmax. Two comparisons inside this grid informed the direction and the schedule; the direction comparison is repeated on the test half at the selected operating point (Table 6), the schedule comparison is not. First, the every-step $\nabla_m S$ arm reached 0.893 and 0.714 at $2.21\times$ and $2.34\times$ the cost of state guidance, against 0.948 and 0.896 at $1.73\times$ and $1.74\times$ for the selected $\hat{g}$ arm: on the dev half the cheaper direction was also the better one, at one η each and so not a matched-magnitude comparison. Second, the taper beat the constant magnitude at the same $\eta = .1$ (0.948 against 0.927, 0.896 against 0.823). From this grid only the selected cell and the two fewer-write schedules of Appendix D.3 were carried to test; no arm was selected on test rows.

The write alone's own selection. A separate grid ran eight write-alone cells at $s_z = 0$ on the same dev rows (3 seeds, 192 rows per arm) — constant and $(1 - t_k)$-tapered $\eta \in \{.05, .1, .2, .3\}$ — with unguided and state-guidance references; the rule was the gated argmax under the full four-part gate. It returned the tapered $\eta = .1$ cell on both checkpoints, under the full gate and under the perplexity-only rule alike (Table 5), so the write-alone row of Table 1 is also the write alone's own selection; its dev-to-test coverage is $0.924 \to 0.905$ on ELF-B and $0.721 \to 0.762$ on ELF-M. On ELF-M every cell at $\eta \geq .2$ fails the perplexity criterion.

Table 5: The write alone's own selection grid on the dev rows (192 rows per arm, 24 blocks), $s_z = 0$, write at every step. Coverage, GPT-2-small mean perplexity, and the full four-part gate (the criterion that fires where it fails); $\star$ marks the selection w^* (gated argmax; the perplexity-only rule selects the same cell). The perplexity bound is this grid's own unguided mean $+2$: 26.66 on ELF-B, 29.32 on ELF-M.

	ELF-B			ELF-M		
arm	coverage	PPL	gate	coverage	PPL	gate
unguided	0.036	24.66	pass	0.029	27.32	pass
state guidance .25	0.570	26.12	pass	0.495	27.58	pass
write alone, constant $\eta = .05$	0.870	25.36	pass	0.656	27.34	pass
write alone, constant $\eta = .1$	0.883	25.43	pass	0.615	28.37	pass
write alone, constant $\eta = .2$	0.823	25.29	pass	0.422	29.61	PPL
write alone, constant $\eta = .3$	0.776	25.98	pass	0.427	31.60	PPL
write alone, tapered $\eta = .05$	0.805	25.18	pass	0.599	27.68	pass
write alone, tapered $\eta = .1$ $\star$	0.924	25.13	pass	0.721	28.12	pass
write alone, tapered $\eta = .2$	0.870	25.29	pass	0.638	29.43	PPL
write alone, tapered $\eta = .3$	0.844	25.63	pass	0.516	30.65	PPL
w^* − state guidance (dev)	$+0.354$ [$+0.276$, $+0.438$]			$+0.227$ [$+0.128$, $+0.331$]		

D.2 THE WRITTEN DIRECTION

On the test rows (6 seeds, 384 rows per arm), the pullback $\nabla_m S$ was written with the method's schedule and η, alone and on top of state guidance at .25, alongside unguided, state guidance and the $\hat{g}$ write alone. Its unguided, state-guidance and write-alone rows reproduce those of Table 1 exactly (Appendix G), so the combined pullback arm is paired against the combined arm of Table 1. Table 6 gives the arms and four contrasts. Written alone the pullback is worse than $\hat{g}$ on both checkpoints and outside the perplexity gate on both (on ELF-B by 0.03: 27.17 against $25.14 + 2$, with no other criterion firing), at $1.49\times$ / $1.60\times$ the cost of state guidance where $\hat{g}$ costs $0.99\times$. On top of state guidance it is not distinguished from $\hat{g}$ on ELF-B (upper bound exactly zero) and is worse and outside the gate on ELF-M, at $2.27\times$ / $2.35\times$. One η and one schedule were compared.

Table 6: The direction held out: the pullback $\nabla_m S$ against $\hat{g}$ at the method's $\eta = .1$ and taper, test rows (384 per arm per checkpoint, 48 blocks). Coverage, GPT-2-small mean perplexity, the full four-part gate, and the paired per-batch cost ratio against the state guidance of the same run. The four contrasts are paired on identical rows; the contrast of the two combined arms pairs these pullback rows with the $\hat{g}$ rows of Table 1, which this run reproduces exactly. The $\hat{g}$ arms are the ones Table 1 reports (combined arm: 0.951 / 0.883, gated).

model	arm	coverage	PPL	gate	cost
ELF-B	unguided	0.048	25.14	pass	$0.22\times$
ELF-B	state guidance .25	0.595	25.95	pass	$1.00\times$
ELF-B	write alone ($\hat{g}$), $s_z = 0$	0.905	26.08	pass	$0.99\times$
ELF-B	pullback write alone ($\nabla_m S$), $s_z = 0$	0.711	27.17	fails	$1.49\times$
ELF-B	state guidance + pullback write ($\nabla_m S$)	0.911	27.08	pass	$2.27\times$
	pullback alone − write alone	-0.194 $[-0.283, -0.120]$			
	pullback alone − state guidance	$+0.116$ $[+0.027, +0.195]$			
	state guidance + pullback − combined	-0.039 $[-0.077, +0.000]$			
	state guidance + pullback − state guidance	$+0.316$ $[+0.254, +0.379]$			
ELF-M	unguided	0.070	27.01	pass	$0.25\times$
ELF-M	state guidance .25	0.617	27.27	pass	$1.00\times$
ELF-M	write alone ($\hat{g}$), $s_z = 0$	0.762	27.46	pass	$0.99\times$
ELF-M	pullback write alone ($\nabla_m S$), $s_z = 0$	0.297	30.04	fails	$1.60\times$
ELF-M	state guidance + pullback write ($\nabla_m S$)	0.776	29.62	fails	$2.35\times$
	pullback alone − write alone	-0.465 $[-0.539, -0.391]$			
	pullback alone − state guidance	-0.320 $[-0.385, -0.259]$			
	state guidance + pullback − combined	-0.107 $[-0.150, -0.065]$			
	state guidance + pullback − state guidance	$+0.159$ $[+0.112, +0.208]$			

The gradients on the unwritten chain. Let $g = \nabla_{\hat{x}} S_{\text{proxy}}(D(\hat{x}))$ be the decoder-only reward gradient at the belief, so that $\nabla_m S = J_m^\top g$. On state-guided chains in the middle of the trajectory (steps 12–16 of $N = 32$ on ELF-M, 24–32 of a 64-step run on ELF-B; 191 and 201 rows), the state proxy's gradient $\nabla_z S_{\text{proxy}}(D(z))$ is orthogonal to $\nabla_m S$ (median cosine 0.000 on both checkpoints), and $\nabla_m S$ is nearly orthogonal to g (0.088 and 0.069) and to the mixture direction $B_k - m_k$ of §5.2. A finite memory step of size .05 along g nonetheless raises the reward about as much as the same step along $\nabla_m S$ on ELF-B (ratio 1.03 $[0.85, 1.27]$) and 0.62 $[0.43, 0.90]$ of it on ELF-M. The vector the method writes, $\hat{g}$, is the same decoder gradient evaluated at the memory input rather than at the belief.

D.3 THE WRITE-COUNT FRONTIER

A separate run on the same test rows evaluated unguided, state guidance, (5) on top of state guidance at every 3rd step (11 writes), every 2nd step (16) and every step (31), and belief guidance. The every-step arm exceeds state guidance and belief guidance on both checkpoints, every lower bound above $+0.05$. Figure 6 draws that run against the write count; every arm passes that run's perplexity gate (unguided mean $+2$: 27.14 on ELF-B, 29.01 on ELF-M). Coverage at 0 / 11 / 16 / 31 writes is 0.595 / 0.874 / 0.910 / 0.951 on ELF-B and 0.617 / 0.779 / 0.790 / 0.883 on ELF-M. Priced on that run's own timing basis, every frontier arm also exceeds oracle best-of-N at its own matched cost under both selection rules. On ELF-L the 11- and 16-write arms ran on the same rows and tapes, every arm inside the full gate, with nothing selected on ELF-L; coverage is 0.609 / 0.796 / 0.807

/ 0.882, the every-step arm being the ELF-L session of Table 1 costed against that session's state guidance, and every step exceeds every 3rd step by $+0.086$ $[+0.044, +0.128]$.

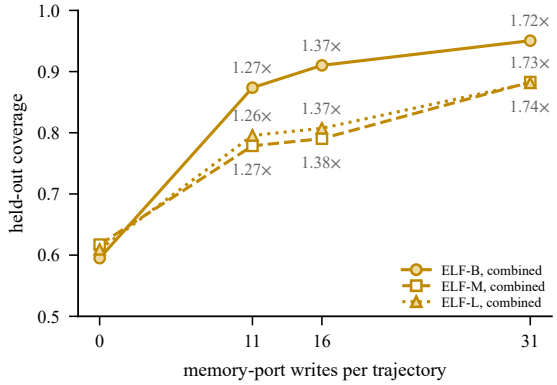

Figure 6: The write-count frontier (Appendix D.3): held-out coverage against the number of memory-port writes per trajectory, on top of state guidance at .25 (0 writes is state guidance itself), three checkpoints. The within-session wall-clock ratio against state guidance is printed at each point; the ELF-L every-step point is costed against its own session's state guidance. Point estimates; contrasts in the text.

E HELD-OUT EVALUATION AND CONTROLS

The evaluation uses the test half of the split of Appendix D.1 and six test seeds. Table 1 reports the controls session; Appendix G gives the sessions that generated the held-out rows and the anchors between them.

E.1 BEST-OF-N

Table 7: Every arm against oracle best-of-N at its own matched cost. The oracle draws from a pool of 8 unguided candidates per row and selects under two rules: fractional coverage, and the hard all-or-nothing verifier scored as its own indicator; Δ: arm minus oracle, with 95 % block-bootstrap intervals. $N = \lceil \text{arm/unguided} \rceil$ in within-session wall-clock seconds on the timing basis of Table 1.

model	arm	N	best-of-N, cov / all	Δ coverage	Δ all-targets
ELF-B	state guidance	5	0.169 / 0.135	$+0.426$ $[+0.353, +0.497]$	$+0.411$ $[+0.341, +0.482]$
	write alone	5	0.169 / 0.135	$+\mathbf{0.736}$ $[+0.693, +0.779]$	$+0.737$ $[+0.690, +0.786]$
	combined	8	0.257 / 0.201	$+0.694$ $[+0.652, +0.734]$	$+0.732$ $[+0.682, +0.779]$
	belief guidance	7	0.223 / 0.174	$+0.527$ $[+0.469, +0.586]$	$+0.523$ $[+0.461, +0.581]$
ELF-M	state guidance	5	0.221 / 0.172	$+0.396$ $[+0.337, +0.453]$	$+0.383$ $[+0.323, +0.440]$
	write alone	4	0.176 / 0.133	$+\mathbf{0.586}$ $[+0.522, +0.650]$	$+0.583$ $[+0.516, +0.654]$
	combined	7	0.306 / 0.242	$+0.577$ $[+0.527, +0.630]$	$+0.612$ $[+0.560, +0.664]$
	belief guidance†	7	0.306 / 0.242	$+0.085$ $[+0.022, +0.148]$	$+0.099$ $[+0.036, +0.161]$
ELF-L	state guidance	4	0.190 / 0.143	$+0.419$ $[+0.346, +0.493]$	$+0.414$ $[+0.341, +0.487]$
	write alone	4	0.190 / 0.143	$+0.592$ $[+0.530, +0.652]$	$+0.609$ $[+0.547, +0.672]$
	combined	7	0.307 / 0.247	$+0.574$ $[+0.526, +0.622]$	$+0.596$ $[+0.547, +0.646]$

Reading the best-of-N comparison. Wall clock is the only convention that charges the decoder gradient, which dominates these arms' cost: it charges the combined arm best-of-8 / best-of-7 / best-of-7, the largest N of any arm. A convention that counts denoiser and decoder passes at one unit each charges the write alone $N = 2$ and the combined arm $N = 3$, and both arms' margins are larger under it, so the wall-clock comparison is the stronger baseline. The oracle selects on the

target-coverage criteria the arms are evaluated on, including a hard all-or-nothing rule, while the method optimises a differentiable proxy, so the baseline is strong rather than a straw man; because a practitioner without the target list cannot select this way, the comparison bounds what the method is worth against test-time sampling rather than describing a deployable baseline.

E.2 Trajectory diagnostics and saturation

The trajectory panels. Figure 5(d–f) reads the preview and acceptance diagnostics recorded at steps 8, 16, 24 and 31 by the controls session (ELF-B, ELF-M) and by the ELF-L session, after the step's state push (the write alone has no push and is read after its write). At $k = 31$ acceptance equals the terminal coverage of Table 1 on eight of the nine arm–checkpoint pairs; on the ELF-M write alone it is 0.759 against 0.762; the diagnostic recomputes the written memory in its own pass rather than reading the sampler's.

Saturation. Rows whose $\mathrm{RMS}(\hat{g}_k) \leq 10^{-6}$ receive no write, so a vanishing gradient is never normalised into a full-magnitude step. This no-op fired on $25.7\,\%$ of the combined arm's writes on ELF-B, $23.8\,\%$ on ELF-M and $30.9\,\%$ on ELF-L — one write in four to one in three, rising with the number of writes ($17.0\,\%$ and $15.3\,\%$ at 11 writes, $20.9\,\%$ and $17.6\,\%$ at 16). For the write alone it fired on $18.8\,\%$, $14.8\,\%$ and $21.6\,\%$, and for the placebo never, since a Gaussian direction has no vanishing rows. The gradient is computed before its RMS is tested, so a suppressed write still costs what an executed one costs: the effective schedule is 21 to 24 writes rather than 31, at the price of 31.

E.3 Emitting the preview

Skipping the final transition and emitting the decode of the last intermediate state, $\mathrm{text}(D(z_{N-1}))$, is the obvious way to keep the preview. We scored it on the held-out rows of Table 1 from the stored texts, without new sampling: the controls session and the ELF-L session store, for every state-guidance row, the preview text at step 31 after the last push. The emitted preview goes through the same four-part gate and the same paired block bootstrap as Table 1 (Table 8). It raises coverage over state guidance on every checkpoint, but fails the gate on perplexity on all three; the structural, entropy and repetition criteria do not fire. The failure is not a tail of a few broken texts: its median perplexity exceeds the unguided median by 3.3 to 5.7, against at most 1.1 for the arms of Table 1. It also covers less than the combined arm on every checkpoint; against the write alone its intervals include zero on ELF-B and ELF-M and favour the write alone on ELF-L.

Table 8: Emitting the preview, on the test rows of Table 1 (384 rows per arm). The gate's perplexity bound is the unguided mean $+2$. Contrasts: paired block bootstrap on identical rows (48 blocks, 2000 resamples), $95\,\%$ intervals.

	ELF-B	ELF-M	ELF-L
coverage			
state guidance	0.595	0.617	0.609
preview emitted	0.871	0.736	0.701
combined	0.951	0.883	0.882
GPT-2-small perplexity, mean / median			
unguided	25.14 / 24.41	27.01 / 26.20	30.34 / 29.34
preview emitted	41.90 / 28.46	32.64 / 29.53	36.62 / 35.05
gate, preview emitted	fails (PPL)	fails (PPL)	fails (PPL)
paired contrasts			
preview − state guidance	$+0.276\,[+0.211, +0.337]$	$+0.118\,[+0.077, +0.161]$	$+0.091\,[+0.044, +0.141]$
combined − preview	$+0.079\,[+0.048, +0.109]$	$+0.147\,[+0.100, +0.193]$	$+0.181\,[+0.138, +0.227]$
write alone − preview	$+0.034\,[-0.007, +0.073]$	$+0.026\,[-0.021, +0.073]$	$+0.082\,[+0.029, +0.133]$

E.4 The method on the positional and exclusion families

The configuration of (5), unchanged and with nothing re-selected, was run on the two further families on the method's own held-out rows: the arms of Table 1, its six test seeds, eight target sets and noise tapes. Positional keeps each set's keyword targets, with the window at token 128. Exclusion uses

the lexicon of Appendix C.4, assigned in a fixed cycle so that each set keeps its number of targets; absence is exact-word, so *likes* does not count as *like*, and the family is not a safety criterion. Each family's proxy ((7)) replaces the keyword proxy in every push and in the write. The unguided rows equal those of Table 1 token for token, and every guided arm runs on its Table 1 tapes. A contrast holds if its lower bound is above zero with both arms inside the full gate.

Table 9: The method, unchanged, on the positional and exclusion families, on the held-out rows and tapes of Table 1 (384 rows per arm per pass, 48 blocks). Coverage: fraction of targets with an accepted occurrence starting before token 128 (positional), fraction of the set's words absent (exclusion). Lower rows: paired contrasts with 95 % block-bootstrap intervals (comb.: the combined arm). †: outside the quality gate, on perplexity; every other arm passes. —: belief guidance not run on ELF-L.

	positional			exclusion		
arm	ELF-B	ELF-M	ELF-L	ELF-B	ELF-M	ELF-L
unguided	0.003	0.004	0.007	0.441	0.339	0.344
state guidance	0.530	0.591	0.514	0.763	0.729	0.729
belief guidance	0.768	0.327	—	0.941†	0.810	—
write alone	0.859	0.634	0.686	0.789	0.556	0.607
combined	0.914	0.846	0.854	0.905	0.835	0.810
write − state	+0.329 [.263, .391]	+0.043 [−.027, .111]	+0.172 [.081, .260]	+0.026 [−.017, .072]	−0.173 [−.233, −.111]	−0.122 [−.174, −.074]
comb. − state	+0.384 [.306, .461]	+0.255 [.201, .309]	+0.340 [.270, .410]	+0.142 [.099, .185]	+0.105 [.060, .156]	+0.081 [.042, .121]
comb. − belief	+0.146 [.078, .219]	+0.520 [.448, .595]	—	−0.036† [−.068, −.008]	+0.025 [−.016, .069]	—
comb. − write	+0.055 [.016, .095]	+0.212 [.165, .257]	+0.168 [.109, .223]	+0.116 [.074, .158]	+0.279 [.234, .323]	+0.203 [.145, .262]

The combined arm beats state guidance and the write alone on all six passes. The write alone beats state guidance on positional ELF-B and ELF-L; its interval includes zero on positional ELF-M and exclusion ELF-B, and on exclusion ELF-M and ELF-L it is below state guidance in every one of the six seeds. Belief guidance is below state guidance on positional ELF-M, also in every seed. On exclusion ELF-B belief guidance reaches higher absence than the combined arm but leaves the gate. Costs against state guidance are close to those of Table 1: 0.99× for the write alone, 1.73–1.79× for the combined arm and 1.50–1.62× for belief guidance. The closest gate margin is the combined arm on exclusion ELF-M, which passes the perplexity criterion by 0.07 (28.94 against the bound 29.01); on ELF-L every arm passes on every family, against an unguided perplexity of 30.34.

E.5 Controls: full tables

The controls session ran the arms of Table 10 serially in one process on the test rows (384 rows per arm per checkpoint) and the state-scale ladder on the dev rows (192 rows per arm, Table 11). Every contrast is a paired block bootstrap on identical rows (48 blocks, 2000 resamples). On ELF-L the same arms ran in a separate session on the rows and noise tapes of Table 1, with nothing selected on ELF-L; its write alone and combined arm are the rows of the ELF-L session of Table 1, whose unguided and state-guidance rows the new session reproduces token for token on all 384 rows.

Comparator selection. The dev rule selects the strongest state scale inside the gate, ties toward the smaller scale; it was fixed with the perplexity criterion alone. Under the full four-part gate it selects .25 on ELF-B and ELF-M and .35 on ELF-L; under the perplexity criterion alone, .25, .50 and .35. Table 11 names the criteria that fire at the larger scales and their margins. Re-run under nearby thresholds (perplexity tolerance 2 or 3, 4-gram repetition tolerance .010 or .015, entropy-gap multiple 1.5 to 1.8), the rule selects .25 on ELF-B at perplexity tolerance 2 and .75 at 3, and .35 and .50 on ELF-L, perplexity being the only criterion that fires on either; on ELF-M the selection moves between .25, .35 and .50 with the entropy-gap and repetition thresholds, so the dev-selected comparator there is a decision of the gate, not of the coverage ladder. Table 1 uses .25 on every checkpoint, which on ELF-L is not the dev selection; Table 12 gives the contrasts against every scale.

Table 10: The control arms on the test rows (384 rows per arm; on ELF-L the ten without belief guidance). Coverage, GPT-2-small mean perplexity, 4-gram repetition, and the within-session wall-clock cost against state guidance at .25 (mean over 96 paired batches). A superscript marks an arm outside the full four-part quality gate and the criteria that fire: p perplexity, e entropy gap to OpenWebText, r repetition; every other arm passes. *lin* marks the $(1 - t_k)$ taper applied to the guidance scale; the placebo writes a unit-RMS Gaussian direction with the write's schedule, $\eta = .1$ and saturation rule, on top of state guidance at .25. —: not run.

	ELF-B				ELF-M				ELF-L			
arm	cov.	PPL	rep-4	cost	cov.	PPL	rep-4	cost	cov.	PPL	rep-4	cost
unguided	0.048	25.14	.036	0.22	0.070	27.01	.013	0.25	0.060	30.34	.009	0.26
state .25	0.595	25.95	.035	1.00	0.617	27.27	.014	1.00	0.609	31.28	.009	1.00
state .35	0.688	26.36	.035	1.00	0.740	27.11	.016	1.00	0.728	31.34	.009	1.00
state .50	0.736	26.90	.035	1.00	0.815	26.92	.018	1.00	0.773	32.23	.010	1.00
state .75	0.828	27.00	.037	1.00	$0.811^{e,r}$	27.74	.028	1.00	$0.832^{e,p}$	32.71	.011	1.00
state .25, lin	0.409	25.92	.036	1.00	0.484	27.09	.014	1.00	0.495	31.28	.008	1.00
state .50, lin	0.646	26.38	.036	1.00	0.740	27.35	.017	1.00	0.736	31.76	.010	1.00
belief .25	0.750	25.92	.035	1.49	0.391^{p}	29.33	.011	1.62	—	—	—	—
belief .25, lin	0.714	25.91	.035	1.49	0.259^{p}	29.30	.011	1.61	—	—	—	—
belief .50, lin	0.855	26.06	.040	1.49	0.374^{p}	29.20	.013	1.61	—	—	—	—
write alone	0.905	26.08	.034	0.98	0.762	27.46	.012	0.99	0.783	31.04	.008	0.99
state .25 + placebo	0.564	26.06	.034	1.01	0.608	27.18	.014	1.00	0.651	31.24	.008	1.00
combined	0.951	26.17	.035	1.76	0.883	28.05	.013	1.74	0.882	31.76	.008	1.73

Table 11: The state-scale ladder on the dev rows, where the scale is selected (192 rows per arm). Gate criteria that fire under the full four-part gate are named with the offending value against its threshold; the entropy-gap multiple is given under the GPT-2 and T5 tokenizers (the criterion requires both to exceed 1.5).

model	scale	coverage	PPL	rep-4	full gate
ELF-B	unguided	0.036	24.66	.036	—
ELF-B	.25	0.570	26.12	.032	pass
ELF-B	.35	0.708	26.81	.032	PPL ($26.81 > 26.66$)
ELF-B	.50	0.766	27.21	.032	PPL
ELF-B	.75	0.852	27.46	.035	PPL
ELF-M	unguided	0.029	27.32	.011	—
ELF-M	.25	0.495	27.58	.014	pass (gap $1.38\times$ / $1.35\times$)
ELF-M	.35	0.549	27.77	.015	entropy gap ($1.59\times$ / $1.53\times$)
ELF-M	.50	0.729	28.33	.021	entropy gap ($1.78\times$ / $1.69\times$), repetition ($.0214 > .0207$)
ELF-M	.75	0.737	30.53	.034	entropy gap, PPL, repetition
ELF-L	unguided	0.029	30.67	.007	—
ELF-L	.25	0.531	31.92	.008	pass
ELF-L	.35	0.690	32.20	.009	pass
ELF-L	.50	0.742	33.04	.008	PPL ($33.04 > 32.67$)
ELF-L	.75	0.810	34.01	.011	PPL

Contrasts against every scale. Table 12 gives the paired contrasts on the test rows. The placebo write against state guidance at .25 is -0.031 $[-0.076, +0.013]$, -0.009 $[-0.056, +0.038]$ and $+0.042$ $[-0.016, +0.102]$ on ELF-B, ELF-M and ELF-L, and the combined arm against the placebo is $+0.387$ $[+0.311, +0.458]$, $+0.275$ $[+0.220, +0.328]$ and $+0.230$ $[+0.177, +0.283]$. Tapered guidance at peak .25 against its constant-scale arm is -0.186 $[-0.237, -0.138]$, -0.133 $[-0.188, -0.079]$ and -0.115 $[-0.173, -0.052]$ for state guidance on ELF-B, ELF-M and ELF-L, and -0.036 $[-0.069, -0.007]$ and -0.132 $[-0.180, -0.083]$ for belief guidance on ELF-B and ELF-M, where both belief arms are outside the gate. Tapered state guidance at peak .50 against constant .25, not in Table 12, is $+0.051$ $[-0.001, +0.099]$ on ELF-B, $+0.122$ $[+0.068, +0.173]$ on ELF-M and $+0.126$ $[+0.081, +0.173]$ on ELF-L.

Table 12: Paired contrasts on the test rows, block bootstrap on identical rows (384 rows, 48 blocks), against every state scale of the ladder and every other comparator. †: comparator outside the full four-part gate on the test rows. —: belief guidance not run (ELF-L).

comparator	ELF-B	ELF-M	ELF-L
write alone − comparator			
state guidance .25	$+0.310$ $[+0.245, +0.375]$	$+0.145$ $[+0.089, +0.204]$	$+0.173$ $[+0.107, +0.241]$
state guidance .35	$+0.217$ $[+0.165, +0.275]$	$+0.022$ $[-0.027, +0.076]$	$+0.055$ $[-0.007, +0.117]$
state guidance .50	$+0.169$ $[+0.120, +0.221]$	-0.053 $[-0.105, +0.001]$	$+0.009$ $[-0.055, +0.070]$
state guidance .75	$+0.077$ $[+0.034, +0.122]$	-0.049 $[-0.103, +0.007]^{\dagger}$	-0.049 $[-0.111, +0.012]^{\dagger}$
unguided	$+0.857$ $[+0.820, +0.892]$	$+0.691$ $[+0.641, +0.742]$	$+0.723$ $[+0.667, +0.779]$
combined − comparator			
state guidance .25	$+0.355$ $[+0.284, +0.422]$	$+0.266$ $[+0.217, +0.314]$	$+0.272$ $[+0.211, +0.333]$
state guidance .35	$+0.263$ $[+0.206, +0.319]$	$+0.143$ $[+0.102, +0.189]$	$+0.154$ $[+0.104, +0.201]$
state guidance .50	$+0.215$ $[+0.158, +0.271]$	$+0.068$ $[+0.026, +0.112]$	$+0.108$ $[+0.064, +0.148]$
state guidance .75	$+0.122$ $[+0.077, +0.169]$	$+0.072$ $[+0.021, +0.125]^{\dagger}$	$+0.049$ $[+0.013, +0.086]^{\dagger}$
tapered state .25	$+0.542$ $[+0.473, +0.609]$	$+0.398$ $[+0.331, +0.465]$	$+0.387$ $[+0.322, +0.456]$
tapered state .50	$+0.305$ $[+0.246, +0.362]$	$+0.143$ $[+0.081, +0.207]$	$+0.146$ $[+0.095, +0.197]$
tapered belief .25	$+0.237$ $[+0.174, +0.305]$	$+0.624$ $[+0.573, +0.674]^{\dagger}$	—
tapered belief .50	$+0.095$ $[+0.049, +0.146]$	$+0.509$ $[+0.452, +0.573]^{\dagger}$	—
write alone	$+0.046$ $[+0.012, +0.078]$	$+0.121$ $[+0.073, +0.164]$	$+0.099$ $[+0.051, +0.148]$
belief guidance .25	$+0.201$ $[+0.143, +0.262]$	$+0.492$ $[+0.434, +0.553]^{\dagger}$	—

F GENERATED TEXT

Both blocks show the first seed and sample 0 of every target set: the first block on ELF-B, the checkpoint and rows of Figure 2(a), the second on ELF-L, the largest checkpoint. Each excerpt is a window of 12 words on each side of the first occurrence of a target, or the first 24 words if no target occurs, cut from a longer text, with a window of 6 words added around every further met target outside it; targets are in bold, the number of targets met is in parentheses, and non-ASCII characters are transliterated. The excerpts are otherwise unedited model outputs: names, events and claims in them are generated and do not refer to real persons or facts, and in two excerpts the names of persons to whom the model attaches a fabricated crime or death are replaced by [name]. At $k = 31$ the acceptance decode and the emitted text agree on the reward on all three rows of the first block and differ in at most one token position. The preview at $k = 16$ is the decode of a partially noised state and does not read as text; it is what a state-side proxy scores. In the second block the four arms share a noise tape, so they often keep one topic and phrasing and differ where a target enters. The rewards are lexical: every guided arm, state guidance included, sometimes places a target where it does not fit the context.

Preview and acceptance along one trajectory. ELF-B, state guidance at .25, seed 101, sample 0 of each target set of §4; decodes after the step-k update. Each line gives whether the target is met, then the excerpt.

Targets: technology.

- $k = 16$, preview (0/1): Le Line vorbe este este a orice easyala sus este, cat scor este OK dar doi Liverpool in alte. Dir de 11. Il si . . .

- $k = 16$, acceptance (0/1): UPDATE: This is a great good download of, and if you're a check out the video, and it is it amazing. I like I ...
- $k = 31$, preview (1/1): ... you plan toward the progress of a first design? As a concept **technology**, I will ask you what steps you need to take to achieve ...
- $k = 31$, acceptance (0/1): UPDATE: This is a great game to watch, and if you enjoyed it, just check out the video and you will enjoy it! As ...

Targets: climate change.

- $k = 16$, preview (1/1): ... Images acum si.ati este savor a in se stood mai in in **climate change**. Salut care in cand bag bot din in scam in drum ...
- $k = 16$, acceptance (0/1): After England's 2-1 win beis City had two of the league team, in 14 the first half saw Washington take a nine-point lead from ...
- $k = 31$, preview (0/1): New England's 1-0 win over Las City with two goals from both sides that went into the first half saw Florida hold a eight-point ...
- $k = 31$, acceptance (1/1): ... to be seen whether the team is the right one to tackle **climate change**. An emergency video shows the ley bull jumping out of the ...

Targets: music, history.

- $k = 16$, preview (2/2): ... am Index pentru.., arrange ce cand sau sau, si, in special sa **music** sa pentru delicate shadow le care. am zeci faca in au fie ... Sigur ge Prepare si danbar. La **history** de dar eaza Olympic ar golden ...
- $k = 16$, acceptance (1/2): ... multimedia on your iPhone and and Android devices. Add the device's mobile **music** apps. The app lets apps for various dishes, such as to dishes ...
- $k = 31$, preview (2/2): ... to play sports games such as soccer games, run running, listen to **music**, **music** sich pentru social occasions, ride a bike tour, walk dans a ... to sell their properties. OVERGARE, N.J., **history** owners of a military jet pack ...
- $k = 31$, acceptance (1/2): ... to play sports games such as football games, run running, listen to **music**, **music** participants for social activities, ride a bike ride, walk into a ...

The four arms on the same rows. ELF-L, the held-out session of Table 1, test seed 1101, sample 0 of each of the eight test target sets; the four arms share the row and its noise tape.

Targets: hospital, lawyer.

- unguided (0/2): "These cuts will have a substantial negative impact on U.S. economic performance, jobs and economic growth," Amherzen Group chief economist John J. Nichols said ...
- state guidance (1/2): ... and no immediate military action is being taken." British nurse killed in **hospital** overwhelmed in Manchester because of Sea de France strain UPDATE A British ...
- write alone (0/2): "These cuts will have a significant negative effect on U.S. job creation, wages and employment growth," Amherst Institute chief economist John J. Nichols said ...
- combined (1/2): ... worsend during this period." A 80-year-old diplomat [name] was taken to the **hospital** for his life after the mansion was darkened by flooding water. [name] ...

Targets: hospital.

- unguided (0/1): Chris Eugene is another well-liked player early in the NBA. All-Star Chamel Johnson had an excellent return this offseason. He's averaged 2.5 points and ...
- state guidance (1/1): ... most from last season. He's played poorly on the defensive side in **hospital**. He has the temptation to run the ball in traffic and has ...
- write alone (1/1): ... and I'd be surprised if he's recruited to Pittsburgh. He is a **hospital** menace at the 6'4 slot and will be an interesting option if ...
- combined (1/1): ... should come in a big snap, being a good defensive keymaker too. **hospital** Carter Carter is another interesting addition to the Oilers too. A veteran ...

Targets: museum.

- unguided (0/1): The vast majority of her working time at HP has been spent filling budgets, managing maintenance, maintenance, and operations. "If you spend less time ...
- state guidance (0/1): For the majority of her working life, she has been on breaks, docking, for hours at a time. She takes the stuff off her ...
- write alone (1/1): ... businesses that are struggling," he said. "Those struggles are the cornerstone behind **museum** Detroit Job Plan." So where do those jobs ideas come from? It's ...

- combined (1/1): … working life, she has been spent vacation, boating, hiking and visiting a **museum**. She takes whatever is left her job in the middle of the …

Targets: restaurant.

- unguided (0/1): As an audio lover, recording video with your iPad has been extremely fun as I've spent much of the time thinking about what I …
- state guidance (1/1): … control of nearby cafes, restaurants, and shops. It depends on type of **restaurant** you'll want to visit, as it's not always possible to watch music …
- write alone (0/1): My iPad's audio (OBI) experience has been excellent, and I've spent much of the time thinking about what I can do with my iPad …
- combined (1/1): … by using an USB jack which will allow you to visit a **restaurant** and walk around comfortably. By using the USB jack you don't even …

Targets: science, budget.

- unguided (0/2): There was much to be said about what happened. But that is strange. Sometimes broadcast media can make an investment in a sport that …
- state guidance (2/2): … [name], who shot quarterback [name] – and the organization consumeed too much **budget**. So planners proposed that the Arena be dedicated to the family of … Portland, Oregon, which is working on **science** an giant new operating headquarters and …
- write alone (2/2): … is far from an affirmation of sport (focusing on, after all, the **science** fiction mission of suppressing people's opinion) or if it is just as … start focusing a on their tax **budget**. The New York Times reported Los …
- combined (2/2): … to reoccur... but now is providing no evidence at all. This is **science** showing how the theory moved away from the affirmative of game-building and, … Coast have started an effort to **budget** "fly controller training." A Senate Republican …

Targets: software.

- unguided (0/1): A car lender is not only the easiest way to save a penny, but it is the cheapest way to save money. There are …
- state guidance (1/1): … 'PHML Editor (Editor).' This gets a lot of attention given the massive **software** design foundation OpenGL has. The process of turning a HTML into a …
- write alone (1/1): … check out a trustworthy shop like Cash Saving One. Beware of smaller **software** OS cards as the smaller monthly discount option is available. How to …
- combined (1/1): … PhML Designing and Programming If you're new to the basics of PhML **software**, or just want to know how to add a new project quickly, …

Targets: supreme court.

- unguided (0/1): Then, I have a plug-in port for my tablet and a Spotify Media player. Since I have the Spotify Media player, I don't have …
- state guidance (1/1): Then **supreme court** it. Doesn't normally you just want a long can of chocolate? …
- write alone (1/1): … you out there, this is the feature that I'm glad is the **supreme court** for device owners, but you need to update it manually in …
- combined (1/1): Then **supreme court** iTunes! Redditors.com is also a popular service, and though I love …

Targets: weather.

- unguided (0/1): "We have a lot of young people in this area that don't care," said Davis, who is raised in the Red Point area. He …
- state guidance (1/1): … small village and didn't smoke all the time. He didn't know how **weather** prepared as a child, he said. It wasn't that he felt disadvantaged …
- write alone (1/1): … of Technology to become a civilian psychologist. "My little brother loved the **weather**, my dad loved the difficulty sleeping and that was often when he …
- combined (1/1): … of organizations travelled across the globe to deal with the spread of **weather** fever in the region. Kasjar Abu, 51, is from the island nation …

G Reproducibility and experiment mapping

All arms use the public OpenWebText checkpoints `checkpoint_95085` (ELF-B, 105M parameters), `checkpoint_76070` (ELF-M, 342M) and `checkpoint_57051` (ELF-L, 652M), the released stochastic (SDE) sampler with churn $\gamma = 1.5$, batch size 4 and sequence length 1024. A

Table 13: Mean sampling seconds per sample, per session. The controls session prices Tables 1 (ELF-B, ELF-M), 7 and 10, and on ELF-L the arms it ran; the ELF-L session prices the ELF-L rows of Table 1 and the write alone and combined arm wherever they appear; the frontier session prices Figure 6; the write-alone selection and pullback session prices Table 6. In the controls session the larger state scales are within 0.003 of state guidance at .25.

session	arm	ELF-B	ELF-M	ELF-L
controls	unguided	0.153	0.380	0.612
	state guidance .25	0.687	1.527	2.370
	write alone	0.673	1.515	—
	combined	1.211	2.662	—
	belief guidance .25	1.020	2.466	—
ELF-L	unguided	—	—	0.623
	state guidance .25	—	—	2.398
	write alone	—	—	2.380
	combined	—	—	4.155
frontier	unguided	0.182	0.398	—
	state guidance .25	0.718	1.550	—
	every-step write (combined)	1.234	2.693	—
pullback	unguided	0.152	0.383	—
	state guidance .25	0.692	1.533	—
	write alone	0.684	1.519	—
	pullback alone	1.027	2.456	—
	state guidance + pullback	1.565	3.608	—

row is one generated sample, and a sample's coverage is the fraction of its targets met. The main suite of §4–§5 uses seeds 101/202/303, three target sets and 16 samples per set and seed: 144 samples per arm, or 192 target rows, since one set has two targets. The ELF-B memory ladder and the preview–acceptance counts of §4 and Appendix B are over these 192 target rows, in 36 blocks of four samples; every other quantity is over samples: the memory-mixture and direction arms use 144, the ELF-M and ELF-L ladders 432 over nine seeds, and the family ladder 432, with whole seeds as its bootstrap unit. The held-out design of §6 uses six test seeds, eight target sets and eight samples, 384 samples per arm in 48 seed $\times$ target-set blocks, and three dev seeds with 192 samples in 24 blocks. Intervals are paired block bootstrap at 95 %.

Sessions and anchors. The held-out rows were generated in serial sessions on the same test half, seeds and noise tapes. For compute, the test pass was cut from 34 to 6 seeds and from 8 to 4 arms and rerun, without inspecting any outcome. A frontier session evaluated (5) at three write counts (Appendix D.3); the controls session regenerated the arms of Table 1 alongside the controls, and that table reports it. The frontier and controls sessions are coverage-identical on every shared arm and seed (3072 of 3072 rows). The write-alone selection and pullback session reproduces the controls session's unguided, state-guidance and write-alone rows exactly (coverage, terminal token ids and tape hashes, on all test and dev rows), which is the condition under which its pullback arms are paired with the controls session's arms. The family session's unguided rows equal the controls session's token for token. The ELF-L session runs on the same tapes, checked on every row, and the ELF-L controls session reproduces its unguided and state-guidance rows token for token on all 384 rows. An earlier session, whose other arms are not reported, reproduces the unguided and state-guidance rows token for token on all six ELF-M seeds and on five of six ELF-B seeds; on ELF-B seed 1303 its terminal tokens differ on all 64 rows of both arms from the later sessions, which agree with each other, while rng keys, schedule and noise hashes, targets and candidate ids all match; we have not resolved the cause. Excluding that seed leaves the combined arm's ELF-B contrasts with state and belief guidance at $+0.366$ and $+0.205$ (from $+0.355$ and $+0.201$).

Timing. Every cost ratio in the paper is within one session, on one RTX 4090 with instrumentation excluded; Table 13 gives the mean sampling seconds per sample behind them. A session's unguided figure is the unit price of a best-of-N candidate for that session's arms. Wall clock is hardware- and implementation-dependent, and these numbers do not transfer to another device.