WHAT A MODEL DOES VERSUS WHERE IT GOES: DECOMPOSING POST-TRAINING BEHAVIORAL CHANGE

Anonymous authors
Paper under double-blind review

ABSTRACT

Post-training is often said to change a behavior, such as how often a model writes "delve" or opens a paragraph with "First", on the strength of counts in each model's own trajectories. Such a count mixes two effects. The post-trained model, the successor, may be more likely than its predecessor to produce the behavior in the same context: a change in propensity. Or its own earlier tokens may lead it into contexts where the behavior is common: a change in reach. For a given event, teacher-forcing the predecessor on the successor's trajectories splits the counted change exactly into these two changes. Across released transitions, from instruction tuning to reinforcement learning from a base model, the split varies widely, and the two changes can have opposite signs: Olmo-3-7B-RL-Zero-Code writes paragraph-opening connectives 4.7 times as often as its predecessor, yet at the successor's own contexts the predecessor assigns them slightly higher probability. The split also accounts for what two interventions do. Restoring the predecessor's probability of the behavior at every position (a local intervention) moves the behavior back toward the predecessor's rate on the transition where propensity dominates; on the two where reach dominates, giving the successor the predecessor's first paragraph (a context intervention) moves it back, and the local intervention does not. For response termination, the local intervention changes which contexts the successor reaches, and this induced change in reach cancels the change in propensity. We recommend reporting the split beside raw counts; it needs only the two checkpoints.

1 INTRODUCTION

Post-training behavioral changes are commonly measured by counting features in each model's own trajectories (Juzek & Ward, 2025b; Juzek et al., 2026; Zeng et al., 2025; Guo et al., 2025). Each count measures the behavior a model realizes on trajectories it generated itself, so a change in the count moves with two things at once.

Suppose a post-trained agent verifies far more often than its predecessor. It may have become more likely to verify at the same debugging state, or it may reach such states more often. Both follow from the same weight update. The claim that the agent became more likely to verify at comparable states concerns only the first. The first is a change in what the model does, its *propensity*; the second is a change in where it goes, its *reach*. Both are defined relative to the behavior counted: a change in what the model does at an earlier position becomes a change in where it goes for a behavior counted later.

The distinction matters for control. A change in propensity sits at the positions where the behavior occurs; a change in reach sits in the earlier choices that lead to those positions, and over long trajectories those choices accumulate.

The same question arises for any behavior counted on a model's own trajectories. Teacher-forcing the predecessor on the successor's stored trajectories (Williams & Zipser, 1989; Lin et al., 2024) gives the predecessor's rate at the successor's contexts. That rate splits the counted change exactly into a change in reach and a change in propensity, a decomposition with the form of Kitagawa (1955) and Oaxaca (1973) (Figure 1a; Section 3). Unlike the covariate shifts such decompositions usually

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

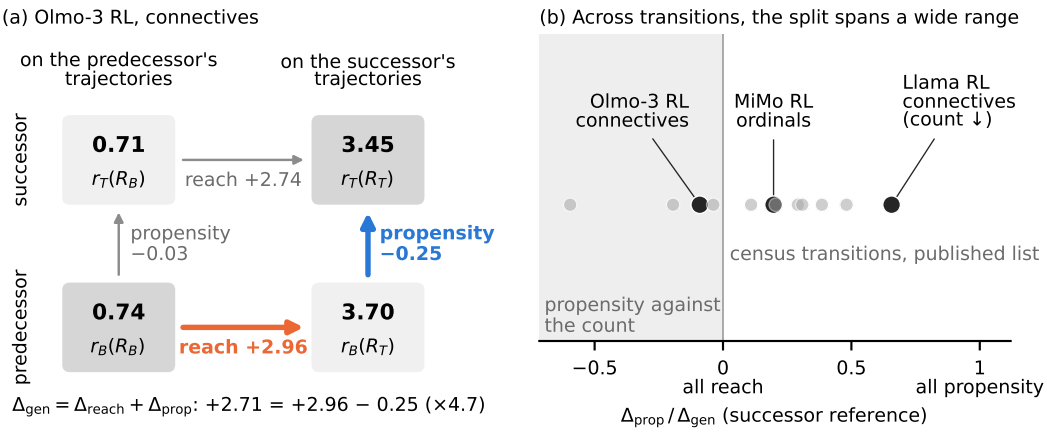


Figure 1: A counted change combines a change in reach and a change in propensity. **(a)** The four rates for the Olmo-3 RL transition's connectives: each model teacher-forced on each model's trajectories. The counted change compares the two models on their own trajectories; the bottom row and then the right column split it into the change in reach (whose trajectories) and the change in propensity (which model), and the thin path around the other two sides is the predecessor-reference split. Reach carries the whole rise, and propensity moves against it on either path. **(b)** $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ for each transition in the census, on the published excess-vocabulary list of Kobak et al. (2025), whose counted change has an interval excluding zero (model-written prompts), and for three case studies: from nearly all reach to mostly propensity, and below zero where propensity moves against the counted change.

address, the contexts here are produced by the very models being compared. We apply the split to a census of ten released instruction-tuning and reinforcement-learning transitions on the published excess-vocabulary list of Kobak et al. (2025), and to connective and ordinal classes and response termination in case studies. How a change divides varies widely (Figure 2 and Section 4.3).

The split also accounts for what two interventions do. The local intervention, which restores the predecessor's probability of the behavior at every position, moves the behavior back on the transition where propensity carries the change. The context intervention, which gives the successor the predecessor's first paragraph, moves the behavior further than the local intervention on the two transitions where reach carries it, and also moves it on the Llama transition, whose change after the first paragraph is carried by reach (Section 5). For response termination, the local intervention changes which contexts the successor reaches, and this induced change in reach cancels the change in propensity.

A counted change measures what a model writes on its own trajectories. How the successor differs at comparable contexts is a second measurement on the same stored trajectories, and it needs only the two checkpoints. We recommend reporting it alongside the count (Section 6).

2 Background and Related Work

Measuring post-training behavior. Lexical overuse is documented across model families (Kobak et al., 2025; Juzek & Ward, 2025a;b; Liang et al., 2024; Paech et al., 2026), and response length and reflection markers are standard diagnostics for policies trained with verifiable rewards (Zeng et al., 2025; Liu et al., 2025; Guo et al., 2025). Juzek & Ward (2025b) and Juzek et al. (2026) go further, attributing part of the overuse to the preference stage by comparing human, base, and instruction-tuned text. Corpus studies track these words in scientific writing over time, and model comparisons count them in each model's own trajectories. A difference between two models' trajectories includes both a change in reach and a change in propensity, and the stage attribution asks which stage of training produced it.

Created or amplified. A parallel debate asks whether post-training creates behaviors or amplifies ones already present. Alignment has been argued to be largely superficial (Zhou et al., 2023; Lin et al., 2024). Reinforcement learning has been found to leave the problems it solves within the

base model's sampling distribution (Yue et al., 2025). Base models already produce self-reflection before reinforcement learning (Liu et al., 2025); verification and backtracking in the base model predict how much it gains from it (Gandhi et al., 2025); and reinforcement learning amplifies a mode learned in pretraining (Zhao et al., 2025). The debate is conducted over answers, sampled behaviors, and training outcomes. The decomposition gives one of its questions a per-context form: how much of a counted change the predecessor already produces at the successor's own contexts.

Comparing models at shared contexts. Evaluating a base model on its aligned successor's trajectories locates where alignment shifts token distributions (Lin et al., 2024), and safety alignment has been found to shift mainly the first few tokens of a response (Qi et al., 2025). For reinforcement learning with verifiable rewards, token-level analyses find the shift concentrated at high-entropy positions (Wang et al., 2025) and sparse (Meng et al., 2026), and Meng et al. (2026) show that swapping a small number of tokens between base and reinforcement-learning trajectories recovers or removes the gain. Three of these studies (Lin et al., 2024; Wang et al., 2025; Meng et al., 2026) compute the predecessor's behavior at the successor's contexts, one bracket of our identity, to ask which tokens and positions change. We use the same operation to ask how a counted behavioral change should be interpreted. Our local intervention (Section 5) is related to the token swaps of Meng et al. (2026). Where these studies locate the decisions that route a response, the split measures how much of a downstream behavior's change is inherited through them.

Self-generated contexts. That a model's own trajectories determine the contexts in which it is later evaluated is familiar from exposure bias (Bengio et al., 2015; Ranzato et al., 2016; Schmidt, 2019), which treats the resulting train–test gap as a problem to correct in training, from imitation learning (Ross et al., 2011), and from the performance difference lemma, which evaluates a policy change under the new policy's state distribution (Kakade & Langford, 2002). Nie (2026) argues that the distribution of states on which post-training applies supervision can matter as much as the objective.

Decompositions. Splitting a difference between two populations into a composition part and a rate part is standard in demography and economics (Kitagawa, 1955; Oaxaca, 1973; Blinder, 1973), and work on distribution shift attributes a model's performance change to covariate and conditional shifts (Zhang et al., 2023; Cai et al., 2026), with nonlinear attributions shown to misassign between the two (Quintero et al., 2025). These decompositions take the covariate distribution as given and study an observed outcome or a predictive model's performance. In causal-mediation terms, with the model as the treatment and the prefix as the mediator, the change in propensity under the successor reference is the total direct effect, the change in reach is the pure indirect effect, and the difference between the two references' changes in propensity (successor minus predecessor) is the mediated interaction (Robins & Greenland, 1992; Pearl, 2001; VanderWeele, 2014); mediation analysis has also been used to probe language models by intervening on internal activations (Vig et al., 2020). Here the contexts are produced by the models being compared, the quantity is a behavior frequency that researchers report, and the cross-world rate $r_B(R_T)$, which mediation analysis usually has to identify under assumptions, is computed directly by teacher forcing, without identification assumptions. The decomposition is an exact identity on a chosen reference set of trajectories (Section 3.1). Appendix A reviews further token-level analyses of post-training.

3 Decomposing Behavioral Change on Self-Generated Trajectories

3.1 Counted Change, Propensity and Reach

Fix an event E, a set of token ids (for example, paragraph-opening connectives such as "First" or "However"); its probability at a context is the probability a model assigns to the set. Let $r_M(R_N)$ be the expected rate per 1,000 tokens that model M assigns to E when teacher-forced on trajectories generated by model N. The counted change is then $\Delta_{\text{gen}} = r_T(R_T) - r_B(R_B)$, and it decomposes exactly:

$$\Delta_{\text{gen}} \;=\; \underbrace{r_B(R_T) - r_B(R_B)}_{\text{change in reach } \Delta_{\text{reach}}} \;+\; \underbrace{r_T(R_T) - r_B(R_T)}_{\text{change in propensity } \Delta_{\text{prop}}}. \tag{1}$$

The change in reach measures what the predecessor would already have done at the successor's contexts. The change in propensity measures how the successor's probability differs from the predecessor's at matched contexts: the successor's own contexts, at which both models are evaluated. In

the verification example of Section 1, the change in propensity asks whether the agent verifies more at the same debugging state. The change in reach asks whether it reaches such states more often.

Each is an average over scored positions (Section 3.2): a negative change in propensity is a lower average over the successor's scored positions, and individual contexts can move either way. With the event's expected count over a fixed horizon, the generation cap, as the return, the identity is the split of the performance difference lemma (Kakade & Langford, 2002): summed over the successor's trajectories, the immediate-reward part of the predecessor's advantage gives the change in propensity, and its continuation-value part gives the change in reach; for rates per 1,000 tokens, the change in reach also absorbs differences in response length.

We summarize how a change divides by the *signed ratio* $\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$, stated with its reference. Between 0 and 1, both changes move with the counted change and the ratio is the fraction carried by propensity; below 0, propensity moves against the counted change; above 1, reach does. The ratio is reported where the counted change's interval excludes zero.

A change in propensity at another position, such as where a paragraph ends, enters the change in reach of an event counted after it (Appendix G.2). The change in reach counts any difference in the preceding context, including a difference one token earlier. The decomposition locates a change relative to the counted position, so the split is relative to the event.

The change in propensity in Equation (1) uses the successor's trajectories. Evaluating it on the predecessor's gives a second reference (Appendix B). We report the successor reference throughout, because the question concerns the successor's own trajectories.

3.2 Estimation on Stored Trajectories

We generate trajectories from both models, store the exact sampled token ids, and teacher-force both models on the same stored ids, evaluating each model's event probability at every scored position. Each teacher-forced distribution is put through the same decoding transform the trajectories were drawn under—temperature 0.7, truncation to the top 0.95 of the mass, renormalization. The transform makes each model's expected count on its own trajectories match its sampled count in expectation (Appendix C.1). A scored position is a token position at which the event's probability is evaluated; the scoring window is the set of scored positions (the first 384 tokens unless stated). Intervals are 95% bootstrap intervals from two thousand draws, paired on prompts. A same-model comparison gives a floor for the change in reach (Appendix C).

Three conventions apply throughout. The split is relative to the event counted and to the reference trajectories (the successor's, unless stated), and to how rates are aggregated (ratios of sums). A direction is stated only where its interval excludes zero. Each rate is computed on the scoring window stated with it, and the split depends on that window (Section 4.3).

4 Propensity and Reach Divide Differently Across Released Transitions

When a behavioral frequency changes after post-training, how much of that change persists at matched contexts?

4.1 Experimental Setup

We study twenty released transitions from the Qwen2.5 (Qwen et al., 2024), Llama-3.1 (Grattafiori et al., 2024), Mistral (Jiang et al., 2023), Gemma-2 (Gemma Team et al., 2024), OLMo-2 (Team OLMo et al., 2025), Olmo-3 (Team Olmo et al., 2025), MiMo (LLM-Core Xiaomi et al., 2025) and DeepSeek-Math (Shao et al., 2024) families, with SimpleRL-Zoo (Zeng et al., 2025) and Tülu-3 (Lambert et al., 2025) successors, spanning instruction tuning, supervised fine-tuning followed by reinforcement learning, and reinforcement learning from a base model (Table 1; Table 10). Every census transition is evaluated on a model-written population of 600 first user turns from UltraChat (Ding et al., 2023), whose user side is itself model-generated, and a human-written population of 600 first user turns written by people, from Chatbot Arena conversations (Chiang et al., 2024), dedu-

Table 1: Transitions in the main text. Rows 1–10 form the census (Section 4.2); rows 10–12 are the case studies (Section 4.3). Eight further transitions are listed in Appendix D. Names are abbreviated: a successor is the name shown followed by the suffix (row 1: Qwen2.5-1.5B-Instruct), except Mistral-7B-Instruct-v0.3 (row 5) and OLMo-2-1124-7B-Instruct and OLMo-2-1124-13B-Instruct (rows 7–8); as predecessors, OLMo-2-7B, OLMo-2-13B, Olmo-3-7B and MiMo-7B are OLMo-2-1124-7B, OLMo-2-1124-13B, Olmo-3-1025-7B and MiMo-7B-Base (Table 10). Populations: M model-written, H human-written, Ma mathematics. Measurement: list = published excess-vocabulary list; conn. = connective class; ordinal = ordinal/sequential class.

	Predecessor → Successor	Pop.	Meas.		Predecessor → Successor	Pop.	Meas.
1	Qwen2.5-1.5B → -Instruct	M, H	list	7	OLMo-2-7B → -Instruct	M, H	list
2	Qwen2.5-7B → -Instruct	M, H	list	8	OLMo-2-13B → -Instruct	M, H	list
3	Qwen2.5-14B → -Instruct	M, H	list	9	Olmo-3-7B → -Instruct	M, H	list
4	Llama-3.1-8B → -Instruct	M, H	list	10	Olmo-3-7B → -RL-Zero-Code	M, H, Ma	list; conn.
5	Mistral-7B-v0.3 → -Instruct	M, H	list	11	MiMo-7B → -RL-Zero	Ma, M	ordinal
6	Gemma-2-9B → -IT	M, H	list	12	Llama-3.1-8B → -SimpleRL-Zoo	M, Ma, H	conn.

plicated against public post-training corpora (Appendix D.2). Where noted, we add a mathematics population of 400 problems.

For breadth, we count the published excess-vocabulary list of Kobak et al. (2025), with words such as "realm", "innovative", "remarkable" and "notable", mapped to single-token ids per tokenizer. For case studies requiring a concrete behavioral class, we use a 46-id connective class of paragraph-opening words ("First", "However", "Moreover"). The class's words were chosen on OLMo-2 generations and its token ids on Olmo-3. We also use a 24-id ordinal/sequential class ("first", "second", "next", "finally"). A control class of comparable word forms ("The", "This", "It") is evaluated beside each (Appendix C). Response termination provides a semantically different event.

Each model generates 9 trajectories per prompt (temperature 0.7, top-p 0.95, 384-token cap). The chat-template rendering puts the prompt inside the successor's chat template (for the Olmo-3 transitions, the Olmo-3-7B-Instruct template; Appendix D.3); the bare rendering gives the predecessor the prompt text alone. In the census, the change in propensity is evaluated under the chat-template rendering, where both models see an identical prefix, and quantities that use the predecessor's own trajectories are reported under both renderings where both exist (Section 4.4). Each case study evaluates both models under one rendering (Appendix D.3).

4.2 Across Released Model Transitions

Figure 2 shows the signed ratio for the published excess-vocabulary list across the census. Where both components move with the count, the signed ratio runs from about a tenth to about half. On two rising transitions per population, the signed ratio is below zero. On every rising transition, human-written prompts put the signed ratio on the same side of 0 and 1 as model-written prompts. At the successor's own contexts, the predecessor already assigns at least three-quarters of the successor's rate on 8 of the 10 transitions on model-written prompts and at least half on all 10 (Appendix F.5).

4.3 The Same Counted Change Can Reflect Very Different Changes in Propensity

Appendix G sets four cases side by side (Table 23) and decomposes further case-study cells (Tables 25 and 28). For the Llama SimpleRL-Zoo transition the connective class falls, and most of that fall persists at matched contexts. Propensity carries 0.66 of the counted change on model-written prompts and 0.70 on mathematics prompts. The human-written population is in Appendix G. The change in propensity sits in the first 64 positions of the response, and after them the two models assign the class about the same probability (Appendix B.4). After the first paragraph, on responses that open with a prose paragraph, the change is carried by reach (Section 5.2). For this transition, a raw count would have been roughly right.

On the MiMo transition, an ordinal/sequential class evaluated on mathematics prompts rises by 1.42 per 1,000 tokens, and propensity carries only 0.20 of it. Read as a change in propensity, the count

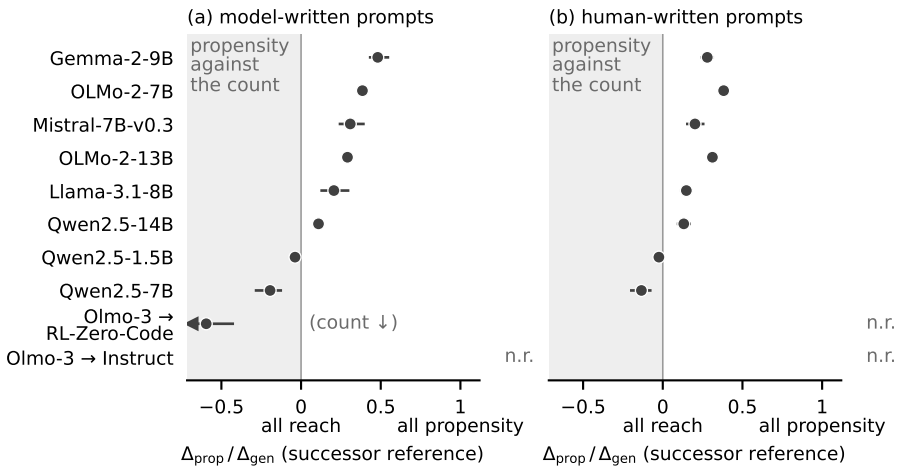


Figure 2: The signed ratio $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (successor reference) with 95% intervals for the published excess-vocabulary list across 10 transitions under the chat-template rendering: (a) model-written prompts, (b) human-written prompts. The axis and shading are those of Figure 1b; in the shaded region, propensity moves against the counted change. Rows are ordered by the model-written signed ratio. n.r. (right of the axis): the counted change's interval includes zero, and no signed ratio is shown. An arrowhead marks an interval that runs past the axis; (count ↓): the counted change falls. Full results in Appendix E.

would overstate it about fivefold. The rest of the rise is reach. Model-written prompts give the same signs (Appendix G).

The sharpest case is the Olmo-3 RL transition. The connective class rises 4.7-fold on model-written prompts, from 0.74 to 3.45 per 1,000 tokens, yet the change in propensity is negative: the change in reach alone exceeds the entire rise. Propensity moves against the count. The successor generates this class far more often while assigning it lower probability at the successor's own contexts. The change in propensity is already negative at positions before the response contains any connective (-0.34 per 1,000 tokens), so the reversal is present before the predecessor has any of the successor's connectives to continue (Appendix B.4). Where the successor's propensity rises is upstream of the connective: at the same contexts, on the predecessor's own first paragraphs, the successor ends the paragraph at 4.68 per 1,000 tokens against the predecessor's 0.96 (Appendix G.2), and at the first token it is more likely to open with a plan (Appendix G.1). For a connective counted one token after the break, that upstream change enters the change in reach, so the split points to the paragraph break rather than to the connective (Appendix I.3). The reversal also holds on human-written prompts, and it survives removing "First" (Appendix G.1). The class's words were chosen on OLMo-2 generations and its token ids on this lineage; on a class chosen on a held-out family, the sign discrepancy rests largely on one word (Appendix L.2).

On the Olmo-3 RL transition, response termination, a behavior with no word list, shows why the split can depend on the generation cap while the change in propensity stays negative at every cap. The chance of stopping depends strongly on the context: it is low in the middle of an answer, high where responses typically end, and low again inside a repetition loop. The generation cap decides how much of each of these contexts the scoring window contains, and the change in reach records that composition. At 384 tokens nearly every successor response reaches the cap, so reach carries almost all of a falling count. At 2,048 tokens, where most responses end on their own, both models' windows contain the positions where their responses end, reach is near zero, and propensity carries most of the fall. At 8,192 tokens the predecessor's responses that never stop, most of them near-repetitive, fill most of its own window, reach turns large, and the count becomes a rise. At every cap the change in propensity is negative: at the same context, the successor is less likely to stop (Table 23; Appendix H).

4.4 Is the Range an Artifact of Measurement Choices?

The range holds under four checks. Rendering: in the census, the change in propensity and the predecessor share at matched contexts, $r_B(R_T)/r_T(R_T)$, use only the successor's trajectories, so they do not depend on how the predecessor generates its own; the counted change and the change in reach do, and we report both renderings (Appendix F). Reference: the case studies keep their signs under the predecessor reference (Table 23), while on a few census rows the sign changes (Appendix B.4). Generation cap: at 2,048 tokens the regenerated census transitions keep their rise, their signed ratio stays on the same side of 0 and 1, and it is closer to zero, by interval, on six of the seven cells (bare rendering; Appendix F.3). Aggregation: with per-prompt averaged rates, the signed ratio still spans about zero to one half across the rising census transitions where both components move with the count (Appendix E.4).

The two references ask for the change in propensity on different context distributions, the successor's or the predecessor's. They differ only when the change depends on context and the two models visit different contexts. Under the average of the two references, the census range where both components move with the count becomes about 0.05 (sign uncertain) to 0.47, the signed ratio is below zero on one rising transition per population (on human-written prompts with an uncertain sign), and the two populations put it on the same side of 0 and 1 on every transition that rises in both (Appendix B.4).

5 What the Split Says About Intervening

5.1 Experimental Setup

Section 4 splits each change on fixed trajectories. Two questions follow for intervening while the model generates. First, does the split carry over online? The local intervention answers this: at every generated position with prefix h, it replaces the successor's probability of the event with the predecessor's, so the successor's next-token distribution becomes

$$\tilde{P}(x \mid h) \;=\; P_B(E \mid h)\, P_T(x \mid E, h) \;+\; \big(1 - P_B(E \mid h)\big)\, P_T(x \mid \neg E, h), \tag{2}$$

and the successor's relative distribution inside the event and outside it is preserved. It applies the logit arithmetic of decoding-time combinations of base and tuned models (Mitchell et al., 2024; Liu et al., 2024) to the event's probability alone. Because the successor under the local intervention assigns the predecessor's probability at every context where the operator is defined (Appendix J.3), its expected online rate is the predecessor's rate on the locally intervened arm's trajectories, $r_B(R_I)$, so, up to sampling,

$$y_I - y_T = -\Delta_{\text{prop}} + \big[r_B(R_I) - r_B(R_T)\big]. \tag{3}$$

The fixed-trajectory account gives the first term. The second is the induced change in reach, the change in reach that the local intervention itself induces. Second, when a change is carried by reach, does changing where the model goes move the behavior? The context intervention, which replaces the upstream paragraph, answers this: in the paragraph exchange, a model continues the first paragraph the other model wrote. The paragraph exchange applies to responses that open with a prose paragraph and is evaluated on the continuation after it (Appendix J.2). For response termination, the context intervention is a prefix transplant: the successor continues the predecessor's response from a fixed token position.

Both act on the distribution after the decoding transform. Figure 3 plots each arm's signed movement toward the predecessor, $m = (y_T - y_I)/(y_T - y_B)$, where y_I is the intervened arm's rate and y_T and y_B are the unmodified rates, each computed against the gap listed for it in Table 47: $m{=}1$ is the predecessor's rate, $m{>}1$ overshoots past it, $m{<}0$ moves away from it. In the paragraph exchange's mirror direction, where the predecessor continues the successor's paragraph, $\bar{m}$ is the fraction of the same gap by which the predecessor's rate moves toward the successor's ($\bar{m}{=}1$ is the successor's rate).

5.2 Intervention Results

Figure 3 shows the results. On the Olmo-3 RL transition, whose counted connective rise is carried almost entirely by the change in reach, supplying the predecessor's first paragraph returns $m = 0.42$ of the gap, while the local intervention moves the rate further from the predecessor ($m = -0.13$).

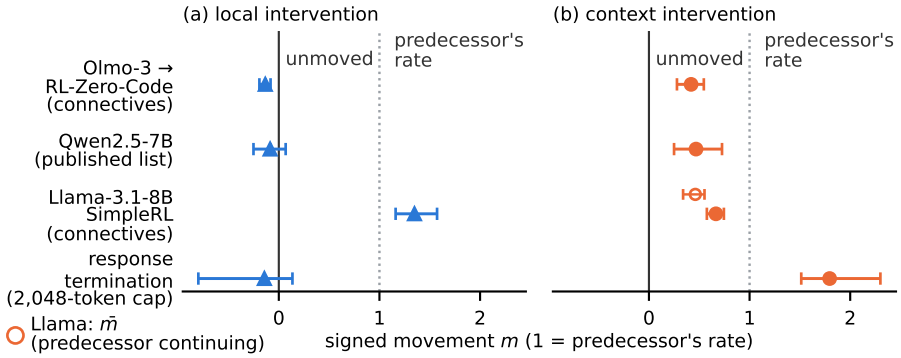


Figure 3: Signed movement toward the predecessor (m): (a) local interventions (triangles) and (b) context interventions (circles), on the same axis. On the Llama row of (b), the filled circle is the successor continuing the predecessor's paragraph and the hollow circle the predecessor continuing the successor's ($\bar{m}$, whose 1 is the successor's rate; Appendix K.4). Intervals are 95% bootstrap, prompt-paired. The Qwen circle's interval is over the prompts its contrast shares; the termination context marker pools three cut points. The arms differ in scoring window, subset and gap, listed in Table 47.

The change in propensity is negative on fixed trajectories, and online the locally intervened arm moves the rate in the direction of Equation (3)'s first term, $-\Delta_{\text{prop}}$. On the Qwen2.5-7B transition, whose change in propensity on the published list is negative against a rising counted change (Table 21), the paragraph exchange returns $m = 0.47$ of the gap, while the local intervention leaves the rate within sampling noise of the successor's, $m = -0.09$ [-0.25, 0.07]; the contrast between the two interventions, on the prompts both share, is 0.64 [0.23, 1.12] (Appendix K.5). Reach carries most of this change under either reference, and the sign of its change in propensity depends on the reference (-0.20 under the successor's, +0.30 under the predecessor's; Appendix B.4).

On the Llama transition, whose counted connective fall is carried by the change in propensity, the local intervention moves the rate by $m = 1.35$: the fixed-trajectory account gives 0.69, and the induced change in reach gives 0.67 (Appendix J.3). The paragraph exchange moves the continuation toward the paragraph's source in both directions: $m = 0.67$ when the successor continues the predecessor's paragraph and $\bar{m} = 0.46$ when the predecessor continues the successor's (Appendix K.4). On the continuation after the first paragraph, evaluated on the model-written prompts, the split gives reach most of the transition's fall (signed ratio 0.13), against 0.46 over the whole of the same responses and 0.66 over all responses (Appendix K.4).

Response termination at the 2,048-token cap shows that the local intervention changes the trajectories on which its own rate is computed. The local intervention sets the predecessor's stop-token mass wherever it is defined (Table 46), yet $m = -0.14$, with an interval that includes zero, and the locally intervened arm's mean length and probability of stopping before the cap stay within sampling noise of the successor's (Appendix K.2). The induced change in reach cancels Equation (3)'s first term: the predecessor's stop rate is 0.66 on the locally intervened arm's trajectories against 0.79 on the successor's (Table 46; Appendix J.3), offsetting the change in propensity of -0.12. The context intervention, a prefix transplant pooled over three cut points, overshoots ($m = 1.80$).

The context intervention moves the control class as well wherever the interval of that movement excludes zero, on Olmo-3 and Llama about as far as the counted class. On Olmo-3 it moves toward the model whose paragraph is supplied; on Llama and Qwen2.5-7B, whose two models' own control-class rates are within sampling noise of each other, it moves the control class as well. Both are consistent with an intervention on where the model goes (Appendix K).

6 Implications

Reporting. We recommend reporting the change in propensity beside any counted change.

Intervention. Equation (3) accounts for each local intervention's result. On the transitions we intervened on, the local intervention moved the change that propensity carries (Llama); the context intervention moved the changes that reach carries (Olmo-3 RL and Llama after the first paragraph), and on Qwen2.5-7B moved the rate further toward the predecessor than the local intervention did (paired contrast 0.64 [0.23, 1.12]); on response termination, Equation (3)'s induced change in reach offset $-\Delta_{\text{prop}}$. Token-level analyses find that a few decisions route the rest of a response (Wang et al., 2025; Meng et al., 2026; Kim & No, 2026), and safety alignment acts mainly on the first tokens of a response (Qi et al., 2025). A behavior downstream of such decisions changes through reach, as the Olmo-3 connective rise follows a change in where the first paragraph ends, evaluated on the predecessor's own first paragraphs (Appendix G.2).

Long generation caps and agents. The verification example of Section 1 is a setting where we expect the distinction to matter. Response termination shows the pattern within our data: as the generation cap grows, the counted change changes sign while the change in propensity keeps its sign (Section 4.3). Measuring agent trajectories this way is future work.

7 Limitations

Interpretation. The change in propensity changes sign between the two references on a few rows (Section 4.4). Matched contexts are written by the successor, and a base predecessor may partly imitate them in context; the change in propensity measures what it does not take up. The decomposition locates a change relative to the counted position on a given set of trajectories. Attributing the change to a part of training, or predicting an intervention's online effect, takes evidence beyond it.

Measurement choices. Quantities that use the predecessor's own trajectories depend on how it is prompted (Section 4.4), and part of the counted change and of the change in reach may reflect a register difference. In the census, base models are teacher-forced under the chat-template rendering, whose control tokens they may not have been trained on. The census is evaluated at the 384-token cap. The longer-cap check covers 4 transitions under the bare rendering. The model-written prompts may overlap the successors' training mixtures, which could inflate either component. The human-written population, deduplicated against public post-training corpora, is the check (Section 4.4). The published excess-vocabulary list was derived from biomedical abstracts, so the census measures that list's change.

Coverage. Each checkpoint is one training run, so intervals cover prompts and trajectories within that run. Each intervention is evaluated on a single transition per lineage, at the generation caps and references we evaluated. The results describe the transitions studied. A prevalence estimate would need transitions sampled from post-training at large.

8 Conclusion

On released transitions, the change in propensity and the change in reach differ in size and can differ in sign, so a count of what a model writes leaves open how it behaves at the same contexts. The split is exact for a chosen event and reference and needs only teacher-forcing passes of the two checkpoints on stored trajectories. Applied to an intervened model's own trajectories, the same decomposition measures how much an intervention changes the contexts it is evaluated on; this induced change in reach is why restoring the stop probability leaves response termination within sampling noise of the successor's. We recommend reporting the change in propensity beside every count.

AI use statement

We used generative AI tools, including Anthropic Claude/Claude Code and OpenAI ChatGPT, in several parts of the research workflow. These tools assisted with literature retrieval and discovery; brainstorming and refinement of the conceptual framing and hypotheses; feedback on methodology and experimental design; implementation and orchestration of experiments, including scoring, testing, analysis, and figure-generation code; maintenance and summarization of the research record; interpretation and stress-testing of experimental results and as annotators in two audits (Appendices F.4 and I.4); and drafting, restructuring, editing, and polishing portions of the manuscript.

We did not use generative AI to generate synthetic datasets or to prove mathematical claims. We reviewed all AI-assisted work, verified reported numerical results against the stored experimental artifacts, checked AI-assisted code through execution and testing, and checked literature claims and citations against their original sources. We take responsibility for the final content of this work, including all text, claims, code, analyses, and artifacts produced with the assistance of generative AI.

Ethics statement

This work uses publicly released models and publicly released datasets only. WildChat (Zhao et al., 2024), a released anonymized chat corpus, and WildJailbreak (Jiang et al., 2024), a released synthetic safety-training set, enter this work only as part of a public supervised fine-tuning mixture and as prompts drawn from it, and the human-written prompts are first user turns from a public release of Chatbot Arena conversations (`lmsys/lmsys-arena-human-preference-55k`; Chiang et al., 2024). No human subjects took part and no new data were collected. We foresee no harm beyond that already posed by the released models; the findings bear on how stylistic and structural artifacts of post-training are measured and interpreted.

Reproducibility statement

The two conventions any reproduction requires—which trajectory reference a change in propensity is evaluated against and which resampling unit an interval uses—are fixed in Sections 3.1 and 3.2. The anonymized supplementary material contains the scorer code, the experiment framework, the macro and figure generators, their tests, the result artifacts needed to regenerate every number, table and figure in the paper, and the regeneration scripts with a README. Computing the decomposition from stored teacher-forced probabilities needs only CPU; the GPU pods for this work, earlier passes included, used at least 189 GPU-hours on NVIDIA A100-SXM4-80GB and NVIDIA A100 80GB PCIe GPUs, at least 24.33 of them on the latter; some earlier pods, among them the one that ran the Olmo-3 paragraph exchange of Appendix I.1, stored no time, and some earlier pods whose hours are not counted ran on NVIDIA RTX A6000 GPUs.

References

Ömer Faruk Akgül, Rajgopal Kannan, Willie Neiswanger, and Viktor Prasanna. Rethinking RL for LLM reasoning: It's sparse policy selection, not capability learning. *arXiv preprint arXiv:2605.06241*, 2026.

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

Alan S. Blinder. Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, 8(4):436–455, 1973. doi: 10.2307/144855.

Kailash Budhathoki, Dominik Janzing, Patrick Blöbaum, and Hoiyi Ng. Why did the distribution change? In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130, pp. 1666–1674. PMLR, 2021.

Tiffany Tianhui Cai, Hongseok Namkoong, and Steve Yadlowsky. Diagnosing model performance under distribution shift. *Operations Research*, 74(2):898–916, 2026. doi: 10.1287/opre.2023.0217.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Angelopoulos, Tianle Li, Dacheng Li, Banghua Zhu, Hao Zhang, et al. Chatbot Arena: An open platform for evaluating LLMs by human preference. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 8359–8388. PMLR, 2024. URL `https://proceedings.mlr.press/v235/chiang24b.html`.

John DiNardo, Nicole M. Fortin, and Thomas Lemieux. Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044, 1996. doi: 10.2307/2171954.

Ning Ding, Yulin Chen, Bokai Xu, Yujia Qin, Shengding Hu, Zhiyuan Liu, Maosong Sun, and Bowen Zhou. Enhancing chat language models by scaling high-quality instructional conversations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 3029–3051, Singapore, 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.emnlp-main.183. URL https://aclanthology.org/2023.emnlp-main.183/.

Kanishk Gandhi, Ayush Chakravarthy, Anikait Singh, Nathan Lile, and Noah D. Goodman. Cognitive behaviors that enable self-improving reasoners, or, four habits of highly effective STaRs. In *Second Conference on Language Modeling*, 2025. URL https://openreview.net/forum?id=QGJ9ttXLTy.

Gemma Team, Morgane Rivière, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, et al. Gemma 2: Improving open language models at a practical size. *arXiv preprint arXiv:2408.00118*, 2024.

Aaron Grattafiori et al. The Llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature*, 645(8081):633–638, 2025. doi: 10.1038/s41586-025-09422-z.

Kexin Huang, Haoming Meng, Junkang Wu, Jinda Lu, Chiyu Ma, Ziqian Chen, Xue Wang, Bolin Ding, et al. Beyond magnitude: Leveraging direction of RLVR updates for LLM reasoning. In *The Fourteenth International Conference on Learning Representations*, pp. 61216–61239, 2026. URL https://openreview.net/forum?id=r6Pw3RiMYL.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, et al. Mistral 7B. *arXiv preprint arXiv:2310.06825*, 2023.

Liwei Jiang, Kavel Rao, Seungju Han, Allyson Ettinger, Faeze Brahman, Sachin Kumar, Niloofar Mireshghallah, Ximing Lu, et al. WildTeaming at scale: From in-the-wild jailbreaks to (adversarially) safer language models. In *Advances in Neural Information Processing Systems 37*, pp. 47094–47165, 2024. doi: 10.52202/079017-1493.

Thomas Stephan Juzek and Zina B. Ward. Why does ChatGPT "delve" so much? Exploring the sources of lexical overrepresentation in large language models. In *Proceedings of the 31st International Conference on Computational Linguistics*, pp. 6397–6411, Abu Dhabi, UAE, 2025a. Association for Computational Linguistics. URL https://aclanthology.org/2025.coling-main.426/.

Thomas Stephan Juzek and Zina B. Ward. Word overuse and alignment in large language models: The influence of learning from human feedback. *arXiv preprint arXiv:2508.01930*, 2025b.

Thomas Stephan Juzek, Xiaoyang Ming, and Jose A. Hernandez. Fully automated identification of lexical alignment and preference-stage shifts in large language models. In *Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026)*, pp. 6116–6131. European Language Resources Association (ELRA), 2026. doi: 10.63317/4ut7ammh7z3h.

Sham M. Kakade and John Langford. Approximately optimal approximate reinforcement learning. In *Proceedings of the Nineteenth International Conference on Machine Learning*, pp. 267–274, 2002.

Soeun Kim and Albert No. Where rollouts begin: Low-load, high-leverage first-token diversification for RLVR. *arXiv preprint arXiv:2605.28295*, 2026.

Evelyn M. Kitagawa. Components of a difference between two rates. *Journal of the American Statistical Association*, 50(272):1168–1194, 1955. doi: 10.1080/01621459.1955.10501299.

Dmitry Kobak, Rita González-Márquez, Ágnes Horvát, and Jan Lause. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances*, 11 (27):eadt3813, 2025. doi: 10.1126/sciadv.adt3813.

Nathan Lambert, Jacob Morrison, Valentina Pyatkin, Shengyi Huang, Hamish Ivison, Faeze Brahman, Lester James V. Miranda, Alisa Liu, et al. Tulu 3: Pushing frontiers in open language model post-training. In *Second Conference on Language Modeling*, 2025. URL https://openreview.net/forum?id=i1uGbfHHpH.

Ming Li, Zhengyuan Yang, Xiyao Wang, Dianqi Li, Linjie Li, Kevin Lin, Tianyi Zhou, and Lijuan Wang. What makes reasoning models different? Follow the reasoning leader for efficient decoding. *arXiv preprint arXiv:2506.06998*, 2025.

Weixin Liang, Zachary Izzo, Yaohui Zhang, Haley Lepp, Hancheng Cao, Xuandong Zhao, Lingjiao Chen, Haotian Ye, et al. Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 29575–29620. PMLR, 2024.

Bill Y. Lin, Abhilasha Ravichander, Ximing Lu, Nouha Dziri, Melanie Sclar, Khyathi Raghavi Chandu, Chandra Bhagavatula, and Yejin Choi. The unlocking spell on base LLMs: Rethinking alignment via in-context learning. In *The Twelfth International Conference on Learning Representations*, pp. 24907–24933, 2024. URL https://openreview.net/forum?id=wxJ0eXwwda.

Alisa Liu, Xiaochuang Han, Yizhong Wang, Yulia Tsvetkov, Yejin Choi, and Noah A. Smith. Tuning language models by proxy. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=dribhnhm1i.

Zichen Liu, Changyu Chen, Wenjun Li, Penghui Qi, Tianyu Pang, Chao Du, Wee Sun Lee, and Min Lin. Understanding R1-Zero-like training: A critical perspective. In *Second Conference on Language Modeling*, 2025. URL https://openreview.net/forum?id=5PAF7PAY2Y.

LLM-Core Xiaomi, Bingquan Xia, Bowen Shen, Cici, Dawei Zhu, Di Zhang, Gang Wang, Hailin Zhang, et al. MiMo: Unlocking the reasoning potential of language model – from pretraining to posttraining. *arXiv preprint arXiv:2505.07608*, 2025.

Haoming Meng, Kexin Huang, Shaohang Wei, Chiyu Ma, Shuo Yang, Xue Wang, Guoyin Wang, Bolin Ding, et al. Sparse but critical: A token-level analysis of distributional shifts in RLVR fine-tuning of LLMs. In *The Fourteenth International Conference on Learning Representations*, pp. 124171–124210, 2026. URL https://openreview.net/forum?id=8vWIXno8LW.

Eric Mitchell, Rafael Rafailov, Archit Sharma, Chelsea Finn, and Christopher D. Manning. An emulator for fine-tuning large language models using small language models. In *The Twelfth International Conference on Learning Representations*, pp. 13229–13244, 2024. URL https://openreview.net/forum?id=Eo7kv0sllr.

Dong Nie. Post-training is about states, not tokens: A state distribution view of SFT, RL, and on-policy distillation. *arXiv preprint arXiv:2605.22731*, 2026.

Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709, 1973. doi: 10.2307/2525981.

Samuel Paech, Allen Roush, Judah Goldfeder, and Ravid Shwartz-Ziv. Antislop: A comprehensive framework for identifying and eliminating repetitive patterns in language models. In *The Fourteenth International Conference on Learning Representations*, pp. 42490–42525, 2026. URL https://openreview.net/forum?id=gLcyM1khyp.

Judea Pearl. Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI 2001)*, pp. 411–420. Morgan Kaufmann, 2001.

Doina Precup, Richard S. Sutton, and Satinder Singh. Eligibility traces for off-policy policy evaluation. In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000), Stanford University, Stanford, CA, USA, June 29–July 2, 2000*, pp. 759–766. Morgan Kaufmann, 2000.

Xiangyu Qi, Ashwinee Panda, Kaifeng Lyu, Xiao Ma, Subhrajit Roy, Ahmad Beirami, Prateek Mittal, and Peter Henderson. Safety alignment should be made more than just a few tokens deep. In *The Thirteenth International Conference on Learning Representations*, pp. 54911–54941, 2025. URL https://openreview.net/forum?id=6Mxhg9PtDE.

Manuel Quintero, William T. Stephenson, Advik Shreekumar, and Tamara Broderick. Common functional decompositions can mis-attribute differences in outcomes between populations. *arXiv preprint arXiv:2504.16864*, 2025. 2nd Workshop on Navigating and Addressing Data Problems for Foundation Models (DATA-FM), ICLR 2025.

Qwen, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, et al. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2024.

Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. Sequence level training with recurrent neural networks. In *International Conference on Learning Representations (ICLR 2016)*, 2016.

James M. Robins and Sander Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2):143–155, 1992. doi: 10.1097/00001648-199203000-00013.

Stéphane Ross, Geoffrey J. Gordon, and J. Andrew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15, pp. 627–635. PMLR, 2011.

Florian Schmidt. Generalization in generation: A closer look at exposure bias. In *Proceedings of the 3rd Workshop on Neural Generation and Translation*, pp. 157–167, 2019. doi: 10.18653/v1/d19-5616.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, et al. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Team Olmo, Allyson Ettinger, Amanda Bertsch, Bailey Kuehl, David Graham, David Heineman, Dirk Groeneveld, Faeze Brahman, et al. Olmo 3. *arXiv preprint arXiv:2512.13961*, 2025.

Team OLMo, Pete Walsh, Luca Soldaini, Dirk Groeneveld, Kyle Lo, Shane Arora, Akshita Bhagia, Yuling Gu, et al. 2 OLMo 2 Furious. *arXiv preprint arXiv:2501.00656v3*, 2025.

Tyler J. VanderWeele. A unification of mediation and interaction: A 4-way decomposition. *Epidemiology*, 25(5):749–761, 2014. doi: 10.1097/EDE.0000000000000121.

Jesse Vig, Sebastian Gehrmann, Yonatan Belinkov, Sharon Qian, Daniel Nevo, Yaron Singer, and Stuart M. Shieber. Investigating gender bias in language models using causal mediation analysis. In *Advances in Neural Information Processing Systems 33*, pp. 12388–12401, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/92650b2e92217715fe312e6fa7b90d82-Abstract.html.

Shenzhi Wang, Le Yu, Chang Gao, Chujie Zheng, Shixuan Liu, Rui Lu, Kai Dang, Xiong-Hui Chen, et al. Beyond the 80/20 rule: High-entropy minority tokens drive effective reinforcement learning for LLM reasoning. In *Advances in Neural Information Processing Systems 38*, pp. 127985–128019, 2025. doi: 10.52202/085713-3850.

Ronald J. Williams and David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1(2):270–280, 1989. doi: 10.1162/neco.1989.1.2.270.

Yang Yue, Zhiqi Chen, Rui Lu, Andrew Zhao, Zhaokai Wang, Yang Yue, Shiji Song, and Gao Huang. Does reinforcement learning really incentivize reasoning capacity in LLMs beyond the base model? In *Advances in Neural Information Processing Systems 38*, pp. 64304–64339, 2025. doi: 10.52202/085713-1933.

Weihao Zeng, Yuzhen Huang, Qian Liu, Wei Liu, Keqing He, Zejun Ma, and Junxian He. SimpleRL-Zoo: Investigating and taming zero reinforcement learning for open base models in the wild. In *Second Conference on Language Modeling*, 2025. URL https://openreview.net/forum?id=vSMCBUgrQj.

Haoran Zhang, Harvineet Singh, Marzyeh Ghassemi, and Shalmali Joshi. "Why did the model fail?": Attributing model performance changes to distribution shifts. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 41550–41578. PMLR, 2023.

Rosie Zhao, Alexandru Meterez, Sham M. Kakade, Cengiz Pehlevan, Samy Jelassi, and Eran Malach. Echo chamber: RL post-training amplifies behaviors learned in pretraining. In *Second Conference on Language Modeling*, 2025. URL https://openreview.net/forum?id=dp4KWuSDzj.

Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. WildChat: 1M ChatGPT interaction logs in the wild. In *The Twelfth International Conference on Learning Representations*, pp. 34590–34605, 2024. URL https://openreview.net/forum?id=Bl8u7ZRlbM.

Chunting Zhou, Pengfei Liu, Puxin Xu, Srinivasan Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, et al. LIMA: Less is more for alignment. In *Advances in Neural Information Processing Systems 36*, pp. 55006–55021, 2023. doi: 10.52202/075280-2400.

Appendix Conventions

Two conventions hold throughout. Rates and their intervals are printed to two decimal places, shares and movements to two, and folds to one; where an interval endpoint would round to zero at two decimals, the point and both endpoints print to three. Every derived quantity is computed from unrounded values, so printed parts need not add up exactly in the last digit. Intervals are 95% bootstrap intervals whose default resampling unit is the prompt (prompt-paired; Appendix C.4); a caption names the resampling unit only where it is another one.

A Extended Related Work

Further token-level analyses of reinforcement learning with verifiable rewards use the signed log-probability difference between base and trained models to identify the direction of the update (Huang et al., 2026), argue that reinforcement learning acts as a sparse selection among tokens the base model already ranks highly (Akgül et al., 2026), and find that a response's first token matters mainly through where it routes the rest of the response (Kim & No, 2026). Li et al. (2025) show that divergence between distilled reasoning models and non-reasoning models concentrates at sentence-initial positions. Outside language models, Budhathoki et al. (2021) attribute a change in a distribution to changes in its causal mechanisms. Each localizes where or why two distributions differ; none of the language-model analyses separates a counted behavior's change into a part carried by the contexts reached and a part at matched contexts.

The counterfactual rate $r_B(R_T)$ can also be approached by reweighting rather than teacher forcing: DiNardo et al. (1996) reweight one population's covariate distribution to another's, and importance-weighted off-policy estimators reweight trajectories by policy ratios (Precup et al., 2000). Teacher forcing evaluates the predecessor's probabilities at the successor's contexts directly and needs no weights.

B Decomposition Details and Alternative References

B.1 Exact Identities and the Predecessor-Reference Quantity

Let $r_B(R_B)$, $r_B(R_T)$, $r_T(R_B)$ and $r_T(R_T)$ denote the four rates. The counted change decomposes as

$$\Delta_{\text{gen}} \;=\; r_T(R_T) - r_B(R_B) \;=\; \underbrace{\big(r_B(R_T) - r_B(R_B)\big)}_{\Delta_{\text{reach}}} \;+\; \underbrace{\big(r_T(R_T) - r_B(R_T)\big)}_{\Delta_{\text{prop}}}, \tag{4}$$

where Δ_{reach} is the change in reach and Δ_{prop} is the change in propensity. The same counted change also decomposes under the predecessor's trajectories:

$$\Delta_{\text{gen}} = \underbrace{(r_T(R_T) - r_T(R_B))}_{r_T(R_T) - r_T(R_B)} + \underbrace{(r_T(R_B) - r_B(R_B))}_{\Delta^B_{\text{prop}}}, \tag{5}$$

where Δ^B_{prop} is the predecessor-reference change in propensity. The quantity $r_T(R_T) - r_T(R_B)$ has no canonical name in this paper and appears only in this appendix. Subtracting Equation (4) from Equation (5) yields

$$\Delta_{\text{prop}} - \Delta^B_{\text{prop}} = (r_T(R_T) - r_B(R_T)) - (r_T(R_B) - r_B(R_B)), \tag{6}$$

an algebraic residual whose magnitude measures how far the two changes in propensity diverge. We call it the interaction term (Equation (6) is an identity, not a third mechanism); Appendix B.4 gives its size across the census and what it suggests about where the two models' differences lie.

Relation to the main-text decomposition. The main text uses successor trajectories as the reference throughout, because the paper asks how to interpret behavior observed in the successor's own trajectories. The predecessor-reference change in propensity Δ^B_{prop} provides a robustness check: when Δ_{prop} and Δ^B_{prop} agree in sign and approximate magnitude, the change in propensity is insensitive to the choice of reference. When they disagree, the interpretation depends on which trajectories the evaluation is conducted on, and we report both.

B.2 Both References, Taken Separately, Across the Census

The census of Section 4.2 uses successor trajectories as the primary reference. Table 2 prints the reach share, one minus the signed ratio under the same reference ($1 - \Delta_{\text{prop}}/\Delta_{\text{gen}}$ or $1 - \Delta^B_{\text{prop}}/\Delta_{\text{gen}}$), under each reference separately, for every resolved transition in the census, across four population-by-rendering combinations.

The upper block counts how many of the rising transitions with a printed reach share have that share's interval entirely above one half under both references, under exactly one, and under the mean of the two references. The lower block prints the per-transition reach shares under each reference and their mean.

B.3 Reference Audit: Successor Versus Predecessor Trajectories

How far do the main-text conclusions depend on the reference? The main text reports the change in propensity computed on successor trajectories. Table 3 prints it under both references side by side for every row that enters the main text.

The census. Across the census under the chat-template rendering, the two references agree in the sign of the change in propensity except on the rows listed in Appendix B.4; under the bare rendering Qwen2.5-14B also changes sign, for the reason given in Appendix F.5. The shares differ in size between references (Table 2).

The case-study transitions. Table 3 prints the change in propensity under each reference for the four main-text case studies. The Olmo-3 RL transition's connective-class change in propensity is negative under both references, with the successor-reference value larger in magnitude. The MiMo ordinal/sequential class is positive under both references on both populations, with similar magnitudes. The Llama SimpleRL connective class is negative under both references, with the successor-reference value larger in magnitude. Response termination's change in propensity is negative under both references on the model-written population, with the successor-reference value smaller in magnitude.

On each case-study row in Table 3 the two changes in propensity agree in sign; the row of the appendix-only Olmo-3-7B-RL-Zero-Mix transition is the one whose two changes in propensity differ in sign. Their magnitudes differ (the predecessor-reference value is smaller for the Olmo-3 connective class and larger for response termination), so the precise share attributed to each component depends on the reference. On the census population-by-rendering grids (Appendix F.5), two transitions per grid have the two changes in propensity on opposite sides of zero (Table 2).

Table 2: The reach share of each census transition's counted change, computed under each reference separately. *counted*: rising transitions whose share is printed; *both*: the share's interval sits above one half under successor and predecessor trajectories alike; *one only*: above one half under one reference but not the other; *avg. (Appendices E and F)* and *avg. (recomputed)* are the count under the mean of the two references on the census's own draws (Figure 4 and Appendices F.1 and F.2) and as recomputed on the draws used here, which differ on rows whose interval sits on one half; *opp. sign* counts transitions, resolved or not, whose two changes in propensity have intervals excluding zero with opposite signs. In the lower block the Olmo-3 RL transition's row for model-written prompts under the chat-template rendering has a resolved *fall* on this list, so it carries shares but enters no count. A dash marks a cell the floor rule of Appendix C.4 withholds. Every interval here is from a resample drawn for this table, not from the census's own draws, which Table 7 uses.

Prompts, rendering	counted	both	one only	avg. (Appendices E and F)	avg. (recomputed)	opp. sign
Model-written, bare	6	3	3	5	4	2
Model-written, template	8	4	4	6	7	2
Human-written, bare	9	8	0	8	8	2
Human-written, template	8	8	0	8	8	2

Transition	On successor trajectories	On predecessor trajectories	Mean of the two
Model-written prompts, bare rendering			
Qwen2.5-7B	1.33 [1.25, 1.42]	0.64 [0.56, 0.70]	0.98 [0.93, 1.03]
Qwen2.5-14B	1.13 [1.08, 1.18]	0.65 [0.61, 0.70]	0.89 [0.85, 0.93]
Llama-3.1-8B	0.77 [0.66, 0.87]	0.55 [0.36, 0.66]	0.66 [0.53, 0.74]
Gemma-2-9B	0.59 [0.55, 0.63]	0.70 [0.65, 0.74]	0.64 [0.60, 0.68]
Mistral-7B-v0.3	—	—	—
OLMo-2-13B	0.68 [0.65, 0.71]	0.40 [0.33, 0.45]	0.54 [0.50, 0.57]
OLMo-2-7B	0.62 [0.59, 0.65]	0.40 [0.34, 0.45]	0.51 [0.47, 0.55]
Model-written prompts, chat-template rendering			
Qwen2.5-7B	1.19 [1.12, 1.29]	0.71 [0.61, 0.78]	0.95 [0.89, 1.01]
Qwen2.5-1.5B	1.04 [1.01, 1.06]	1.01 [1.00, 1.02]	1.03 [1.01, 1.04]
Qwen2.5-14B	0.89 [0.86, 0.92]	0.61 [0.56, 0.66]	0.75 [0.72, 0.78]
Llama-3.1-8B	0.79 [0.69, 0.87]	0.65 [0.52, 0.74]	0.72 [0.63, 0.79]
Gemma-2-9B	0.52 [0.44, 0.58]	0.73 [0.67, 0.77]	0.62 [0.56, 0.67]
Mistral-7B-v0.3	0.69 [0.60, 0.77]	0.50 [0.36, 0.60]	0.60 [0.50, 0.67]
OLMo-2-13B	0.71 [0.68, 0.74]	0.41 [0.36, 0.45]	0.56 [0.53, 0.59]
OLMo-2-7B	0.61 [0.58, 0.64]	0.45 [0.38, 0.50]	0.53 [0.49, 0.57]
Olmo-3 → RL-Zero-Code	1.60 [1.42, 2.01]	1.41 [1.30, 1.68]	1.50 [1.36, 1.85]
Human-written prompts, bare rendering			
Qwen2.5-7B	1.32 [1.23, 1.46]	0.71 [0.60, 0.80]	1.02 [0.95, 1.09]
Qwen2.5-1.5B	1.10 [1.03, 1.20]	1.01 [0.96, 1.07]	1.05 [1.00, 1.13]
Qwen2.5-14B	1.03 [0.98, 1.09]	0.70 [0.63, 0.76]	0.87 [0.82, 0.91]
Llama-3.1-8B	0.86 [0.81, 0.89]	0.86 [0.84, 0.88]	0.86 [0.83, 0.88]
Gemma-2-9B	0.72 [0.68, 0.77]	0.87 [0.84, 0.89]	0.79 [0.77, 0.82]
Mistral-7B-v0.3	0.81 [0.75, 0.87]	0.79 [0.74, 0.82]	0.80 [0.76, 0.83]
OLMo-2-13B	0.67 [0.65, 0.70]	0.62 [0.58, 0.65]	0.64 [0.62, 0.67]
OLMo-2-7B	0.60 [0.57, 0.62]	0.64 [0.61, 0.67]	0.62 [0.60, 0.64]
Olmo-3 → RL-Zero-Code	−0.757 [−5.170, 0.005]	−0.27 [−3.44, 0.32]	−0.51 [−4.39, 0.16]
Human-written prompts, chat-template rendering			
Qwen2.5-7B	1.13 [1.07, 1.21]	0.76 [0.69, 0.81]	0.95 [0.90, 0.99]
Qwen2.5-1.5B	1.02 [1.00, 1.05]	1.00 [0.99, 1.01]	1.01 [1.00, 1.02]
Qwen2.5-14B	0.87 [0.82, 0.91]	0.59 [0.51, 0.66]	0.73 [0.67, 0.78]
Llama-3.1-8B	0.85 [0.81, 0.89]	0.87 [0.85, 0.89]	0.86 [0.84, 0.89]
Gemma-2-9B	0.72 [0.68, 0.77]	0.95 [0.93, 0.96]	0.83 [0.81, 0.86]
Mistral-7B-v0.3	0.80 [0.74, 0.86]	0.78 [0.74, 0.82]	0.79 [0.75, 0.83]
OLMo-2-13B	0.69 [0.66, 0.71]	0.65 [0.63, 0.68]	0.67 [0.65, 0.69]
OLMo-2-7B	0.62 [0.59, 0.64]	0.74 [0.71, 0.77]	0.68 [0.66, 0.70]

B.4 Why the Two References Disagree

The two references' changes in propensity need not agree: on the Olmo-3 RL transition, they are -0.25 [-0.27, -0.22] and -0.03 [-0.04, -0.02] respectively, and on 2 transitions per population the two fall on opposite sides of zero.

Identity. With $g(c) = p_T(E \mid c) - p_B(E \mid c)$, the successor-reference change in propensity is the average of g over the successor's trajectories and the predecessor-reference change is its average over the predecessor's; their difference is the interaction term of Equation (6). It is nonzero only if g varies across contexts and the two trajectory distributions weight those contexts differently.

Table 3: The change in propensity under each reference for the main-text case studies and the appendix-only Olmo-3-7B-RL-Zero-Mix transition. The successor-reference column is $r_T(R_T) - r_B(R_T)$ and the predecessor-reference column is $r_T(R_B) - r_B(R_B)$. Rates per 1,000 tokens with 95% bootstrap intervals.

Transition, class, population	On successor trajectories	On predecessor trajectories
Olmo-3 RL, connectives, model-written	−0.25 [−0.27, −0.22]	−0.03 [−0.04, −0.02]
Olmo-3-7B-RL-Zero-Mix, connectives, model-written	−0.17 [−0.19, −0.15]	0.04 [0.03, 0.06]
MiMo, ordinal/sequential, mathematics	0.28 [0.19, 0.36]	0.30 [0.22, 0.37]
MiMo, ordinal/sequential, model-written	0.62 [0.58, 0.66]	0.61 [0.56, 0.65]
Llama SimpleRL, connectives, model-written	−0.69 [−0.74, −0.64]	−0.69 [−0.73, −0.64]
Llama SimpleRL, connectives, mathematics	−1.54 [−1.67, −1.43]	−1.33 [−1.42, −1.23]
Llama SimpleRL, connectives, human-written	−0.97 [−1.05, −0.90]	−0.77 [−0.83, −0.72]
Response termination, model-written	−0.01 [−0.02, −0.01]	−0.10 [−0.11, −0.08]

Size. Table 4 gives the interaction term and the Shapley (mean-of-two) components for all 30 rows we report. The interaction excludes zero on 24 of 30 rows and, among rows whose Δ_{gen} interval excludes zero, reaches 0.49 [0.39, 0.63] of Δ_{gen}. The change in propensity changes sign between the two references on 6 rows at the point estimate: Olmo-3-7B-Instruct on both populations, the appendix-only Olmo-3-7B-RL-Zero-Mix transition, Qwen2.5-7B on both populations, and Qwen2.5-1.5B on human-written prompts.

Hypothesis. We hypothesize that the successor's contexts contain its own earlier structure, including earlier occurrences of the event and plan-announcing openings, under which the predecessor's event probability rises by continuing that structure more than the successor's does. Such contexts are common in the successor's trajectories and rare in the predecessor's, which would make the successor-reference change in propensity more negative.

Evidence consistent with this hypothesis. On the Olmo-3 RL transition, 0.31 [0.29, 0.32] of the successor's scored positions precede any occurrence of the class, against the predecessor's 0.90 [0.90, 0.91]; the rise is concentrated at paragraph openings (Appendix I.3).

Evidence against the simplest form. Splitting responses by whether the first paragraph announces a plan, the interaction term of −0.22 [−0.24, −0.19] divides into −0.06 [−0.14, 0.04] from composition and −0.16 [−0.27, −0.08] within strata (Table 5), so plan announcement alone does not account for it. The control class splits mostly by composition (composition 0.72 [0.58, 0.89] of a gap of 0.81 [0.73, 0.90], a share of 0.89 [0.72, 1.10]), so the split is not specific to the counted class.

Earlier-occurrence strata and position bins. Table 6 stratifies the change in propensity by the number of earlier occurrences of the class in the same response and by 64-position bins, under both references, for three case-study cells: Olmo-3 and Llama, whose main-text paragraphs cite it, and MiMo.

On Olmo-3, the change in propensity is negative at every stratum under the successor reference, and it is largest in magnitude of the three history strata before the response contains any token of the connective class ($h = 0$: −0.34 [−0.38, −0.29]). On Llama, the change in propensity concentrates in the first 64 positions (−3.21 [−3.41, −3.00]), and after them the predecessor share stays within about a tenth of one (predecessor share 1.10 [1.07, 1.13] at positions 64–127, falling to 0.93 [0.91, 0.96] at 320–383). On MiMo, the change in propensity changes sign with occurrence history: it is negative before the first ordinal ($h = 0$: −0.60 [−0.87, −0.37]), then positive after ($h = 1$: 0.47 [0.37, 0.57]; $h \geq 2$: 0.54 [0.45, 0.63]). This sign change holds under both references.

Table 4: The interaction term and the Shapley components for all 30 rows we report. Δ_{gen} is the counted change; Δ_{prop} is the successor-reference change in propensity; Δ^{B}_{prop} is the predecessor-reference change; $I = \Delta_{\text{prop}} - \Delta^{B}_{\text{prop}}$ is the interaction; prop_{Sh} and reach_{Sh} are the Shapley (mean-of-two-references) components in rate units. Two lines per row: point estimates, then 95% bootstrap intervals. $*$ marks rows on which the two references' changes in propensity differ in sign. Rates per 1,000 tokens.

Case	Δ_{gen}	Δ_{prop}	Δ^{B}_{prop}	I	prop_{Sh}	reach_{Sh}
Gemma-2-9B	2.79	1.34	0.77	0.58	1.06	1.74
model-written	[2.34, 3.28]	[1.18, 1.52]	[0.68, 0.86]	[0.42, 0.74]	[0.95, 1.18]	[1.34, 2.16]
Llama-3.1-8B	1.03	0.21	0.35	−0.14	0.28	0.74
model-written	[0.73, 1.32]	[0.12, 0.30]	[0.30, 0.42]	[−0.24, −0.05]	[0.22, 0.34]	[0.47, 1.03]
Mistral-7B-v0.3	0.84	0.26	0.42	−0.16	0.34	0.50
model-written	[0.67, 1.01]	[0.19, 0.33]	[0.35, 0.49]	[−0.25, −0.08]	[0.29, 0.40]	[0.35, 0.65]
OLMo-2-13B	4.19	1.22	2.48	−1.26	1.85	2.33
model-written	[3.82, 4.54]	[1.07, 1.37]	[2.30, 2.66]	[−1.45, −1.07]	[1.72, 1.98]	[2.03, 2.62]
OLMo-2-7B	3.84	1.48	2.13	−0.65	1.81	2.04
model-written	[3.47, 4.19]	[1.33, 1.63]	[1.95, 2.31]	[−0.86, −0.46]	[1.68, 1.94]	[1.72, 2.35]
Olmo-3 → RL-Zero-Code	−0.56	0.33	0.23	0.10	0.28	−0.85
model-written	[−0.78, −0.32]	[0.31, 0.36]	[0.21, 0.25]	[0.07, 0.13]	[0.26, 0.30]	[−1.07, −0.60]
Olmo-3 → Instruct	−0.21	−0.32	0.93	−1.25	0.30	−0.51
model-written*	[−0.46, 0.03]	[−0.43, −0.23]	[0.74, 1.11]	[−1.43, −1.08]	[0.18, 0.42]	[−0.72, −0.31]
Qwen2.5-14B	2.26	0.25	0.88	−0.63	0.56	1.70
model-written	[1.99, 2.54]	[0.17, 0.32]	[0.78, 0.99]	[−0.74, −0.53]	[0.49, 0.64]	[1.46, 1.94]
Qwen2.5-1.5B	1.77	−0.07	−0.022	−0.04	−0.04	1.81
model-written	[1.50, 2.06]	[−0.11, −0.02]	[−0.043, 0.000]	[−0.09, −0.01]	[−0.07, −0.02]	[1.54, 2.10]
Qwen2.5-7B	1.02	−0.20	0.30	−0.50	0.05	0.97
model-written*	[0.79, 1.25]	[−0.27, −0.13]	[0.21, 0.39]	[−0.59, −0.41]	[−0.02, 0.12]	[0.76, 1.19]
Gemma-2-9B	3.31	0.93	0.18	0.75	0.55	2.76
human-written	[2.93, 3.73]	[0.77, 1.09]	[0.14, 0.23]	[0.59, 0.90]	[0.47, 0.64]	[2.41, 3.14]
Llama-3.1-8B	2.14	0.32	0.27	0.05	0.29	1.84
human-written	[1.87, 2.41]	[0.22, 0.41]	[0.23, 0.32]	[−0.05, 0.13]	[0.24, 0.35]	[1.60, 2.11]
Mistral-7B-v0.3	1.06	0.21	0.23	−0.02	0.22	0.84
human-written	[0.92, 1.22]	[0.15, 0.28]	[0.20, 0.27]	[−0.09, 0.05]	[0.18, 0.27]	[0.71, 0.99]
OLMo-2-13B	4.32	1.35	1.50	−0.150	1.42	2.90
human-written	[3.99, 4.69]	[1.22, 1.48]	[1.37, 1.63]	[−0.305, 0.005]	[1.32, 1.53]	[2.62, 3.20]
OLMo-2-7B	4.27	1.63	1.11	0.53	1.37	2.90
human-written	[3.93, 4.62]	[1.51, 1.76]	[1.00, 1.22]	[0.38, 0.68]	[1.27, 1.46]	[2.63, 3.19]
Olmo-3 → RL-Zero-Code	0.06	0.33	0.22	0.11	0.27	−0.21
human-written	[−0.10, 0.22]	[0.30, 0.36]	[0.20, 0.24]	[0.08, 0.14]	[0.25, 0.29]	[−0.37, −0.05]
Olmo-3 → Instruct	−0.17	−0.24	0.56	−0.80	0.16	−0.33
human-written*	[−0.37, 0.03]	[−0.33, −0.15]	[0.41, 0.70]	[−0.94, −0.66]	[0.06, 0.25]	[−0.49, −0.17]
Qwen2.5-14B	1.71	0.23	0.71	−0.48	0.47	1.25
human-written	[1.45, 1.96]	[0.15, 0.31]	[0.59, 0.82]	[−0.58, −0.38]	[0.38, 0.55]	[1.00, 1.49]
Qwen2.5-1.5B	2.13	−0.05	0.01	−0.06	−0.023	2.15
human-written*	[1.85, 2.44]	[−0.10, −0.01]	[−0.01, 0.02]	[−0.10, −0.01]	[−0.049, 0.003]	[1.86, 2.47]
Qwen2.5-7B	1.23	−0.17	0.30	−0.46	0.07	1.17
human-written*	[1.01, 1.47]	[−0.24, −0.09]	[0.23, 0.37]	[−0.54, −0.38]	[0.01, 0.13]	[0.95, 1.39]
Olmo-3 → RL-Zero-Code	2.71	−0.25	−0.03	−0.22	−0.14	2.85
connectives	[2.60, 2.82]	[−0.27, −0.22]	[−0.04, −0.02]	[−0.24, −0.19]	[−0.15, −0.12]	[2.74, 2.96]
MiMo-7B	1.42	0.28	0.30	−0.02	0.29	1.13
ordinals, mathematics	[1.14, 1.75]	[0.19, 0.36]	[0.22, 0.37]	[−0.08, 0.04]	[0.21, 0.36]	[0.86, 1.45]
MiMo-7B	1.51	0.62	0.61	0.01	0.61	0.89
ordinals, model-written	[1.30, 1.72]	[0.58, 0.66]	[0.56, 0.65]	[−0.02, 0.05]	[0.57, 0.66]	[0.69, 1.11]
Llama-3.1-8B SimpleRL	−1.06	−0.69	−0.69	−0.01	−0.69	−0.37
connectives, model-written	[−1.18, −0.92]	[−0.74, −0.64]	[−0.73, −0.64]	[−0.04, 0.02]	[−0.73, −0.64]	[−0.48, −0.25]
Llama-3.1-8B SimpleRL	−2.19	−1.54	−1.33	−0.22	−1.43	−0.76
connectives, mathematics	[−2.68, −1.67]	[−1.67, −1.43]	[−1.42, −1.23]	[−0.32, −0.12]	[−1.53, −1.34]	[−1.22, −0.24]
Llama-3.1-8B SimpleRL	−0.77	−0.97	−0.77	−0.20	−0.87	0.11
connectives, human-written	[−1.10, −0.40]	[−1.05, −0.90]	[−0.83, −0.72]	[−0.25, −0.15]	[−0.93, −0.81]	[−0.20, 0.45]
Olmo-3 → RL-Zero-Code	−0.34	−0.01	−0.10	0.08	−0.05	−0.28
termination, 384-token cap	[−0.38, −0.29]	[−0.02, −0.01]	[−0.11, −0.08]	[0.07, 0.10]	[−0.06, −0.05]	[−0.32, −0.24]
Olmo-3 → RL-Zero-Code	−0.09	−0.12	−0.10	−0.018	−0.11	0.03
termination, 2,048-token cap	[−0.15, −0.03]	[−0.13, −0.11]	[−0.11, −0.09]	[−0.032, −0.005]	[−0.12, −0.10]	[−0.03, 0.08]
Olmo-3 → RL-Zero-Code	0.27	−0.12	−0.04	−0.08	−0.08	0.35
termination, 8,192-token cap	[0.23, 0.31]	[−0.13, −0.11]	[−0.05, −0.04]	[−0.09, −0.06]	[−0.09, −0.07]	[0.30, 0.39]
Olmo-3 → RL-Zero-Mix	2.73	−0.17	0.04	−0.21	−0.06	2.79
connectives (app. only)*	[2.62, 2.83]	[−0.19, −0.15]	[0.03, 0.06]	[−0.24, −0.19]	[−0.08, −0.05]	[2.68, 2.90]

Table 5: The interaction term on the Olmo-3 RL transition, decomposed by plan announcement. Two blocks: the counted class (connectives) and the control class. Each block gives the composition and within-strata terms of the interaction term; the shares of positions are the same for both classes. The composition term measures how much the shift in plan prevalence accounts for the interaction; the within-strata term measures the residual; Prop. diff. is the stratum's change in propensity under the successor (succ.) and predecessor (pred.) reference. The control class's gap is carried mostly by composition. Rates per 1,000 tokens; shares are fractions of scored positions in the stratum. 95% bootstrap intervals on totals.

	Share of positions		per 1,000 tokens				
Stratum	Succ.	Pred.	Prop. diff., succ.	Prop. diff., pred.	Composition	Within strata	Gap
Connective class							
Plan	0.98	0.04	−0.25	−0.14	−0.18	−0.05	—
No plan	0.02	0.96	−0.24	−0.02	0.13	−0.11	—
Total	—	—	—	—	−0.06 [−0.14, 0.04]	−0.16 [−0.27, −0.08]	−0.22 [−0.24, −0.19]
Control class							
Plan	0.98	0.04	−0.48	−0.20	−0.32	−0.14	—
No plan	0.02	0.96	−0.86	−1.34	1.04	0.23	—
Total	—	—	—	—	0.72 [0.58, 0.89]	0.09 [−0.08, 0.24]	0.81 [0.73, 0.90]

The propensity share and its two groups. Beside the signed ratio (Section 3.1), the appendix reports the *propensity share* $s_{\text{prop}} = |\Delta_{\text{prop}}|/(|\Delta_{\text{reach}}| + |\Delta_{\text{prop}}|)$, interpreted within one of two groups, assigned by the signs of the two point estimates relative to the count. When both components move with the count (the *same-direction* group), the denominator equals the counted change and $s_{\text{prop}} = \Delta_{\text{prop}}/\Delta_{\text{gen}}$, the signed ratio, which lies between 0 and 1. When they move in opposite directions (the *opposite-direction* group), the denominator is the gross movement $|\Delta_{\text{reach}}|+|\Delta_{\text{prop}}|$, which exceeds the counted change. s_{prop} is then a share of that movement, and which component opposes the count is part of the result: the signed ratio is below 0 where propensity opposes it and above 1 where reach does. Where a component's interval includes zero, its sign, and hence the group, is uncertain, and we mark the case with ‡. A share read without its group is ambiguous: one half means an even split in the first group and exact cancellation in the second.

The mean of the two references. Table 7 prints the decomposition under the mean of the two references for every census transition. Qwen2.5-7B, opposite-direction under the successor reference, becomes same-direction under the mean on both populations (sign uncertain on model-written prompts); the non-rising Olmo-3 Instruct transition's component against the count changes from reach to propensity.

C Estimation and Calibration

C.1 Teacher-Forcing Procedure and the Decoding Transform

The estimation procedure described in Section 3.2 requires both models to be teacher-forced on both sets of stored trajectories. At every scored position the model's next-token distribution is put through the decoding transform, the same transform the trajectories were sampled under (temperature 0.7, truncation to the top 0.95 of the probability mass, renormalization), so that a model's expected event count on its own trajectories matches its sampled count in expectation. Both the change in reach and the change in propensity inherit this transform, because each is a difference of two rates computed under it. Appendix C.3 repeats the decomposition on the untransformed next-token distributions to confirm that the results do not depend on the transform.

A rate is an expected event count per thousand scored positions of the whole response, pooled across prompts as a ratio of sums. Four such passes, one per cell of Equation (1), give the rates the decomposition needs.

C.2 Calibration, Same-Model Floors, and Controls

Three checks precede every decomposition. The calibration check teacher-forces each model on its own stored trajectories and requires the expected count of its class's token ids, over the count it

Table 6: The change in propensity by earlier-occurrence stratum and by 64-position bin, under both references, for three case-study cells. h is the number of earlier occurrences of the class in the same response before the scored position (on the successor's trajectories for Δ_{prop}, on the predecessor's for Δ^{B}_{prop}). n_T is the number of scored positions contributed by successor trajectories. Predecessor share is $r_B(R_T)/r_T(R_T)$ at the given positions. Rates per 1,000 tokens with 95% bootstrap intervals.

Cell	Stratum	Δ_{prop} (succ. ref.)	Δ^{B}_{prop} (pred. ref.)	Pred. share	n_T
Olmo-3 RL, connectives, model-written					
	Whole	−0.25 [−0.27, −0.22]	−0.03 [−0.04, −0.02]		
	$h = 0$	−0.34 [−0.38, −0.29]	−0.02 [−0.03, −0.01]		627,744
	$h = 1$	−0.21 [−0.24, −0.18]	−0.12 [−0.17, −0.06]		1,065,713
	$h \geq 2$	−0.19 [−0.24, −0.15]	−0.05 [−0.13, 0.03]		363,892
	0–63	−0.25 [−0.32, −0.19]	−0.05 [−0.07, −0.03]	1.03 [1.02, 1.03]	345,481
	64–127	−0.33 [−0.39, −0.27]	−0.03 [−0.04, −0.01]	1.07 [1.06, 1.09]	345,174
	128–191	−0.13 [−0.18, −0.09]	0.00 [−0.02, 0.01]	1.06 [1.04, 1.09]	344,793
	192–255	−0.11 [−0.15, −0.08]	−0.02 [−0.04, −0.01]	1.09 [1.06, 1.12]	343,932
	256–319	−0.28 [−0.33, −0.23]	−0.03 [−0.05, −0.02]	1.19 [1.15, 1.22]	340,961
	320–383	−0.37 [−0.42, −0.32]	−0.03 [−0.05, −0.01]	1.21 [1.18, 1.24]	337,008
Llama SimpleRL, connectives, model-written					
	Whole	−0.69 [−0.74, −0.64]	−0.69 [−0.73, −0.64]		
	$h = 0$	−0.80 [−0.85, −0.74]	−0.84 [−0.91, −0.77]		1,229,031
	$h = 1$	−0.25 [−0.33, −0.18]	−0.44 [−0.51, −0.38]		207,273
	$h \geq 2$	−0.45 [−0.52, −0.39]	−0.39 [−0.45, −0.32]		159,356
	0–63	−3.21 [−3.41, −3.00]	−3.34 [−3.53, −3.12]	1.86 [1.75, 2.00]	343,707
	64–127	−0.13 [−0.17, −0.09]	−0.18 [−0.23, −0.14]	1.10 [1.07, 1.13]	326,389
	128–191	−0.03 [−0.07, 0.01]	−0.03 [−0.08, 0.02]	1.02 [1.00, 1.05]	299,288
	192–255	0.04 [−0.01, 0.09]	0.01 [−0.03, 0.06]	0.98 [0.95, 1.01]	259,075
	256–319	0.11 [0.06, 0.16]	0.14 [0.09, 0.18]	0.94 [0.92, 0.96]	209,611
	320–383	0.12 [0.07, 0.16]	0.17 [0.12, 0.21]	0.93 [0.91, 0.96]	157,590
MiMo RL, ordinal/sequential (24-id class), mathematics					
	Whole	0.28 [0.19, 0.36]	0.30 [0.22, 0.37]		
	$h = 0$	−0.60 [−0.87, −0.37]	−0.38 [−0.60, −0.19]		289,831
	$h = 1$	0.47 [0.37, 0.57]	0.47 [0.39, 0.56]		393,323
	$h \geq 2$	0.54 [0.45, 0.63]	0.52 [0.43, 0.61]		697,844
	0–63	−0.78 [−1.09, −0.50]	−0.46 [−0.73, −0.23]	1.08 [1.05, 1.11]	230,400
	64–127	0.41 [0.26, 0.56]	0.32 [0.18, 0.46]	0.97 [0.96, 0.98]	230,338
	128–191	0.56 [0.44, 0.69]	0.43 [0.31, 0.55]	0.94 [0.93, 0.96]	230,187
	192–255	0.48 [0.36, 0.61]	0.52 [0.41, 0.63]	0.95 [0.94, 0.96]	230,144
	256–319	0.54 [0.41, 0.66]	0.49 [0.39, 0.60]	0.94 [0.93, 0.96]	230,093
	320–383	0.46 [0.33, 0.59]	0.51 [0.39, 0.63]	0.95 [0.94, 0.96]	229,836

Table 7: The census decomposition under the mean of the two references, chat-template rendering. s is the propensity share s_{prop}, printed with its group (Appendix B.4). ▲: the group changes from the successor reference to the mean; ▽: the component against the count changes; ‡ (mean column): sign uncertain; (n.r.): counted-change interval includes zero; a dash marks a share withheld under Appendix C.4 (iv). Rates per 1,000 tokens with 95% bootstrap intervals from the census's own draws, not the resample of Table 2.

Transition	Δ_{gen}	Succ. ref. group, s	Mean of references Δ_{prop}	Mean of references Δ_{reach}	Mean group, s [95% CI]	
Model-written prompts						
Qwen2.5-1.5B	1.77 [1.50, 2.06]	opp., 0.04	−0.04	1.81	opp., 0.02 [0.01, 0.04]	
Qwen2.5-7B	1.02 [0.79, 1.25]	opp., 0.14	0.05	0.97	same, 0.05 [0.00, 0.11]‡	▲
Qwen2.5-14B	2.26 [1.99, 2.54]	same, 0.11	0.56	1.70	same, 0.25 [0.22, 0.28]	
Llama-3.1-8B	1.03 [0.73, 1.32]	same, 0.21	0.28	0.74	same, 0.28 [0.21, 0.37]	
Mistral-7B-v0.3	0.84 [0.67, 1.01]	same, 0.31	0.34	0.50	same, 0.40 [0.34, 0.50]	
Gemma-2-9B	2.79 [2.34, 3.28]	same, 0.48	1.06	1.74	same, 0.38 [0.33, 0.43]	
OLMo-2-13B	4.19 [3.82, 4.54]	same, 0.29	1.85	2.33	same, 0.44 [0.41, 0.47]	
OLMo-2-7B	3.84 [3.47, 4.19]	same, 0.39	1.81	2.04	same, 0.47 [0.43, 0.51]	
Olmo-3 → RL-Zero-Code	−0.56 [−0.78, −0.32]	opp., 0.27	0.28	−0.85	opp., 0.25 [0.21, 0.32]	
Olmo-3 → Instruct	−0.21 [−0.46, 0.03] n.r.	opp., —	0.30	−0.51	opp., —	▽
Human-written prompts						
Qwen2.5-1.5B	2.13 [1.85, 2.44]	opp., 0.02	−0.02	2.15	opp., 0.01 [0.00, 0.02]‡	
Qwen2.5-7B	1.23 [1.01, 1.47]	opp., 0.11	0.07	1.17	same, 0.05 [0.01, 0.10]	▲
Qwen2.5-14B	1.71 [1.45, 1.96]	same, 0.13	0.47	1.25	same, 0.27 [0.22, 0.33]	
Llama-3.1-8B	2.14 [1.87, 2.41]	same, 0.15	0.29	1.84	same, 0.14 [0.11, 0.16]	
Mistral-7B-v0.3	1.06 [0.92, 1.22]	same, 0.20	0.22	0.84	same, 0.21 [0.17, 0.25]	
Gemma-2-9B	3.31 [2.93, 3.73]	same, 0.28	0.55	2.76	same, 0.17 [0.14, 0.19]	
OLMo-2-13B	4.32 [3.99, 4.69]	same, 0.31	1.42	2.90	same, 0.33 [0.31, 0.35]	
OLMo-2-7B	4.27 [3.93, 4.62]	same, 0.38	1.37	2.90	same, 0.32 [0.30, 0.34]	
Olmo-3 → RL-Zero-Code	0.06 [−0.10, 0.22] n.r.	opp., —	0.27	−0.21	opp., —	
Olmo-3 → Instruct	−0.17 [−0.37, 0.03] n.r.	opp., —	0.16	−0.33	opp., —	▽

Table 8: The calibration check: each model teacher-forced on its own stored trajectories. The pooled ratio is the model's expected count of its own class ids over the count it sampled; every scored row sits within a tenth of one.

Measurement	Pooled ratio	Spearman
Olmo-3 RL case study, predecessor and successor	0.99–1.00	0.99–1.00
Formatting markers (six passes)	1.00	stored
Qwen2.5-7B SimpleRL (both passes)	1.00	—
Full per-model list, all rows (worst)	0.94	—

sampled, to lie within a tenth of one, with the two counts ranking the words at a Spearman correlation of at least 0.95 (Table 8). The rendering-equality check requires every rendered prompt to be byte-identical within each cell's own prefix; under the bare rendering the predecessor's and successor's prefixes differ by design. The stricter round-trip check requires every response's token ids to re-tokenize with no mismatch.

Same-model floors. A same-model comparison (one model measured against a second draw of its own trajectories, where the true change is zero) puts the OLMo-2-0425-1B lineage floor on the screened prompts at 0.19 [−0.03, 0.42] per 1,000 tokens and the Olmo-3-1025-7B one at −0.03 [−0.15, 0.07], both intervals including zero. The OLMo-2-1B SFT → DPO transition's own floor on the model-written prompts is −0.01 [−0.08, 0.06], 0.04 of its connective-class change. Of the Qwen2.5-7B same-model floors this paper prints, one excludes zero: the Qwen2.5-7B → SimpleRL-Zoo transition's, on the connective class, at 0.51 [0.02, 1.19] of the change. A same-model floor that is not below 0.10 of a change indicates that the measurement's accounting is inexact there, and the shares of that change are withheld wherever they would be printed (Appendix C.4); a transition without a floor of its own on a population is held to its lineage's floor (the transitions trained from the same pretrained model) for that event there, and where the lineage has no floor for that event, population, generation cap and scoring window, the shares are printed without one.

The control class. The control class (42 comparable word forms that the decomposition's case studies do not predict will move) moves with the connective class on two of the seven case-study

Table 9: The connective class's decomposition under the two next-token distributions, on both prompt populations. The transformed distribution (after the decoding transform) is the one the rest of the paper uses; the raw distribution is the model's untransformed next-token distribution. The last column is the paired difference between the two distributions' $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (mean of the two references) values, computed on the same bootstrap draws. Rates per 1,000 tokens with 95% bootstrap intervals.

Prompts	Distribution	Counted change
Model-written	transformed	2.71 [2.60, 2.83]
	raw	2.60 [2.50, 2.70]
Screened	transformed	2.27 [2.05, 2.46]
	raw	2.14 [1.93, 2.32]

Prompts	Distribution	On successor trajectories	On predecessor trajectories
Model-written	transformed	−0.25 [−0.27, −0.22]	−0.03 [−0.04, −0.02]
	raw	−0.18 [−0.20, −0.16]	−0.04 [−0.05, −0.04]
Screened	transformed	−0.24 [−0.26, −0.22]	−0.05 [−0.06, −0.04]
	raw	−0.20 [−0.22, −0.18]	−0.07 [−0.08, −0.06]

Prompts	Distribution	$\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (mean of the two references)	Raw minus transformed
Model-written	transformed	−0.05 [−0.06, −0.05]	
	raw	−0.04 [−0.05, −0.04]	0.007 [0.005, 0.008]
Screened	transformed	−0.06 [−0.07, −0.06]	
	raw	−0.06 [−0.07, −0.06]	0.001 [−0.001, 0.003]

cells that Appendix C.4 (iv) reads. The control class's movement is reported beside every case-study number in the appendix, so that a reader can assess specificity row by row.

C.3 The Decomposition on Raw Next-Token Distributions

The decoding transform is applied to both models, and the change in propensity could in principle be a property of that transform rather than of the models. Table 9 rules this out. Computed on the untransformed next-token distributions, the connective class's counted change on the model-written population is 2.60 [2.50, 2.70] per 1,000 tokens against 2.71 [2.60, 2.82] under the transform, and both changes in propensity stay below zero: −0.18 [−0.20, −0.16] per 1,000 tokens on successor trajectories and −0.04 [−0.05, −0.04] on predecessor trajectories. Raw counted changes and their components sit closer to zero than transformed ones almost everywhere, as expected from a temperature below one and a truncated tail; the signed ratios barely move. The paired difference between the two distributions' $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (mean of the two references) values is 0.007 [0.005, 0.008] of the counted change on the model-written population and 0.001 [−0.001, 0.003] on the screened population. On the control class the same gap is −0.03 [−0.03, −0.02], and on response termination −0.04 [−0.05, −0.03].

C.4 Resampling Units and the Bootstrap

Every interval in the paper is a 95% central bootstrap interval at 2,000 draws, except the two exchange tables of Appendices G.4 and I.1 (10,000).

Prompt-paired intervals. For every decomposition rate, every census share, the local-intervention corrections, and the plan-announcement strata, the resampling unit is the prompt. One multiplicity draw $k \sim \text{Multinomial}(n, 1/n)$ over the n prompts recomputes each rate as $1000 \sum_i k_i a_i / \sum_i k_i d_i$, where a_i is the event count and d_i the scored-position count on prompt i. One draw is shared by every word and every cell, so that dependence across words and across cells is preserved within each replicate. The Llama paragraph exchange (Appendix K.4) also uses prompt-paired intervals, as Figure 3 states. The Qwen2.5-7B contrast between the two interventions (Appendix K.5) is prompt-paired because the locally intervened arm has no exchange prefixes, so the prompt is the only unit both arms share: one draw over the 94 shared prompts reweights both, each prefix at its prompt's weight; the exchange's m in Figure 3 and Table 47 is read on the same draws.

Prefix-paired intervals. For the other paragraph exchanges and their m and $\bar{m}$ (Appendices G.4, I.1 and I.2), the resampling unit is the eligible prefix. The construction is the same, with the prefix replacing the prompt and continuation tokens replacing whole-response tokens.

Conventions. The following conventions apply throughout.

(i) Intervals are as stated at the top of this subsection. Prompt-paired and prefix-paired intervals are not compared, except where a contrast is computed on the prompts both share, which is named where printed.
(ii) Rates carry their own denominators: the decomposition rates of Section 4 are per 1,000 tokens of the whole response, pooled as a ratio of sums; exchange cells are per thousand continuation tokens; locally intervened arms are per thousand tokens of their scoring window; censored stop rates are per thousand at-risk positions. Rates with different denominators are never differenced, and a counted change is never compared with an exchange cell.
(iii) Every m, $\bar{m}$ or C is quoted with the gap it divides by (Table 47).
(iv) No share is printed where the counted change's interval includes zero or where the same-model floor is not below 0.10 of the change, and a cell whose same-model floor excludes zero and exceeds half of its change is not read; the predecessor share at matched contexts is printed only where $r_T(R_T)$ has an interval lower bound of at least 0.5 per 1,000 tokens.
(v) A decomposition is scored only after the calibration, rendering-equality and round-trip checks of Appendix C.2 hold.
(vi) The successor reference is used unless a display labels the predecessor reference or the mean of the two references.

D Models, Prompt Populations, and Rendering

D.1 Model Revisions and Training Provenance

Table 1 lists the twelve transitions with a row in the main text; Table 10 lists all twenty, drawn from eight model families; the DeepSeek-Math-7B → SimpleRL-Zoo transition is used in Appendices G.5 and L.2. Table 10 gives the exact Hugging Face repository path for the two endpoints of each transition; the Olmo-3 instruct chain's intermediate stages are named in Appendix G.3. Commit revisions are recorded in each artifact's provenance block.

All models were obtained from public Hugging Face repositories. The Llama-3.1 and Gemma-2 checkpoints are loaded through ungated mirrors (`NousResearch` and `unsloth`, respectively) because the original repositories require access approval; the Llama-3.1 and Gemma-2-9B-IT mirror files are identical to the originals (same SHA-256), and the Gemma-2-9B base mirror stores the same parameters in bfloat16 where the original stores float32. Three Olmo-3 successors share a single predecessor checkpoint (`allenai/Olmo-3-1025-7B`); Olmo-3-7B-RL-Zero-Code and Olmo-3-7B-RL-Zero-Mix are scored against one stored set of predecessor trajectories, and Olmo-3-7B-Instruct against its own. The four SimpleRL-Zoo successors (Zeng et al., 2025) follow a public GRPO recipe (Shao et al., 2024) with verifiable rewards and no supervised fine-tuning stage.

D.2 Prompt Populations and Construction

Four prompt populations appear in the paper.

The *model-written* population is 600 first user turns of UltraChat conversations (Ding et al., 2023) (the `train_sft` split of `HuggingFaceH4/ultrachat_200k`, a filtered subset), drawn once in the order of a seeded hash over the 179,351 rows that survived script, length and duplication filters among 207,865 scanned. UltraChat's user side is itself model-generated, so this population is not a sample of real user prompts.

The *human-written* population is 600 first user turns a person wrote, drawn from `lmsys/lmsys-arena-human-preference-55k`, a public release of Chatbot Arena preference data (Chiang et al., 2024), through the same pipeline with two deduplication stages added: one excludes any prompt that occurs in the model-written population, and one excludes any prompt whose normalized first user turn matches one in `allenai/tulu-3-sft-mixture`, `allenai/tulu-3-sft-olmo-2-mixture` or the

Table 10: Exact model repositories. Every checkpoint was loaded from Hugging Face; exact commit revisions are recorded in the released experiment artifacts. The Olmo-3 → RL-Zero-Code and Olmo-3 → RL-Zero-Mix rows share a single set of stored predecessor trajectories. OLMo-2 and Olmo-3 are counted as separate families because they are separate pretraining runs. The OLMo-2-1B predecessor is the stage-2 ingredient-3 branch, the SFT model's initialization, not the repository's main revision. Ungated mirrors are used for Gemma-2 and Llama-3.1 because the original repositories require access approval.

Family	Repository path	Training route
Census transitions (released instruction-tuned successors)		
Qwen2.5-1.5B	`Qwen/Qwen2.5-1.5B` → `Qwen/Qwen2.5-1.5B-Instruct`	instruction tuning
Qwen2.5-7B	`Qwen/Qwen2.5-7B` → `Qwen/Qwen2.5-7B-Instruct`	instruction tuning
Qwen2.5-14B	`Qwen/Qwen2.5-14B` → `Qwen/Qwen2.5-14B-Instruct`	instruction tuning
Llama-3.1-8B	`NousResearch/Meta-Llama-3.1-8B` → `NousResearch/Meta-Llama-3.1-8B-Instruct`	instruction tuning
Mistral-7B-v0.3	`mistralai/Mistral-7B-v0.3` → `mistralai/Mistral-7B-Instruct-v0.3`	instruction tuning
Gemma-2-9B	`unsloth/gemma-2-9b` → `unsloth/gemma-2-9b-it`	instruction tuning
OLMo-2-7B	`allenai/OLMo-2-1124-7B` → `allenai/OLMo-2-1124-7B-Instruct`	instruction tuning
OLMo-2-13B	`allenai/OLMo-2-1124-13B` → `allenai/OLMo-2-1124-13B-Instruct`	instruction tuning
Olmo-3 transitions (three successors from one predecessor)		
Olmo-3 → Instruct	`allenai/Olmo-3-1025-7B` → `allenai/Olmo-3-7B-Instruct`	SFT → DPO → RL
Olmo-3 → RL-Zero-Code	`allenai/Olmo-3-1025-7B` → `allenai/Olmo-3-7B-RL-Zero-Code`	RL from base
Olmo-3 → RL-Zero-Mix	`allenai/Olmo-3-1025-7B` → `allenai/Olmo-3-7B-RL-Zero-Mix`	RL from base
Additional RL-from-base transitions		
MiMo-7B	`XiaomiMiMo/MiMo-7B-Base` → `XiaomiMiMo/MiMo-7B-RL-Zero`	RL from base
Llama-3.1-8B SimpleRL	`NousResearch/Meta-Llama-3.1-8B` → `hkust-nlp/Llama-3.1-8B-SimpleRL-Zoo`	RL from base
Qwen2.5-7B SimpleRL	`Qwen/Qwen2.5-7B` → `hkust-nlp/Qwen-2.5-7B-SimpleRL-Zoo`	RL from base
Mistral-7B-v0.1 SimpleRL	`mistralai/Mistral-7B-v0.1` → `hkust-nlp/Mistral-7B-v0.1-SimpleRL-Zoo`	RL from base
DeepSeek-Math-7B SimpleRL	`deepseek-ai/deepseek-math-7b-base` → `hkust-nlp/DeepSeek-Math-7B-SimpleRL-Zoo`	RL from base
Additional intermediate-stage and SFT-only transitions (appendix case studies)		
OLMo-2-1B SFT	`allenai/OLMo-2-0425-1B@stage2-ingredient3-step23852-tokens51B` → `allenai/OLMo-2-0425-1B-SFT`	SFT
OLMo-2-7B SFT	`allenai/OLMo-2-1124-7B` → `allenai/OLMo-2-1124-7B-SFT`	SFT
OLMo-2-1B SFT → DPO	`allenai/OLMo-2-0425-1B-SFT` → `allenai/OLMo-2-0425-1B-DPO`	DPO
Tülu-3-8B SFT	`NousResearch/Meta-Llama-3.1-8B` → `allenai/Llama-3.1-Tulu-3-8B-SFT`	SFT

`train_sft` split of `HuggingFaceH4/ultrachat_200k`. The Olmo-3 SFT mixture was not included in deduplication, so the two Olmo-3 rows are deduplicated against the Tülu-3 mixtures and UltraChat-200k only. The model-written population is drawn from UltraChat and is not deduplicated against any successor's mixture.

The *mathematics* population is 400 NuminaMath-TIR problems (competition, school-exam and word problems with numerical answers), drawn from the decontaminated NuminaMath-TIR subset of `allenai/tulu-3-sft-olmo-2-mixture-0225`, selected on task rather than on any model's trajectories.

The *screened* population is 600 user turns from the training mixture of `allenai/OLMo-2-0425-1B-SFT` (`allenai/tulu-3-sft-olmo-2-mixture-0225`), kept because the mixture's own answer opened a paragraph with a connective. That is a filter on the training data, never on any model's trajectories, but it raises the event's base rate. The screened population is used only in the paragraph exchange and in the case-study decompositions whose rows are labeled screened; nothing rests on it alone.

Effect size depends on the population: the same transition's decomposition can differ between model-written and human-written prompts. The numbers in this paper describe the stated prompt populations and measure no prevalence over prompts in general.

D.3 Rendering and the Teacher-Forcing Round-Trip Check

Every cell is teacher-forced on the stored sampled token ids, so that no cell is scored on re-encoded text. A separate check that re-encodes an arm's stored text and compares the token ids flags between a thirtieth and a little over half of an arm's responses. On the worst arm (Llama-3.1-Tulu-3-8B-SFT on the model-written population), 3,009 of the flagged responses are a trailing end-of-sequence token the generation engine omits from the stored text, 70 are a decoder's space clean-up before punctuation, and 0 are a genuine token divergence. Dropping every response that ended at the sequence end moves the Olmo-3 RL transition's counted change by 0.04 per 1,000 tokens, leaving it distinguishable from zero. The Llama-3.1-8B $\rightarrow$ SimpleRL-Zoo transition is evaluated under its literal training prompt (`Question:\n{user}\nAnswer:\nLet's think step by step.\n`) on both arms. The MiMo transition is evaluated under the bare rendering on both arms for the ordinal/sequential class (Appendix G.4) and under the same literal SimpleRL prompt on both arms for the reflection class (Appendix G.5). On the model-written and mathematics populations the Olmo-3-1025-7B transitions have a single rendering: both arms are evaluated under the Olmo-3-7B-Instruct chat template, with the same rendered prompt at generation and at teacher forcing in all four cells, so their model-written census rows are reported under the chat template alone (Table 15); on the human-written population both renderings are evaluated (Table 16). The rendering grafts the Instruct chat template onto both models; the Olmo-3-7B-RL-Zero-Code repository ships a ChatML template that adds an OLMo system turn and ends the generation prompt with an opening `<think>` tag, so responses begin inside a thinking span; within the cap, no response closes it. Under the repository template, the successor's shift toward connectives at paragraph openings that follow prose is smaller but remains: $+0.76$ [0.64, 0.90] nats against $+0.93$ [0.81, 1.06] under the grafted template (log ratio of the connective share among those openings; 23-word class, screened prompts, 384-token cap).

E Full Released-Model Census

E.1 Two Published Excess-Vocabulary Lists on Three Released Transitions

Table 11 gives the full decomposition of two published excess-vocabulary lists (Kobak et al., 2025; Juzek & Ward, 2025a) on three released base-to-instruct transitions, each rendered under its own chat template on both arms. A word contributes an id only where one of its literal forms is a single token under both tokenizers, with the same id and an exact round trip; that rule keeps 187 of the published list's 291 words and 12 of the second list's 21, identically on all three transitions. Both lists rise on every transition. Where each rise sits differs: the change in reach accounts for the excess-vocabulary rise on Qwen2.5-7B, while roughly half of the OLMo-2-7B rise (mean of the two references) is a change in propensity.

All 6 of the 6 list-by-transition cells rise with an interval excluding zero. Over the eight behavioral classes scored on these passes, and every cell whose counted change rises with an interval excluding zero and that Appendix C.4 (iv) reads, the change in reach is the larger of the two on most cells. This paper's own connective class does not reproduce the Olmo-3 RL transition's rise on Llama-3.1-8B, where it does not separate from zero, or on Qwen2.5-7B, where it falls, by -0.18 [-0.25, -0.10] per 1,000 tokens. On OLMo-2-7B it is not read (Appendix C.4 (iv)): the transition's own same-model floor excludes zero and is 3.96 times the class's change.

E.2 The Census Across 10 Released Transitions, Chat-Template Rendering

Figure 4 and Table 12 extend the three-transition table to 10 released transitions on the published list, evaluated on the model-written population with both arms rendered under the successor's chat template (for the Olmo-3 transitions, the Olmo-3-7B-Instruct template; Appendix D.3). Each transition keeps only the forms its own two tokenizers agree on, so the resolved id sets differ by row: most rows resolve 242 single-token ids, Gemma-2-9B resolves 364 and Mistral-7B-v0.3 only 47. Of the 10 transitions, 8 rise, and on all 8 the change in reach is the larger component by point. On the second list (Juzek & Ward, 2025a) the change in propensity lies at or below zero under both references on 2 of its resolved rows, against 1 on the published list. Two rows do not rise: the Olmo-3-7B-Instruct transition, and the Olmo-3 RL transition studied in Section 4.3, whose counted change on the list is negative.

Table 11: Two published excess-vocabulary lists on three released base-to-instruct transitions, each rendered under its own chat template. Every list rises on every transition; the third row of each block gives $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (mean of the two references), showing that it differs by transition. The last rows are this paper's own behavioral classes on the same passes. Rates per 1,000 tokens with 95% bootstrap intervals; fold is the successor's rate over the predecessor's; $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ is withheld where the counted change does not separate from zero, and a dash marks a signed ratio the floor rule of Appendix C.4 withholds.

Class	Llama-3.1-8B	Qwen2.5-7B	OLMo-2-7B	Quantity
Published list	1.03 [0.73, 1.32]	1.02 [0.79, 1.25]	3.84 [3.47, 4.19]	change per 1,000 tokens
	1.4 [1.3, 1.6]	1.3 [1.2, 1.4]	2.8 [2.5, 3.1]	fold
	0.28 [0.21, 0.37]	0.05 [−0.02, 0.11]	0.47 [0.43, 0.51]	$\Delta_{\text{prop}}/\Delta_{\text{gen}}$
Second list (Juzek & Ward, 2025a)	0.19 [0.12, 0.25]	0.08 [0.03, 0.12]	0.38 [0.32, 0.43]	change per 1,000 tokens
	2.2 [1.7, 2.9]	1.2 [1.1, 1.4]	4.4 [3.4, 6.0]	fold
	0.19 [0.09, 0.31]	—	0.50 [0.42, 0.60]	$\Delta_{\text{prop}}/\Delta_{\text{gen}}$
Control class	−1.31 [−1.76, −0.87]	−1.72 [−2.06, −1.37]	−0.61 [−1.07, −0.21]	change per 1,000 tokens
Formatting markers	25.20 [22.16, 28.13]	25.86 [23.51, 28.19]	30.37 [27.50, 33.09]	change per 1,000 tokens
	0.06 [0.06, 0.07]	0.05 [0.04, 0.05]	0.10 [0.09, 0.11]	$\Delta_{\text{prop}}/\Delta_{\text{gen}}$
Connective class	−0.01 [−0.10, 0.08]	−0.18 [−0.25, −0.10]	−0.06 [−0.21, 0.08]	change per 1,000 tokens
Response ends	−0.60 [−0.73, −0.48]	−0.76 [−0.85, −0.68]	−1.14 [−1.25, −1.03]	change per 1,000 tokens
Length	36 [27, 45]	33 [29, 39]	75 [68, 81]	change, tokens

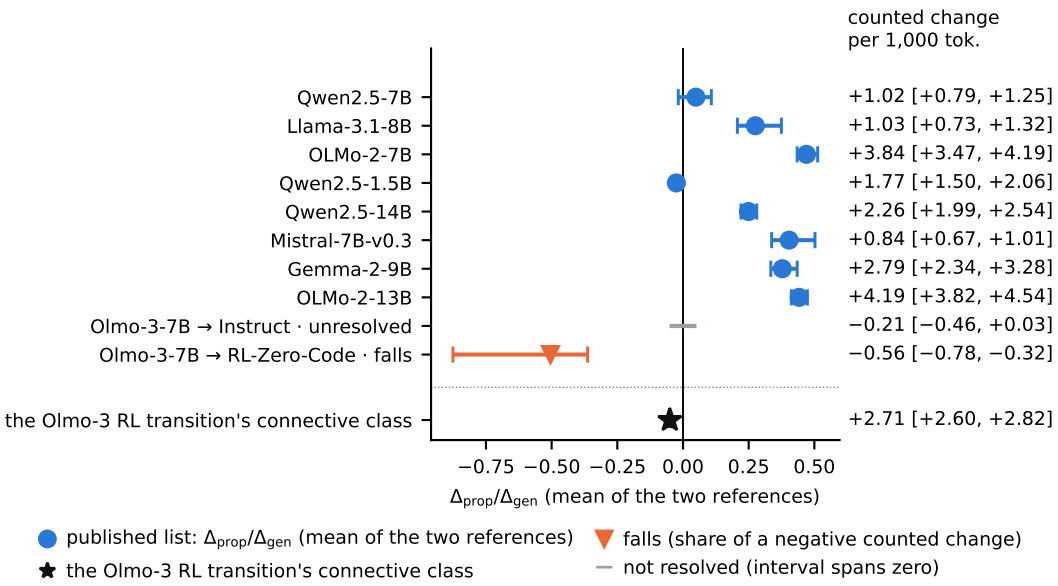


Figure 4: 10 released transitions on the published excess-vocabulary list (Kobak et al., 2025), under the successor's chat-template rendering (for the Olmo-3 transitions, the Olmo-3-7B-Instruct template; Appendix D.3). Each row is $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (mean of the two references) of that transition's counted change; two rows do not rise (the Olmo-3-7B-Instruct transition and the Olmo-3 RL transition of Section 4.3). Counted changes per 1,000 tokens appear at the right edge.

E.3 Vocabulary Mappings and the Word Sets

Each class is a set of token ids resolved per transition under that transition's tokenizer (Table 13).

Connective class. The 23 words of Table 13, each in its capitalized and its lower-case form without a leading space; each of the 46 forms contributes the first token of its encoding, so mid-sentence occurrences with a leading space are not counted. On the OLMo, Llama, Qwen and MiMo tokenizers the 46 forms give 46 distinct ids; some forms are several tokens long, and their id is the

Table 12: The same 10 transitions as numbers, under the successor's chat-template rendering (for the Olmo-3 transitions, the Olmo-3-7B-Instruct template; Appendix D.3). $\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ is the signed ratio (Section 3.1) under the successor reference; a negative value means the change in propensity moves against the count. Rates per 1,000 tokens with 95% bootstrap intervals.

Transition	Change per 1,000 tokens	$\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ (succ. ref.)	Larger component
Qwen2.5-7B	1.02 [0.79, 1.25]	−0.19 [−0.29, −0.12]	reach
Qwen2.5-1.5B	1.77 [1.50, 2.06]	−0.04 [−0.06, −0.01]	reach
Qwen2.5-14B	2.26 [1.99, 2.54]	0.11 [0.08, 0.14]	reach
Llama-3.1-8B	1.03 [0.73, 1.32]	0.21 [0.12, 0.30]	reach
Gemma-2-9B	2.79 [2.34, 3.28]	0.48 [0.43, 0.55]	reach
Mistral-7B-v0.3	0.84 [0.67, 1.01]	0.31 [0.24, 0.40]	reach
OLMo-2-13B	4.19 [3.82, 4.54]	0.29 [0.26, 0.32]	reach
OLMo-2-7B	3.84 [3.47, 4.19]	0.39 [0.35, 0.42]	reach
Olmo-3 → Instruct‡	−0.21 [−0.46, 0.03]	—‖	no rise
Olmo-3 → RL-Zero-Code	−0.56 [−0.78, −0.32]	−0.60 [−1.05, −0.42]	falls

‡ Sign uncertain: one component's interval includes zero.
‖ Opposite-direction group (Appendix B.4); reach moves against the count; sign uncertain.

word-initial piece (for example the id for "furthermore" is "f" and the id for "therefore" is "there", which is also a control-class id). On the Mistral-7B-v0.1 tokenizer every id carries a word-boundary marker.

How it was chosen. The words are the single-word members of a list of 34 discourse connectives written down before a test of whether excess-vocabulary markers belong to a wider class, on generations from OLMo-2-1124-7B and its SFT successor; the list was fixed before that test ran and excluded the markers under test. The id rule was fixed on the Olmo-3-1025-7B tokenizer before the first decomposition was computed. No word was chosen by its counted change or its change in propensity.

Ordinal/sequential class. The six forms first, second, third, next, then and finally, the ordinal subset of the connective class, fixed before the MiMo family was examined. Each form enters in four variants (lower-case and capitalized, with and without a leading space) where the variant is a single token under both tokenizers of the transition with an exact round trip; on MiMo all 24 variants resolve, and the ids were frozen before any trajectory was generated.

Control class. The 21 frequent paragraph-initial words of Table 13 that have no discourse-relational function, fixed before running as a control for the connective list and disjoint from it. Beside the connective class it is resolved by the connective rule (42 ids); beside the ordinal class, by the ordinal rule (84 ids).

Formatting markers and the published lists follow Table 13 and Appendix E.1.

Restricted to the 31 connective-class ids whose form is a single complete-word token, which drops the word-initial pieces and with them the id shared with the control class, the counted change is above zero and the change in propensity below zero under both references on all three Olmo-3 → RL-Zero-Code cells (population by rendering), the counted change and the change in propensity under both references are below zero on all three Llama-3.1-8B → SimpleRL-Zoo cells, each with an interval excluding zero, and on none of the six cells does a sign or an excludes-zero status differ from the full class's.

E.4 Per-Prompt Averaged Rates

Does the census depend on how rates are aggregated? The main-text census reports the signed ratio as a ratio of sums. Table 14 recomputes it as the propensity share (Appendix B.4), which equals the signed ratio on same-direction rows, on the same 10 transitions under the chat-template rendering with per-prompt averaged rates. On both populations, the per-prompt propensity share is within a tenth of the ratio-of-sums value on every rising transition. Two rising rows change group between the two aggregations: Qwen2.5-1.5B on model-written prompts (s_{prop} from 0.04 [0.01, 0.06] to 0.002 [0.001, 0.048], sign uncertain) and on human-written prompts (s_{prop} from 0.023 [0.004, 0.043] to 0.010 [0.001, 0.047], sign uncertain). On human-written prompts, Olmo-3-7B-

Table 13: The three fixed behavioral classes used throughout: the connective class, the control class and the formatting markers (Appendix E.3).

Connective forms	Control forms	Formatting markers
Additionally, additionally, Also, also, Consequently, consequently, Finally, finally, First, first, Furthermore, furthermore, However, however, Indeed, indeed, Instead, instead, Likewise, likewise, Meanwhile, meanwhile, Moreover, moreover, Nevertheless, nevertheless, Next, next, Nonetheless, nonetheless, Overall, overall, Second, second, Similarly, similarly, Then, then, Therefore, therefore, Third, third, Thus, thus, Ultimately, ultimately	A, a, An, an, Both, both, Each, each, It, it, Its, its, Many, many, Most, most, One, one, Our, our, Some, some, Such, such, The, the, Their, their, There, there, These, these, They, they, This, this, We, we, You, you, Your, your	`**`, `␣**`, `#`, `␣#`, `-`, `␣-`, `*`, `␣*`, `0`, `1`, `2`, `3`, `4`, `5`, `6`, `7`, `8`, `9`

Table 14: The propensity share under two aggregations: ratios of sums (the main-text convention) and per-prompt averaged rates, on the published excess-vocabulary list under the chat-template rendering. ‡: sign uncertain (one component's interval includes zero). A group is listed wherever a share is printed, and beside a share the floor rule of Appendix C.4 withholds (a dash); "same" = same-direction, "opp." = opposite-direction. r.o.s. = ratio of sums, p.p. = per prompt. 95% bootstrap intervals.

Transition	s_{prop} (ratio of sums)	s_{prop} (per prompt)	Group (r.o.s.)	Group (p.p.)
Model-written prompts				
Gemma-2-9B	0.48 [0.43, 0.55]	0.51 [0.45, 0.60]	same	same
Llama-3.1-8B	0.21 [0.12, 0.30]	0.24 [0.14, 0.35]	same	same
Mistral-7B-v0.3	0.31 [0.24, 0.40]	0.35 [0.27, 0.46]	same	same
OLMo-2-13B	0.29 [0.26, 0.32]	0.31 [0.28, 0.34]	same	same
OLMo-2-7B	0.39 [0.35, 0.42]	0.41 [0.38, 0.44]	same	same
Olmo-3 → RL-Zero-Code	0.27 [0.23, 0.34]	0.27 [0.23, 0.33]	opp.	opp.
Olmo-3 → Instruct‡	—	—	—	—
Qwen2.5-14B	0.11 [0.08, 0.14]	0.13 [0.09, 0.17]	same	same
Qwen2.5-1.5B‡	0.04 [0.01, 0.06]	0.002 [0.001, 0.048]	opp.	same
Qwen2.5-7B	0.14 [0.10, 0.18]	0.10 [0.03, 0.17]	opp.	opp.
Human-written prompts				
Gemma-2-9B	0.28 [0.24, 0.32]	0.31 [0.27, 0.36]	same	same
Llama-3.1-8B	0.15 [0.11, 0.19]	0.22 [0.15, 0.29]	same	same
Mistral-7B-v0.3	0.20 [0.15, 0.26]	0.27 [0.19, 0.35]	same	same
OLMo-2-13B	0.31 [0.29, 0.34]	0.36 [0.32, 0.39]	same	same
OLMo-2-7B	0.38 [0.36, 0.41]	0.43 [0.40, 0.45]	same	same
Olmo-3 → RL-Zero-Code	—	—	—	—
Olmo-3 → Instruct‡	—	—	—	same
Qwen2.5-14B	0.13 [0.09, 0.18]	0.18 [0.10, 0.25]	same	same
Qwen2.5-1.5B‡	0.023 [0.004, 0.043]	0.010 [0.001, 0.047]	opp.	same
Qwen2.5-7B‡	0.11 [0.06, 0.15]	0.08 [0.01, 0.15]	opp.	opp.

Instruct's fall resolves only under per-prompt averaging (same-direction, sign uncertain; the floor rule of Appendix C.4 withholds its share). Each has a component whose interval includes zero.

F Rendering, Population, and Generation-Cap Robustness

F.1 Bare vs. Chat-Template Rendering, with Per-Transition Floors

Does the census depend on prompting each predecessor under a chat template it was not trained on?

Table 15 evaluates the 8 released base-to-instruct transitions again with the predecessor prompted bare (as a completion model, with no chat template) while the successor keeps its own template. Each transition additionally draws its predecessor a second time at a different seed, giving that transition its own same-model floor: a comparison whose true change is zero. Under the bare rendering the published list rises on 7 of the 8 transitions. Under the mean of the two references, the change in reach is the larger component by point on every rising transition whose bare share is printed, and its interval lies entirely above one half on each of them except OLMo-2-7B; under both

Table 15: The eight rising census transitions under both predecessor renderings, with each transition's own same-model floor, and the two non-rising rows under the chat-template rendering alone. $\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ is the signed ratio (Section 3.1), computed under the successor reference; the floor is one template draw of a predecessor against a second draw, where the true change is zero. $*$ marks the Olmo-3 RL transition, whose floor and ratio are its connective class's own same-model floor over that class's change, since this row's counted change on the list is negative. Rates per 1,000 tokens with 95% bootstrap intervals; a dash marks a quantity the row does not define or that the floor rule of Appendix C.4 withholds. ‡: sign uncertain: one component's interval includes zero.

Transition	Rendering	Change per 1,000 tokens	$\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ (succ. ref.)	Larger component
Qwen2.5-7B	template	1.02 [0.79, 1.25]	−0.19 [−0.29, −0.12]	reach
	bare	1.19 [0.99, 1.41]	−0.33 [−0.42, −0.25]	reach
Qwen2.5-1.5B	template	1.77 [1.50, 2.06]	−0.04 [−0.06, −0.01]	reach
	bare	0.09 [−0.07, 0.24]	—	does not divide
Qwen2.5-14B	template	2.26 [1.99, 2.54]	0.11 [0.08, 0.14]	reach
	bare	1.94 [1.72, 2.19]	−0.13 [−0.18, −0.08]	reach
Llama-3.1-8B	template	1.03 [0.73, 1.32]	0.21 [0.12, 0.30]	reach
	bare	0.93 [0.67, 1.21]	0.23 [0.14, 0.33]	reach
Gemma-2-9B	template	2.79 [2.34, 3.28]	0.48 [0.43, 0.55]	reach
	bare	3.25 [2.83, 3.68]	0.41 [0.37, 0.46]	reach
Mistral-7B-v0.3	template	0.84 [0.67, 1.01]	0.31 [0.24, 0.40]	reach
	bare	0.80 [0.64, 0.97]	—	—
OLMo-2-13B	template	4.19 [3.82, 4.54]	0.29 [0.26, 0.32]	reach
	bare	4.08 [3.66, 4.47]	0.32 [0.29, 0.35]	reach
OLMo-2-7B	template	3.84 [3.47, 4.19]	0.39 [0.35, 0.42]	reach
	bare	3.91 [3.53, 4.30]	0.38 [0.34, 0.41]	reach
Olmo-3 → Instruct‡	template	−0.21 [−0.46, 0.03]	—‖	no rise
Olmo-3 → RL-Zero-Code	template	−0.56 [−0.78, −0.32]	−0.60 [−1.05, −0.42]	falls

Transition	Rendering	Floor per 1,000 tokens	Floor ÷ rise
Qwen2.5-7B	template	−0.04 [−0.18, 0.10]	0.04
	bare		0.03
Qwen2.5-1.5B	template	−0.07 [−0.20, 0.07]	0.04
	bare		0.74
Qwen2.5-14B	template	0.05 [−0.08, 0.18]	0.02
	bare		0.02
Llama-3.1-8B	template	0.08 [−0.09, 0.26]	0.08
	bare		0.09
Gemma-2-9B	template	−0.21 [−0.39, −0.04]	0.07
	bare		0.06
Mistral-7B-v0.3	template	0.08 [−0.02, 0.18]	0.10
	bare		0.10
OLMo-2-13B	template	0.03 [−0.08, 0.13]	0.01
	bare		0.01
OLMo-2-7B	template	−0.05 [−0.19, 0.09]	0.01
	bare		0.01
Olmo-3 → Instruct	template	—	—
Olmo-3 → RL-Zero-Code	template	0.04 [−0.05, 0.13] *	0.02 *

‖ Opposite-direction group (Appendix B.4); reach moves against the count; sign uncertain.

renderings at once the interval lies above one half on 5 transitions. Each transition's same-model floor is under a tenth of that transition's own bare rise on 6 of the 8, the exceptions being Mistral-7B-v0.3 (whose ratio is just above the tenth, so no bare share is printed) and Qwen2.5-1.5B (whose bare rise does not separate from zero).

Qwen2.5-1.5B is instructive: its template-rendered rise of 1.77 [1.50, 2.06] per 1,000 tokens falls to 0.09 [−0.07, 0.24] per 1,000 tokens (an interval spanning zero) once the predecessor is prompted bare, because the predecessor's own rate on this list roughly doubles when it is prompted the way it was trained. A reported counted change of a published list is a statement about the transition and about how the predecessor was prompted. The one census transition whose change in propensity changes sign between renderings, Qwen2.5-14B on model-written prompts, does so because under the bare rendering the predecessor is teacher-forced under a different prompt from the successor, so the two are matched only on the continuation.

Table 16: The 10 census transitions on human-written prompts, under both predecessor renderings, with each transition's own same-model floor. The 8 released base-to-instruct transitions divide as they do on the model-written population. $\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ is the signed ratio (Section 3.1) under the successor reference; a negative value means the change in propensity moves against the count. Rates per 1,000 tokens with 95% bootstrap intervals; a dash marks a quantity the row does not define (a counted change that does not separate from zero has no signed ratio). ‡: sign uncertain: one component's interval includes zero.

Transition	Bare counted change	Bare $\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ (succ. ref.)	Larger component
Qwen2.5-7B	0.94 [0.71, 1.16]	−0.32 [−0.46, −0.22]	reach
Qwen2.5-1.5B	0.65 [0.42, 0.90]	−0.10 [−0.20, −0.03]	reach
Qwen2.5-14B	1.52 [1.26, 1.80]	−0.03 [−0.09, 0.02]	reach
Llama-3.1-8B	2.14 [1.89, 2.40]	0.14 [0.10, 0.19]	reach
Gemma-2-9B	3.15 [2.78, 3.52]	0.28 [0.23, 0.32]	reach
Mistral-7B-v0.3	1.11 [0.95, 1.26]	0.19 [0.13, 0.25]	reach
OLMo-2-13B	4.26 [3.91, 4.61]	0.33 [0.30, 0.36]	reach
OLMo-2-7B	4.18 [3.86, 4.53]	0.40 [0.38, 0.43]	reach
Olmo-3 → Instruct	−0.14 [−0.33, 0.05]	—	no rise
Olmo-3 → RL-Zero-Code	0.22 [0.04, 0.38]	1.76 [0.99, 6.78]	propensity

Transition	Template counted change	Template $\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ (succ. ref.)	Larger component
Qwen2.5-7B	1.23 [1.01, 1.47]	−0.13 [−0.21, −0.07]	reach
Qwen2.5-1.5B	2.13 [1.85, 2.44]	−0.024 [−0.047, −0.003]	reach
Qwen2.5-14B	1.71 [1.45, 1.96]	0.13 [0.09, 0.18]	reach
Llama-3.1-8B	2.14 [1.87, 2.41]	0.15 [0.11, 0.19]	reach
Gemma-2-9B	3.31 [2.93, 3.73]	0.28 [0.24, 0.32]	reach
Mistral-7B-v0.3	1.06 [0.92, 1.22]	0.20 [0.15, 0.26]	reach
OLMo-2-13B	4.32 [3.99, 4.69]	0.31 [0.29, 0.34]	reach
OLMo-2-7B	4.27 [3.93, 4.62]	0.38 [0.36, 0.41]	reach
Olmo-3 → Instruct‡	−0.17 [−0.37, 0.03]	—‖	no rise
Olmo-3 → RL-Zero-Code	0.06 [−0.10, 0.22]	—‖	no rise

Transition	Floor	Fl. ÷ bare	Fl. ÷ template
Qwen2.5-7B	−0.05 [−0.18, 0.09]	0.05	0.04
Qwen2.5-1.5B	−0.02 [−0.17, 0.13]	0.03	0.01
Qwen2.5-14B	0.12 [−0.02, 0.26]	0.08	0.07
Llama-3.1-8B	−0.02 [−0.11, 0.07]	0.01	0.01
Gemma-2-9B	−0.05 [−0.17, 0.06]	0.01	0.01
Mistral-7B-v0.3	0.03 [−0.05, 0.11]	0.02	0.02
OLMo-2-13B	−0.07 [−0.16, 0.02]	0.02	0.02
OLMo-2-7B	0.02 [−0.10, 0.14]	0.01	0.01
Olmo-3 → Instruct	−0.04 [−0.14, 0.06]	0.26	0.22
Olmo-3 → RL-Zero-Code	0.00 [−0.11, 0.11]	0.02	0.07

‖ Opposite-direction group (Appendix B.4); reach moves against the count. Olmo-3-7B-Instruct is additionally sign uncertain.

F.2 Human-Written vs. Model-Written Prompt Populations

Does the census depend on the prompts being model-written?

The same 10 transitions are evaluated on 600 first user turns that a person wrote, drawn from a public release of arena preference data, deduplicated against two public Tülu-3 SFT mixtures and UltraChat-200k (Appendix D.2), through the same pipeline as the model-written population. Under the bare rendering the list rises on 9 of the 10 transitions. Among the rising transitions the change in reach is the larger component on 8, its interval lies entirely above one half on 8, and on 8 it does so under both renderings at once, on every released base-to-instruct transition. Each transition's same-model floor is under a tenth of that transition's bare rise on 9 of 10, the one exception being the row whose rise does not separate from zero.

Table 16 gives the full table. One row divides differently: the Olmo-3 RL transition's list rate rises by 0.22 [0.04, 0.38] per 1,000 tokens on the human-written prompts with both changes in propensity positive, where on the model-written population the same list falls. That result is narrow: under the chat template the same transition's counted change does not separate from zero (0.06 [−0.10, 0.22]). Under the successor reference and under the mean of the two, on none of the 8 released base-to-instruct transitions is the change in propensity the larger component on either population under either rendering; under the predecessor reference it is the larger component for OLMo-2-7B and OLMo-2-13B on model-written prompts (Table 2).

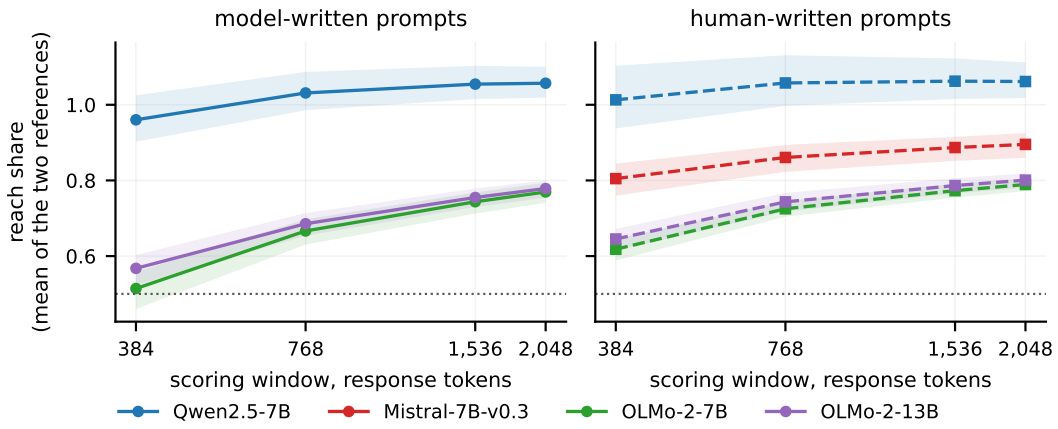


Figure 5: The reach share (mean of the two references) of the published list's rise climbs with the scoring window on every cell but one (Qwen2.5-7B on human-written prompts settles back at the last window by less than its own interval width). Each line is one census transition on nested windows of the same 2,048-token trajectories; the marker is the reach share of that window's rise with a 95% bootstrap band; the dotted line is one half. Left, model-written prompts; right, human-written. Mistral-7B-v0.3's model-written row is not drawn: the floor rule of Appendix C.4 withholds its share at every window.

F.3 384-Token vs. 2,048-Token Census

Does the census depend on the 384-token generation cap?

We re-generated 4 transitions (Qwen2.5-7B, Mistral-7B-v0.3, OLMo-2-7B and OLMo-2-13B, each against its instruction-tuned successor) at a 2,048-token cap on both prompt populations. The same trajectories were decomposed over nested windows of 384, 768, 1,536 and 2,048 response tokens (Table 17 and Figure 5).

The list still rises at the longer cap: the counted change at 2,048 tokens has an interval above zero on 4 of 4 transitions. On the model-written prompts, where the same-model floor is scored, the 384-token window of the new trajectories reproduces the census's 384-token reach share (the same side of one half) on 4 of 4 transitions, which confirms that the longer trajectories are a repeat of one measurement rather than a second measurement.

The reach share is reported here under the mean of the two references; the main-text census (Section 4.2) reports the signed ratio under the successor reference, and Table 2 gives both references' shares. On every cell of Table 17 the reach share at 2,048 tokens is above its own value at 384 tokens. The reach share rises with the window, and at 2,048 tokens part of the predecessor's window is held by rollouts still running at the cap, whose unique-token ratio is low (the degeneration block of Table 17).

Two things travel with the longer window. First, on each transition some of the predecessor's trajectories are still running when the window ends, and their unique-token ratio is low: the degeneration block of Table 17 gives the share of trajectories still running, their mean unique-token ratio, the share carrying a repeated block, and the share of the arm's tokens they hold. Second, Mistral-7B-v0.3's model-written rows carry no share at any window: that transition's same-model floor is 0.24 of its own rise at 2,048 tokens (above the bar of Appendix C.4 (iv)).

The census at the 2,048-token cap under the bare rendering. Table 18 gives the decomposition of the same four transitions at the 384- and 2,048-token windows of the 2,048-token trajectories, evaluated under the bare rendering. All eight cells keep their group (Appendix B.4) across both windows, and the propensity share is lower at 2,048 than at 384 tokens on six of the seven cells whose share is printed, all but Qwen2.5-7B on model-written prompts, where the difference includes zero.

Table 17: The published list's reach share over nested windows of the same 2,048-token trajectories, four census transitions on both prompt populations, at the two ends of the sweep; Figure 5 draws every window. The share at 2,048 tokens exceeds its value at 384 tokens on every cell, so the reach share rises with the window; at the longer cap, part of the predecessor's window is held by rollouts still running at the cap, whose unique-token ratio is low. The second block gives each transition's rise at the longer cap and the same-model floor over that rise; the third block is the degeneration table, computed on the predecessor's trajectories still running when the window ends; the loop share's denominator is the trajectories that reached the cap, not all trajectories. The reach share is one minus the $\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ under the mean of the two references; 95% bootstrap intervals; windows are nested prefixes of the same trajectories. Mistral-7B-v0.3's model-written rows are the transition whose same-model floor is not small, and they carry no share.

Transition	Prompts	384 response tokens	2,048 response tokens
Qwen2.5-7B	model-written	0.96 [0.90, 1.02]	1.06 [1.02, 1.10]
	human-written	1.01 [0.94, 1.10]	1.06 [1.02, 1.11]
Mistral-7B-v0.3	model-written	—	—
	human-written	0.80 [0.76, 0.84]	0.90 [0.86, 0.93]
OLMo-2-7B	model-written	0.51 [0.46, 0.56]	0.77 [0.74, 0.79]
	human-written	0.62 [0.59, 0.65]	0.79 [0.77, 0.81]
OLMo-2-13B	model-written	0.57 [0.52, 0.60]	0.78 [0.75, 0.80]
	human-written	0.65 [0.61, 0.67]	0.80 [0.78, 0.82]

Transition	Prompts	Rise at 2,048	Floor ÷ rise
Qwen2.5-7B	model-written	1.21 [0.98, 1.44]	0.05
	human-written	1.21 [0.88, 1.56]	0.07
Mistral-7B-v0.3	model-written	0.75 [0.50, 0.97]	0.24
	human-written	1.19 [0.96, 1.42]	0.01
OLMo-2-7B	model-written	4.41 [3.94, 4.88]	0.02
	human-written	4.59 [4.21, 4.98]	0.04
OLMo-2-13B	model-written	4.55 [4.09, 5.01]	0.07
	human-written	4.53 [4.07, 4.96]	0.02

Transition	Prompts	At the cap	Unique-token ratio	Loop share (of those at the cap)	Token share
Qwen2.5-7B	model-written	58 of 1,800 (3.2%)	0.18	1 of 58 (1.7%)	0.09
	human-written	149 of 1,800 (8.3%)	0.06	8 of 149 (5.4%)	0.35
Mistral-7B-v0.3	model-written	556 of 1,800 (30.9%)	0.10	23 of 556 (4.1%)	0.59
	human-written	851 of 1,800 (47.3%)	0.06	117 of 851 (13.7%)	0.78
OLMo-2-7B	model-written	498 of 1,800 (27.7%)	0.09	19 of 498 (3.8%)	0.63
	human-written	710 of 1,800 (39.4%)	0.07	55 of 710 (7.7%)	0.77
OLMo-2-13B	model-written	478 of 1,800 (26.6%)	0.12	16 of 478 (3.3%)	0.58
	human-written	685 of 1,800 (38.1%)	0.10	47 of 685 (6.9%)	0.74

F.4 The Genre Audit of the Predecessor's Bare Trajectories

Does the change in reach on bare-rendered predecessors reflect a difference in discourse genre between the two models' trajectories?

A predecessor model handed a bare user turn need not answer it. If its trajectories were mostly document continuations rather than answers, the change in reach would reflect a change in the kind of text the two models write, not a change in the contexts reached within one kind. Two classifiers were built to test this, sorting each bare-rendered predecessor trajectory into one of three classes: *answering* (the continuation addresses the user's request in an assistant register), *continuing* (it carries on the document or writes another user turn), and *other* (degenerate, empty, or neither register). Both classifiers use the same three definitions.

The first classifier is a cascade of surface clauses; the second is a greedy-decoding pass of a pinned open-weights model (Qwen/Qwen3-8B). Both were checked against a second reader on the same 120 trajectories. The second reader is also a model. Tables 19 and 20 give the two confusion matrices. Cohen's κ is 0.13 [0.01, 0.24] for the rule (0.42 observed agreement against 0.33 expected) and 0.31 [0.16, 0.47] for the judge (0.63 against 0.46), both below the 0.60, the agreement a subset analysis would need. The judge's disagreements concentrate on one boundary: 18 trajectories it labels as continuing the reader labels as answering, and 18 the other way about, of 44 disagreements in all.

Table 18: The four regenerated census transitions at the 384- and 2,048-token windows of the same 2,048-token trajectories, evaluated under the bare rendering. Columns give the counted change, the change in reach, the change in propensity, the propensity share (with its 95% bootstrap interval) and the group. $*$ marks a cell whose same-model floor is not below a tenth of that window's rise, so no share is printed. Rates per 1,000 tokens.

Transition	Pop.	Window	Δ_{gen}	Δ_{reach}	Δ_{prop}	s_{prop}	Group
Qwen2.5-7B	model-written	384	1.18	1.57	−0.38	0.20 [0.16, 0.24]	opposite-direction
Qwen2.5-7B	model-written	2,048	1.21	1.55	−0.34	0.18 [0.15, 0.21]	opposite-direction
Qwen2.5-7B	human-written	384	0.91	1.20	−0.29	0.19 [0.14, 0.25]	opposite-direction
Qwen2.5-7B	human-written	2,048	1.21	1.49	−0.28	0.16 [0.12, 0.20]	opposite-direction
OLMo-2-7B	model-written	384	3.92	2.42	1.50	0.38 [0.34, 0.43]	same-direction
OLMo-2-7B	model-written	2,048	4.41	3.47	0.94	0.21 [0.18, 0.24]	same-direction
OLMo-2-7B	human-written	384	4.18	2.46	1.72	0.41 [0.38, 0.44]	same-direction
OLMo-2-7B	human-written	2,048	4.59	3.23	1.36	0.30 [0.27, 0.32]	same-direction
Mistral-7B-v0.3	model-written	384	0.60	0.37	0.23	*	same-direction
Mistral-7B-v0.3	model-written	2,048	0.75	0.49	0.26	*	same-direction
Mistral-7B-v0.3	human-written	384	1.15	0.93	0.22	0.19 [0.12, 0.26]	same-direction
Mistral-7B-v0.3	human-written	2,048	1.19	1.02	0.18	0.15 [0.09, 0.21]	same-direction
OLMo-2-13B	model-written	384	4.16	2.95	1.21	0.29 [0.25, 0.33]	same-direction
OLMo-2-13B	model-written	2,048	4.55	3.80	0.74	0.16 [0.14, 0.19]	same-direction
OLMo-2-13B	human-written	384	4.28	2.86	1.42	0.33 [0.30, 0.36]	same-direction
OLMo-2-13B	human-written	2,048	4.53	3.45	1.07	0.24 [0.21, 0.27]	same-direction

Table 19: The surface rule's labels against the second reader's, on the same 120 trajectories. Rows are the rule's label and columns the reader's; the rule's rows are equal by construction, the reader's columns are not.

Rule's label	Reader: answering	Reader: continuing	Reader: other
Answering	24	15	1
Continuing	17	22	1
Other	27	9	4

Neither classifier reaches the agreement a subset analysis would need, so no class share, no subset recomputation and no before-and-after comparison enters this paper. The judge model's determinism check also failed: 1 of 200 re-run greedy labels differed. The quantity $\Delta_{\mathrm{prop}} = r_T(R_T) - r_B(R_T)$ uses no predecessor-generated trajectory, so a genre difference in those trajectories does not reach a statement made in Δ_{prop} alone; the predecessor share at matched contexts, $r_B(R_T)/r_T(R_T)$, shares that property (Appendix F.5).

F.5 The Predecessor Share at Matched Contexts

The change in propensity $\Delta_{\mathrm{prop}} = r_T(R_T) - r_B(R_T)$ uses only the successor's trajectories: $r_T(R_T)$ is the successor teacher-forced on its own stored trajectories, and $r_B(R_T)$ is the predecessor teacher-forced on the same trajectories. Both quantities are computed on the successor's trajectories, so a statement made in Δ_{prop} alone holds whatever genre the bare predecessor writes; the genre objection of Appendix F.4 does not reach it. The same is true of the predecessor share at matched contexts, $r_B(R_T)/r_T(R_T)$: the predecessor's rendering still matters (it determines the prompt under which the predecessor is teacher-forced), but only the successor's trajectories enter.

Table 21 gives $r_T(R_T)$, $r_B(R_T)$, Δ_{prop} and $r_B(R_T)/r_T(R_T)$ for every transition on the published excess-vocabulary list, across four grids (two prompt populations, each under both predecessor renderings). The printing condition of Appendix C.4 (iv) is met on all 38 cells.

Summary counts. Table 22 collects the interval-based counts per grid. The predecessor share at matched contexts has an interval lower bound at or above three-quarters on 8 of 10 transitions (model-written, chat-template rendering); on the three other grids the count is 6 (model-written bare, of 8), 7 (human-written bare) and 7 (human-written template). On every transition of every grid, the ratio's interval lower bound is at or above one half. On exactly 3 transitions per grid, the change in propensity has an interval at or below zero: Qwen2.5-7B, Qwen2.5-1.5B and Olmo-3-7B-Instruct on the three grids that carry the Olmo-3 transitions, with Qwen2.5-14B in place of Olmo-3-7B-Instruct

Table 20: The judge model's labels against the same reader's, on the same trajectories. Rows are the judge's label and columns the reader's.

Judge's label	Reader: answering	Reader: continuing	Reader: other
Answering	50	18	0
Continuing	18	23	3
Other	0	5	3

on the model-written bare grid, which has no Olmo-3 rows. The absolute change in propensity is at most half of $r_T(R_T)$ by interval on every transition.

Rendering sensitivity. Qwen2.5-14B on the model-written population is the one transition whose change in propensity lies on opposite sides of zero under the two renderings: Δ_{prop} is 0.25 [0.17, 0.32] per 1,000 tokens under the chat template and -0.24 [-0.33, -0.15] per 1,000 tokens under the bare prompt. On the human-written population the same transition's sign differs by point only, with an interval spanning zero under the bare prompt.

Bare-grid serialization. Under the bare rendering the two models are teacher-forced on the same stored continuation under different prompts: the predecessor sees a bare prompt and the successor its own chat template. The change in propensity on the bare grids therefore also contains the predecessor's prompt-format effect at fixed text. That effect is at most 0.08 per 1,000 tokens on model-written prompts and 0.07 per 1,000 tokens on human-written ones, except on Qwen2.5-7B (0.19 per 1,000 tokens model-written) and Qwen2.5-14B (0.49 per 1,000 tokens model-written). On Qwen2.5-14B the bare sign reversal disappears under identical prefixes: the chat-template change in propensity is 0.25 [0.17, 0.32] per 1,000 tokens, positive, against the bare value above. Under the chat-template rendering both models see identical token prefixes, so the chat-template grids are free of this effect.

The Olmo-3 RL transition on this grid. On the connective class, the Olmo-3 RL transition's change in propensity is -0.25 [-0.27, -0.22] per 1,000 tokens (the predecessor assigning 3.70 per 1,000 tokens against the successor's 3.45), for a predecessor share at matched contexts of 1.07 [1.06, 1.08]: the predecessor assigns more than the successor at the successor's own contexts, while the counted change is $+2.71$, a 4.7-fold increase.

G Additional Case-Study Results

G.1 The Olmo-3 RL Transition, Full Decomposition

Table 24 gives five OLMo-2 and Olmo-3 transitions on the screened prompts, with the counted change, the change in reach and the change in propensity computed under both references. On the top row the change in reach exceeds the counted change and the change in propensity is below zero under both references. Table 25 extends the same decomposition to 13 transition-by-population cells, 11 on the model-written prompts and two on the screened prompts. Olmo-3-7B-RL-Zero-Mix, trained from the same predecessor on a mixed task set, shows an equally large counted change, 2.73 [2.62, 2.83] per 1,000 tokens, with a change in propensity of -0.17 [-0.19, -0.15] that turns positive under the predecessor reference (0.04 [0.03, 0.06]; Table 25). The components move in opposite directions: propensity accounts for 0.05 [0.05, 0.06] of the gross movement (Appendix B.4) and moves against the count. At the response's first token, where both models are given the same prompt, the successor is more likely than the predecessor to open with a plan-opening marker: the markers' rate rises by 1.72 [1.67, 1.78] per 1,000 tokens on the model-written prompts, and their change in reach is -0.01 [-0.02, -0.01] of that rise there and -0.04 [-0.06, -0.03] on the screened prompts, below zero on both, so the change in propensity carries slightly more than all of it.

The reversal also holds on the human-written prompts, under both renderings (Table 26): the counted change is above zero and the change in propensity is below zero under both references, with $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ of -0.13 [-0.15, -0.11] under the chat-template rendering, against -0.09 [-0.10, -0.08] on the model-written prompts; under the bare rendering the floor rule of Appendix C.4 with-

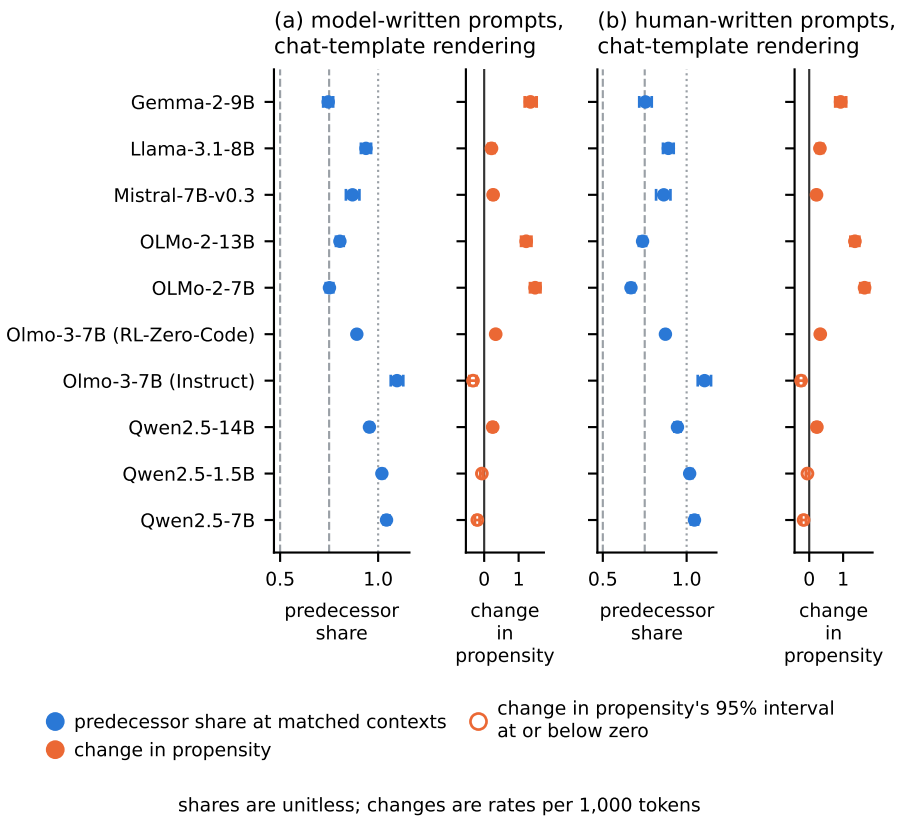


Figure 6: The predecessor share at matched contexts, $r_B(R_T)/r_T(R_T)$, for every census transition under the chat-template rendering. This quantity uses no predecessor-generated trajectory: neither $r_B(R_T)$ nor $r_T(R_T)$ depends on the predecessor's own trajectories. The dashed line marks one half; every transition sits above it. Table 21 gives the numbers.

holds it. It survives removing "First" (Table 27): "First" carries 0.34 [0.29, 0.39] of the class's successor-reference change in propensity on the model-written prompts and 0.59 [0.53, 0.65] on the human-written prompts under the chat template, below one on both, and 0.70 [0.67, 0.72] and 0.75 [0.70, 0.80] of the class's counted change on the same two cells, also below one, so the remaining word forms together keep a counted change above zero and a change in propensity below zero; under the bare rendering the floor rule withholds these shares.

Table 28 summarizes every case-study transition with the component that carries each counted change and the principal limitation of each.

G.2 Where the Olmo-3 Divergence Sits Within the Response

Figure 7 shows that the Olmo-3 RL transition's divergence from its predecessor is dense at the response opening without being confined there. Figure 7 gives the full position-class profile. Paragraph 1's first sentence holds 0.11 [0.11, 0.11] of all the positive divergence on a small fraction of the tokens, while the break token and the counted connective opener each hold under one percent. Paragraphs after the second (the figure's *later* class) hold 0.60 [0.59, 0.60] of it, spread over most of the tokens, and what sits there is mostly not structural: formatting, punctuation, mathematics, code and sentence-initial positions together hold 0.31 [0.31, 0.32], while ordinary content-word positions hold 0.69 [0.68, 0.69] at a low excess per token, at 0.9 times those paragraphs' mean excess per token against 2.2 at sentence openings. 0.24 [0.23, 0.24] of the divergence lies before the first paragraph break on the model-written prompts and 0.27 [0.27, 0.28] on the screened ones.

On the connectives the successor itself wrote at the second paragraph's opening we cannot tell what it assigns apart from what the predecessor does, and the mean log-likelihood ratio there is 0.002

Table 21: The predecessor share at matched contexts: the published excess-vocabulary list on every census transition, across two prompt populations and two predecessor renderings. This quantity uses no predecessor-generated trajectory; the predecessor's rendering still determines the prompt under which it is teacher-forced. $r_T(R_T)$ and $r_B(R_T)$ are rates per 1,000 tokens; $\Delta_{\text{prop}} = r_T(R_T) - r_B(R_T)$ is the change in propensity; $r_B(R_T)/r_T(R_T)$ is the predecessor share at matched contexts. 95% bootstrap intervals. The printing condition of Appendix C.4 (iv) is met on all 38 cells.

Transition	$r_T(R_T)$	$r_B(R_T)$	Δ_{prop} per 1,000 tokens	$r_B(R_T)/r_T(R_T)$
Model-written prompts, predecessor rendered under the chat template				
Gemma-2-9B	5.28	3.93	1.34 [1.18, 1.52]	0.75 [0.72, 0.77]
Llama-3.1-8B	3.41	3.20	0.21 [0.12, 0.30]	0.94 [0.91, 0.96]
Mistral-7B-v0.3	1.99	1.73	0.26 [0.19, 0.33]	0.87 [0.84, 0.90]
OLMo-2-13B	6.25	5.03	1.22 [1.07, 1.37]	0.80 [0.78, 0.82]
OLMo-2-7B	5.96	4.48	1.48 [1.33, 1.63]	0.75 [0.73, 0.77]
Olmo-3 → RL-Zero-Code	3.08	2.74	0.33 [0.31, 0.36]	0.89 [0.88, 0.90]
Olmo-3 → Instruct	3.37	3.70	−0.32 [−0.43, −0.23]	1.10 [1.06, 1.13]
Qwen2.5-14B	5.58	5.33	0.25 [0.17, 0.32]	0.96 [0.94, 0.97]
Qwen2.5-1.5B	3.51	3.58	−0.07 [−0.11, −0.02]	1.02 [1.01, 1.03]
Qwen2.5-7B	4.65	4.85	−0.20 [−0.27, −0.13]	1.04 [1.03, 1.06]

Transition	$r_T(R_T)$	$r_B(R_T)$	Δ_{prop} per 1,000 tokens	$r_B(R_T)/r_T(R_T)$
Model-written prompts, predecessor prompted bare (8 transitions; two Olmo-3 transitions absent)				
Gemma-2-9B	5.28	3.94	1.33 [1.17, 1.50]	0.75 [0.72, 0.77]
Llama-3.1-8B	3.41	3.20	0.21 [0.12, 0.30]	0.94 [0.91, 0.96]
Mistral-7B-v0.3	1.99	1.72	0.27 [0.19, 0.33]	0.87 [0.83, 0.90]
OLMo-2-13B	6.25	4.96	1.30 [1.15, 1.44]	0.79 [0.77, 0.81]
OLMo-2-7B	5.96	4.49	1.47 [1.31, 1.63]	0.75 [0.73, 0.77]
Qwen2.5-14B	5.58	5.82	−0.24 [−0.33, −0.15]	1.04 [1.03, 1.06]
Qwen2.5-1.5B	3.51	3.62	−0.11 [−0.15, −0.07]	1.03 [1.02, 1.04]
Qwen2.5-7B	4.65	5.04	−0.39 [−0.46, −0.32]	1.08 [1.07, 1.10]

Transition	$r_T(R_T)$	$r_B(R_T)$	Δ_{prop} per 1,000 tokens	$r_B(R_T)/r_T(R_T)$
Human-written prompts, predecessor prompted bare				
Gemma-2-9B	3.77	2.89	0.88 [0.71, 1.04]	0.77 [0.73, 0.81]
Llama-3.1-8B	2.90	2.59	0.31 [0.22, 0.40]	0.89 [0.86, 0.92]
Mistral-7B-v0.3	1.57	1.36	0.21 [0.14, 0.28]	0.87 [0.82, 0.91]
OLMo-2-13B	5.13	3.73	1.39 [1.25, 1.54]	0.73 [0.71, 0.75]
OLMo-2-7B	4.92	3.24	1.67 [1.54, 1.82]	0.66 [0.64, 0.68]
Olmo-3 → RL-Zero-Code	2.57	2.19	0.38 [0.35, 0.42]	0.85 [0.84, 0.86]
Olmo-3 → Instruct	2.24	2.55	−0.31 [−0.40, −0.22]	1.14 [1.10, 1.18]
Qwen2.5-14B	4.13	4.18	−0.05 [−0.13, 0.03]	1.01 [0.99, 1.03]
Qwen2.5-1.5B	2.88	2.94	−0.06 [−0.11, −0.02]	1.02 [1.01, 1.04]
Qwen2.5-7B	3.53	3.83	−0.30 [−0.38, −0.22]	1.09 [1.06, 1.11]

Transition	$r_T(R_T)$	$r_B(R_T)$	Δ_{prop} per 1,000 tokens	$r_B(R_T)/r_T(R_T)$
Human-written prompts, predecessor rendered under the chat template				
Gemma-2-9B	3.77	2.84	0.93 [0.77, 1.09]	0.75 [0.72, 0.79]
Llama-3.1-8B	2.90	2.58	0.32 [0.22, 0.41]	0.89 [0.86, 0.92]
Mistral-7B-v0.3	1.57	1.36	0.21 [0.15, 0.28]	0.86 [0.82, 0.90]
OLMo-2-13B	5.13	3.78	1.35 [1.22, 1.48]	0.74 [0.72, 0.76]
OLMo-2-7B	4.92	3.28	1.63 [1.51, 1.76]	0.67 [0.65, 0.69]
Olmo-3 → RL-Zero-Code	2.57	2.25	0.33 [0.30, 0.36]	0.87 [0.86, 0.88]
Olmo-3 → Instruct	2.24	2.48	−0.24 [−0.33, −0.15]	1.11 [1.07, 1.15]
Qwen2.5-14B	4.13	3.90	0.23 [0.15, 0.31]	0.95 [0.93, 0.96]
Qwen2.5-1.5B	2.88	2.93	−0.05 [−0.10, −0.01]	1.02 [1.00, 1.03]
Qwen2.5-7B	3.53	3.70	−0.17 [−0.24, −0.09]	1.05 [1.03, 1.07]

Table 22: Summary counts for the predecessor share at matched contexts of the published excess-vocabulary list. Each row is one population-by-rendering grid; the columns count transitions meeting various interval-based criteria. Counts are based on the predecessor share at matched contexts ($r_B(R_T)/r_T(R_T)$) and the change in propensity (Δ_{prop}). The two Olmo-3 transitions are absent from the model-written bare grid, so that grid has 8 transitions.

Grid	n	$\frac{r_B(R_T)}{r_T(R_T)} \geq 0.75$	$\frac{r_B(R_T)}{r_T(R_T)} \geq 0.5$	$\Delta_{\text{prop}} \leq 0$	$\lvert\Delta_{\text{prop}}\rvert \leq \frac{1}{2} r_T(R_T)$
Mod.-wr., tmpl.	10	8	10	3	10
Mod.-wr., bare	8	6	8	3	8
Human-wr., bare	10	7	10	3	10
Human-wr., tmpl.	10	7	10	3	10

[−0.003, 0.008] nats on the model-written prompts over 3,813 scored positions. Over every scored position of that class the change in propensity is negative under both references on both populations, and the whole 1.91 [1.86, 1.96] per 1,000 tokens rise in connective openers there is the change in

Table 23: Counted change and its decomposition for four case studies on three transitions, with response termination at three generation caps (per 1,000 tokens). Δ^{B}_{prop} is the change in propensity under the predecessor reference, and $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ the signed ratio (successor reference). Two lines per row: point estimates, then 95% intervals. ‡: a component's interval includes zero.

Case	Δ_{gen}	Δ_{reach}	Δ_{prop}	Δ^{B}_{prop}	$\Delta_{\text{prop}}/\Delta_{\text{gen}}$
Olmo-3 connectives	2.71	2.96	−0.25	−0.03	−0.09
	[2.60, 2.82]	[2.85, 3.07]	[−0.27, −0.22]	[−0.04, −0.02]	[−0.10, −0.08]
MiMo ordinal	1.42	1.14	0.28	0.30	0.20
	[1.14, 1.75]	[0.88, 1.45]	[0.19, 0.36]	[0.22, 0.37]	[0.14, 0.26]
Llama connectives	−1.06	−0.36	−0.69	−0.69	0.66
	[−1.18, −0.92]	[−0.47, −0.25]	[−0.74, −0.64]	[−0.73, −0.64]	[0.60, 0.74]
Termination, 384	−0.34	−0.33	−0.01	−0.10	0.03
	[−0.38, −0.29]	[−0.37, −0.28]	[−0.02, −0.01]	[−0.11, −0.08]	[0.02, 0.05]
Termination, 2,048	−0.09	0.04	−0.12	−0.10	1.41‡
	[−0.15, −0.03]	[−0.03, 0.09]	[−0.13, −0.11]	[−0.11, −0.09]	[0.81, 4.11]
Termination, 8,192	0.27	0.39	−0.12	−0.04	−0.44
	[0.23, 0.31]	[0.34, 0.43]	[−0.13, −0.11]	[−0.05, −0.04]	[−0.51, −0.37]

Table 24: The five OLMo-2 and Olmo-3 transitions, decomposed (screened prompts, all scored positions). The second column is the counted change and the columns after it show which component it came from. Change in reach is the predecessor's own conditionals evaluated at the successor's contexts; predecessor-reference change in propensity is the successor's change evaluated on the *predecessor's* trajectories; the successor-reference change is that plus the residual column, and the last column is the mean of the two references. On the top row the change in reach is larger than the counted change and the change in propensity is below zero under both references. Counted changes carry 95% intervals; the component columns are shares of the row's own counted change; OLMo-2-7B SFT pools three generation seeds; the counted change of OLMo-2-1B SFT → DPO is negative (†), so its four columns give rates per 1,000 tokens instead of shares. A dash marks a share the floor rule of Appendix C.4 withholds, except in OLMo-2-7B SFT's last column, where the mean of the two references is not computed.

Transition	Δ_{gen} (per 1,000 tokens)	Δ_{reach}	Δ^{B}_{prop}	residual	Δ_{prop} (mean)
Olmo-3 → RL-Zero-Code	2.27 [2.06, 2.46]	1.11	−0.02	−0.09	−0.06
Olmo-3 → RL-Zero-Mix	1.89 [1.69, 2.07]	1.07	0.02	−0.08	−0.03
OLMo-2-1B SFT	0.56 [0.30, 0.92]	—	—	—	—
OLMo-2-7B SFT	0.45 [0.18, 0.66]	1.04	0.13	−0.17	—
OLMo-2-1B SFT → DPO†	−0.28 [−0.65, −0.05]	−0.26	−0.08	0.06	−0.05

reach (1.03 [1.03, 1.04] of it). After such a paragraph the predecessor supplies one word above all others: 0.99 [0.99, 1.00] of the connectives it writes continuing that paragraph are "first", the opening item of the announced list. The first-break decision moves on the predecessor's own first paragraphs as well: the predecessor opens a break at 0.96 per 1,000 tokens on its own first paragraphs and the successor at 4.68 on those same paragraphs.

G.3 Scoping the Reversal to RL from a Base: the Released Instruct Chain

Table 10 describes the Olmo-3 RL transition as reinforcement learning from a base. The released Olmo-3-7B instruction-tuned chain (predecessor → Instruct-SFT → Instruct-DPO → Instruct-RL) is decomposed stage by stage on the same connective class (Table 29). Olmo-3-7B-RL-Zero-Code announces a plan in its first paragraph on 4,608 of the 5,321 responses with a first paragraph break (87%), against the predecessor's 127 of 2,027 (6%). Neither the instruction-tuned stages nor the released RL stage produces the plan-announcing paragraph: their counts are 1 of 3,471 (<1%), 32 of 4,287 (1%) and 36 of 4,344 (1%), respectively. The change in propensity is positive under both references on both the preference stage (0.16 [0.12, 0.19] successor-reference, 0.14 [0.08, 0.21] predecessor-reference) and the released RL stage (0.09 [0.07, 0.12] successor-reference, 0.09 [0.07, 0.11] predecessor-reference), the opposite sign from Olmo-3-7B-RL-Zero-Code's. Stage and recipe differ together, so this identifies no single cause; it is a counterexample to universality and the reason the main text scopes the reversal to reinforcement learning from a base.

Table 25: Thirteen transition × population cells, 11 on the model-written prompts and two on the screened prompts. On the Olmo-3 RL rows the change in reach exceeds the counted change; on the Llama SimpleRL row it does not. Each row is one transition on one prompt population. *mod.-wr.* is the model-written population and *screened* the screened population. The second column is the counted change per 1,000 tokens; the four middle columns are the change in reach, the predecessor-reference change in propensity and the residual (the last two sum to the successor-reference change), and the mean of the two references, each as a share of that row's own counted change with its 95% interval beneath it. Site prev. is the count of responses carrying the measured scored position at all, with the percentage, predecessor on the point line and successor on the interval line; percentages are of each model's 5,400 trajectories. Every row is computed on the connective class except the two under "Formatting-marker set"; a model name too long for one line continues on the row's second line. The two screened rows are a second generation draw on the screened population; their counted changes differ by sampling from the first draw's, which Table 28 and Figure 12 print. †: the counted change's interval includes zero, so no share is defined, or the counted change is negative, so the four columns give rates per 1,000 tokens instead. An unmarked dash marks a share or rate the floor rule of Appendix C.4 withholds.

Transition	Pop.	Δ_{gen}	Δ_{reach}	Δ^{B}_{prop}	residual	Δ_{prop} (mean)	Site prev.
Connective class							
Olmo-3 →	mod.-wr.	2.71	1.09	−0.01	−0.08	−0.05	1,138 (21.1%)
RL-Zero-Code		[2.60, 2.82]	[1.08, 1.10]	[−0.01, −0.01]	[−0.09, −0.07]	[−0.06, −0.05]	4,854 (89.9%)
Olmo-3 →	mod.-wr.	2.73	1.06	0.02	−0.08	−0.02	1,138 (21.1%)
RL-Zero-Mix		[2.62, 2.83]	[1.05, 1.07]	[0.01, 0.02]	[−0.09, −0.07]	[−0.03, −0.02]	4,615 (85.5%)
OLMo-2-1B SFT	mod.-wr.	0.38	—	—	—	—	188 (3.5%)
		[0.28, 0.49]					1,209 (22.4%)
OLMo-2-1B SFT	mod.-wr.	−0.14	−0.12	−0.15	0.13	−0.09	1,209 (22.4%)
→ DPO†		[−0.22, −0.06]	[−0.20, −0.04]	[−0.20, −0.11]	[0.10, 0.17]	[−0.12, −0.06]	1,465 (27.1%)
OLMo-2-7B SFT†	mod.-wr.	0.09	—	—	—	—	622 (11.5%)
		[−0.04, 0.21]					1,040 (19.3%)
Qwen2.5-7B	mod.-wr.	−0.43	—	—	—	—	1,462 (27.1%)
SimpleRL		[−0.61, −0.24]					1,684 (31.2%)
Qwen2.5-7B	screened	−0.37	−0.32	−0.028	−0.032	−0.04	1,898 (35.1%)
SimpleRL†		[−0.61, −0.14]	[−0.55, −0.08]	[−0.061, 0.005]	[−0.064, 0.001]	[−0.07, −0.02]	2,676 (49.6%)
Llama-3.1-8B	mod.-wr.	−1.06	−0.36	−0.69	−0.01	−0.69	207 (3.8%)
SimpleRL†		[−1.18, −0.92]	[−0.47, −0.25]	[−0.73, −0.64]	[−0.04, 0.02]	[−0.73, −0.64]	245 (4.5%)
Mistral-7B-v0.1	mod.-wr.	−0.63	−0.15	−0.31	−0.17	−0.40	723 (13.4%)
SimpleRL†		[−0.93, −0.32]	[−0.45, 0.16]	[−0.35, −0.27]	[−0.21, −0.13]	[−0.43, −0.36]	1,768 (32.7%)
Tülu-3-8B SFT†	mod.-wr.	−0.01	—	—	—	—	13 (0.2%)
		[−0.12, 0.10]					1,192 (22.1%)
Tülu-3-8B SFT	screened	0.42	1.26	0.12	−0.38	−0.072	130 (2.4%)
		[0.24, 0.62]	[1.12, 1.51]	[0.06, 0.23]	[−0.71, −0.22]	[−0.176, −0.002]	2,030 (37.6%)
Formatting-marker set							
OLMo-2-1B SFT	mod.-wr.	19.00	0.97	0.04	0.00	0.03	188 (3.5%)
		[16.36, 21.70]	[0.96, 0.98]	[0.03, 0.04]	[−0.01, 0.01]	[0.03, 0.04]	1,209 (22.4%)
OLMo-2-1B SFT	mod.-wr.	11.15	0.95	0.06	−0.014	0.06	1,209 (22.4%)
→ DPO		[9.66, 12.62]	[0.94, 0.96]	[0.05, 0.08]	[−0.025, −0.003]	[0.04, 0.07]	1,465 (27.1%)

G.4 A Second Family: MiMo, Ordinal/Sequential Class

MiMo-7B-RL-Zero is a case in which the counted change has the same sign as the change in propensity but overstates its magnitude. The decomposition covers both the mathematics and the model-written populations (Table 30), the paragraph exchange (Table 31), and the context intervention on the text-defined window.

The ordinal class ("first", "second", "third", "next", "then" and "finally") has 4 surface variants per form, giving 24 ids. The class rises on both populations and the change in reach carries most of the rise: 1.42 [1.14, 1.75] per 1,000 tokens on the mathematics prompts, of which 0.80 [0.74, 0.86] is the change in reach, and 1.51 [1.30, 1.72] on the model-written ones. The change in propensity does not reverse here: it is positive under both references, 0.28 [0.19, 0.36] per 1,000 tokens evaluated at the successor's contexts and 0.30 [0.22, 0.37] at the predecessor's, and larger again on the model-written prompts. On the model-written population, propensity carries 0.41 [0.36, 0.47] of the counted change. The control class mapped by the same rule moves the other way under the mean

Table 26: The Olmo-3 RL transition's connective class on the human-written prompts under both renderings, beside the model-written prompts (chat-template rendering; the top row of Table 25). The counted change is above zero and the change in propensity below zero under both references in all three rows. Rates per 1,000 tokens with 95% intervals; the four cells, points only, are each model teacher-forced on each model's trajectories, b the predecessor and t the successor, so the first and last cells are each model on its own trajectories. A dash marks a share the floor rule of Appendix C.4 withholds.

Prompts, rendering	b on b	b on t	t on b	t on t	Counted change	Change in reach
model-written, chat template	0.74	3.70	0.71	3.45	2.71 [2.60, 2.82]	2.96 [2.85, 3.07]
human-written, bare	1.86	3.03	1.67	2.68	0.81 [0.66, 0.99]	1.17 [1.01, 1.35]
human-written, chat template	0.93	2.90	0.89	2.68	1.74 [1.61, 1.89]	1.97 [1.83, 2.12]

Prompts, rendering	Δ_{prop} (succ. ref.)	$\Delta^{B}_{\mathrm{prop}}$ (pred. ref.)	$\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ (succ. ref.)
model-written, chat template	−0.25 [−0.27, −0.22]	−0.03 [−0.04, −0.02]	−0.09 [−0.10, −0.08]
human-written, bare	−0.36 [−0.39, −0.33]	−0.20 [−0.22, −0.17]	—
human-written, chat template	−0.23 [−0.25, −0.21]	−0.05 [−0.06, −0.04]	−0.13 [−0.15, −0.11]

Table 27: The Olmo-3 RL transition's successor-reference change in propensity in the connective class, split by word form, on the model-written prompts (chat-template rendering) and the human-written prompts (both renderings). The upper block gives each form's change in propensity per 1,000 tokens; the five rows after "First" add up to the class with "First" removed. The lower block gives each form's share of the class's change in propensity; the share of "First" is below one where it is printed, and a dash marks a share the floor rule of Appendix C.4 withholds. "first" is the line-initial lowercase form; the pooled row holds the class's remaining ids. 95% intervals.

Word form	model-written, chat template	human-written, bare	human-written, chat template
Class, all ids	−0.25 [−0.27, −0.22]	−0.36 [−0.39, −0.33]	−0.23 [−0.25, −0.21]
"First"	−0.08 [−0.10, −0.07]	−0.23 [−0.26, −0.21]	−0.13 [−0.15, −0.12]
"Also"	−0.11 [−0.12, −0.09]	−0.08 [−0.09, −0.06]	−0.04 [−0.06, −0.03]
"Next"	−0.08 [−0.09, −0.07]	−0.02 [−0.03, −0.01]	−0.05 [−0.05, −0.04]
"Then"	0.05 [0.04, 0.06]	−0.003 [−0.010, 0.005]	0.03 [0.02, 0.04]
"first"	0.000 [−0.001, 0.000]	0.000 [−0.002, 0.001]	0.000 [−0.001, 0.001]
Remaining ids, pooled	−0.02 [−0.03, −0.02]	−0.02 [−0.03, −0.02]	−0.03 [−0.04, −0.02]

Share of the class's change	model-written, chat template	human-written, bare	human-written, chat template
"First"	0.34 [0.29, 0.39]	—	0.59 [0.53, 0.65]
"Also"	0.44 [0.39, 0.50]	0.21 [0.18, 0.25]	0.19 [0.14, 0.24]
"Next"	0.32 [0.28, 0.36]	0.05 [0.04, 0.07]	0.21 [0.18, 0.24]
"Then"	−0.20 [−0.25, −0.16]	0.01 [−0.01, 0.03]	−0.12 [−0.16, −0.09]
"first"	0.002 [0.000, 0.004]	0.001 [−0.002, 0.004]	0.002 [−0.003, 0.006]
Remaining ids, pooled	0.09 [0.07, 0.11]	0.07 [0.05, 0.09]	0.13 [0.11, 0.16]

of the two references (−2.65 [−2.85, −2.45] per 1,000 tokens on the mathematics prompts), so the positive term belongs to this class rather than to the mapping.

The paragraph exchange moves the class with the supplied paragraph (Table 31). Either model opens most next paragraphs with an ordinal on these prompts (the predecessor does on 0.66 of its continuations of its own first paragraphs) so the predecessor-to-successor difference that m and $\bar{m}$ divide by is small, 0.008 [0.005, 0.011] on the projection the exchange is scored by. Within that difference the transfer is complete in both directions, $\bar{m} = 1.05$ [0.95, 1.20] with the predecessor continuing and $m = 1.00$ [0.87, 1.15] with the successor continuing, and the raw class-opener share gives the same reading on wider intervals. A small difference transferred completely is weaker evidence than a large one; what the exchange establishes here is the direction.

The context intervention on the text-defined window confirms the direction. The eligible subset is 2,911 of 3,600 pairs over 391 of 400 prompts. The successor handed the predecessor's first paragraph opens its next paragraphs with the class at 2.38 [2.29, 2.47] per 1,000 tokens of that window, against 2.54 [2.44, 2.64] unmodified and 2.39 [2.31, 2.48] for the predecessor, moving the rate 1.09 of the gap, just past the predecessor by point (folded C 0.91 [0.38, 1.00] of the predecessor-to-successor gap).

Table 28: Every case-study transition, with the component the counted change is carried by. Counted changes are rate changes per 1,000 tokens, with 95% intervals where the measurement stores one and points for the three formatting-marker rows on the OLMo transitions. *scr.*, *m.w.* and *math.* are the screened, model-written and mathematics prompts, *conn.* the connective class; *larger component* names the larger of the two components, and a dash marks a cell the floor rule of Appendix C.4 withholds.

transition	pr.	event	counted change	larger component	principal limitation
OLMo-2-1B SFT	scr.	conn.	0.56 [0.30, 0.92]	—	floor a third of it
OLMo-2-7B SFT	scr.	conn.	0.45 [0.18, 0.66]	reach	—
OLMo-2-1B SFT → DPO	scr.	conn.	−0.28 [−0.65, −0.05]	—	a fall; one reference
Tülu-3-8B SFT	scr.	conn.	0.47 [0.28, 0.66]	reach	zero on model-written
OLMo-2-1B SFT	scr.	markers	4.88	—	one reference
OLMo-2-1B SFT → DPO	scr.	markers	10.27	reach	one reference
Olmo-3 → RL-Zero-Code	scr.	markers	−5.61	—	a fall
Llama-3.1-8B	m.w.	markers	25.20 [22.16, 28.13]	reach	predecessor template
Llama-3.1-8B SimpleRL	m.w.	conn.	−1.06 [−1.18, −0.92]	propensity	a fall; the count is right
Mistral-7B-v0.1 SimpleRL	m.w.	conn.	−0.63 [−0.93, −0.32]	not separable	control falls harder
Qwen2.5-7B SimpleRL	m.w.	conn.	−0.43 [−0.61, −0.24]	—	—
Olmo-3 → RL-Zero-Mix	m.w.	conn.	2.73 [2.62, 2.83]	reach	mean of the two references only
MiMo-7B	math.	ordinals	1.42 [1.14, 1.75]	reach	change in propensity positive; count right in direction

Table 29: Formatting markers on three transitions outside the released Olmo-3-7B instruction-tuned chain, then the chain itself. The first block gives the formatting markers (Table 13) on OLMo-2-1B SFT, OLMo-2-1B SFT → DPO and Olmo-3 → RL-Zero-Code; a dash marks a share the floor rule of Appendix C.4 withholds. The rows below it give the chain, stage by stage: the counted change in the connective class and the share of each stage's responses with a first paragraph break that open with a plan-announcing paragraph. Olmo-3-7B-RL-Zero-Code announces a plan on 4,608 of the 5,321 responses with a first paragraph break (87%); neither the SFT, DPO nor released RL stage produces it. The second and third stages' changes in propensity are positive, the opposite sign from Olmo-3-7B-RL-Zero-Code's.

Quantity	Statistic	Value
Markers, OLMo-2-1B SFT	counted change	4.88
Markers, OLMo-2-1B SFT	reach share	—
Markers, OLMo-2-1B SFT	$\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (mean of the two references)	—
Markers, OLMo-2-1B SFT → DPO	$\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (mean of the two references)	0.07 [0.06, 0.08]
Markers, RL-Zero-Code	counted change	−5.61
Markers, RL-Zero-Code	reach share	—
Markers, RL-Zero-Code	$\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (mean of the two references)	—
Instruct chain, predecessor → SFT	counted change	0.04 [−0.10, 0.17]
Instruct chain, SFT → DPO	counted change	0.11 [0.01, 0.20]
Instruct chain, DPO → RL	counted change	0.07
Instruct chain	predecessor plan count	127 of 2,027 (6.3%)
Instruct chain	SFT plan count	1 of 3,471 (<0.1%)
Instruct chain	DPO plan count	32 of 4,287 (0.7%)
Instruct chain	RL plan count	36 of 4,344 (0.8%)
RL-Zero-Code	plan count	4,608 of 5,321 (86.6%)

What this family adds is one half of the Olmo-3 finding. The class rises, the change in reach carries most of that rise, and the behavior follows the paragraph the model is handed. The change in propensity is positive, so reading the counted change on this transition as a change in propensity would have had the direction right and the size wrong. Across families it is the sign of the change in propensity that varies, not the size of the change in reach.

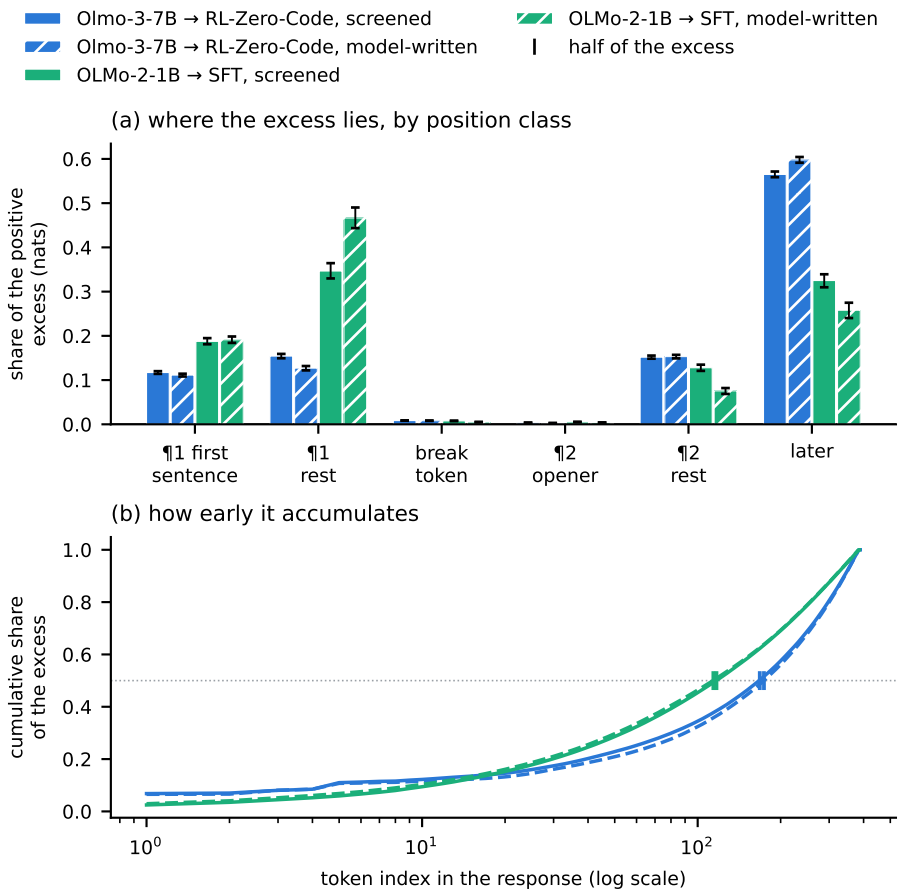


Figure 7: Where the two models diverge along the response. Paragraph 1's first sentence holds a tenth of the whole excess on a small fraction of the tokens, while the break token and the counted connective opener each hold under one percent. **(a)** Six position classes' shares of the positive log-likelihood difference along the successor's trajectories, color the transition and hatch the population; **(b)** the cumulative share by token index, the tick where half has accumulated.

G.5 A Reflection Class Across Five Families

Reflection keywords used in the reinforcement-learning literature (Guo et al., 2025; Zeng et al., 2025) can be tested on five released RL-from-base transitions (Table 32). The class ("wait", "alternatively" and "verify" in three surface variants each) is evaluated on the mathematics prompts at the 384-token cap. On three of the five transitions both the predecessor and the successor write the class at under 0.05 per 1,000 tokens on their own trajectories, a rate no decomposition can distinguish from noise. On Mistral-7B-v0.1 only 2 of the ids survive the round-trip rule against the 6 the class needs, because the successor's tokenizer splits the leading-space spellings the predecessor gives as one token.

On MiMo-7B-RL-Zero the class rises from 0.39 to 4.87 per 1,000 tokens on each model's own trajectories, and it divides both ways: the change in reach is 0.55 [0.53, 0.56] of the counted change and the change in propensity is positive under both references, 2.03 [1.87, 2.18] at the successor's contexts and 0.30 [0.24, 0.36] at the predecessor's. Read as a change in propensity, the counted change of this class on this transition would have the direction right and the size wrong, as the ordinal class does on the same transition (Appendix G.4). What rises here is one word: "wait" carries 0.90 of the class by expected count and 4.10 per 1,000 tokens of the 4.48 [4.17, 4.81] counted change, against 0.39 for "alternatively" and -0.01 for "verify". The control class, mapped by the same rule on the same passes, moves the other way at matched contexts, so the positive change in propensity belongs to the reflection class rather than to the mapping.

Table 30: The ordinal class on a second family, decomposed, with the two control classes beneath it. The class rises on both populations with the change in reach the larger part, and its change in propensity is positive under both references rather than negative. Rates per 1,000 tokens with 95% intervals; the four cells are each model teacher-forced on each arm's trajectories, b the predecessor and t the successor; change in propensity at the successor's contexts and at the predecessor's, and the mean of the two; the last column is the successor-reference change in reach as a share of the row's counted change. The control rows carry the standing control words; their cells and shares are not reported.

Class, prompts	b on b	b on t	t on b	t on t	Counted change
Ordinal class, mathematics	9.25	10.39	9.55	10.67	1.42 [1.14, 1.75]
model-written	6.30	7.19	6.91	7.81	1.51 [1.30, 1.72]
Control class, mathematics	—	—	—	—	2.16 [1.24, 3.09]
model-written	—	—	—	—	0.15 [−1.06, 1.37]
Head-token control, mathematics	—	—	—	—	−0.61 [−0.91, −0.31]
model-written	—	—	—	—	−0.36 [−0.48, −0.23]

Class, prompts	At the successor's contexts	At the predecessor's contexts
Ordinal class, mathematics	0.28 [0.19, 0.36]	0.30 [0.22, 0.37]
model-written	0.62 [0.58, 0.66]	0.61 [0.56, 0.65]
Control class, mathematics	—	—
model-written	—	—
Head-token control, mathematics	—	—
model-written	—	—

Class, prompts	Mean change in propensity	Reach share
Ordinal class, mathematics	0.29 [0.21, 0.36]	0.80 [0.74, 0.86]
model-written	0.61 [0.57, 0.66]	0.59 [0.53, 0.64]
Control class, mathematics	−2.65 [−2.85, −2.45]	—
model-written	−4.15 [−4.31, −3.99]	—
Head-token control, mathematics	−0.09 [−0.12, −0.06]	—
model-written	−0.07 [−0.09, −0.05]	—

G.6 Breadth Summary

Table 28 above summarizes every case-study transition in one row. On the cells whose shares the floor rule of Appendix C.4 does not withhold, the counted change is larger in the change in reach on all but two: Llama-3.1-8B SimpleRL, where it is larger in the change in propensity, and Mistral-7B-v0.1 SimpleRL, where the two components cannot be separated. The one transition whose change in propensity carries the counted change (Llama-3.1-8B SimpleRL) is examined further below.

G.7 The Llama SimpleRL Transition Across Three Populations

Section 4.3 presents the Llama-3.1-8B SimpleRL transition as the case in which the change in propensity carries the counted change. That result was computed on the model-written prompts. Table 33 sets the same connective class, the same 46 ids used on the Olmo-3 RL transition, beside two further prompt populations.

The connective class falls on all three populations and the change in propensity is the majority component on all three. Under the successor reference, $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ on the model-written prompts is 0.66 [0.60, 0.74], on the mathematics prompts 0.70 [0.58, 0.93], and on the human-written prompts 1.27 [0.91, 2.34]. Every signed ratio in the table is taken over a negative counted change: the class falls, not rises, on all three populations. On the human-written prompts the signed ratio exceeds one because the change in reach there is positive at the point and its interval does not exclude zero, so it moves against the fall and the change in propensity accounts for slightly more than the full counted change. The same-model floor on the human-written prompts is −0.04 [−0.42, 0.35] per

Table 31: The paragraph exchange on the second family, predecessor and successor each continuing the other's first paragraphs, and the context intervention evaluated on the window after the supplied paragraph. The opener share follows whose paragraph is continued rather than who continues it, m and $\bar{m}$ both sit at one, and the intervention returns the successor's rate in that window to the predecessor's. The opener share is the share of continuations whose next paragraph opens with one of the class's ids; m (successor continuing) and $\bar{m}$ (predecessor continuing) are the shares of the predecessor-to-successor difference that move with the paragraph; D is that difference; the projection readout is the exchange's own score, fitted on a disjoint half of the prompts. Intervals are 95% from a 10,000-draw prefix-paired bootstrap holding the difference direction fixed; the last five rows are rates per 1,000 tokens of the window after the first paragraph break found in the response's text, with intervals from a 2,000-draw prompt-paired bootstrap.

Quantity	Value
Predecessor continuing its own first paragraphs	0.66
Predecessor continuing the successor's	0.72
Successor continuing the predecessor's	0.66
Successor continuing its own	0.71
$\bar{m}$ (predecessor continuing), projection readout	1.05 [0.95, 1.20]
m (successor continuing), projection readout	1.00 [0.87, 1.15]
the difference D it is taken over	0.008 [0.005, 0.011]
$\bar{m}$ (predecessor continuing), class-opener share	1.23 [1.01, 2.11]
m (successor continuing), class-opener share	1.14 [0.86, 1.86]
the difference D it is taken over	0.04 [0.01, 0.08]
Ordinal openers in the window, predecessor	2.39 [2.31, 2.48]
the successor, unmodified	2.54 [2.44, 2.64]
the successor on the predecessor's first paragraph	2.38 [2.29, 2.47]
Context-intervention correction, C	0.91 [0.38, 1.00]
Signed movement, m	1.09
the gap $y_T - y_B$ it is taken over	0.15 [0.07, 0.23]

Table 32: Three reflection keywords from the literature on five transitions trained from a base, mathematics prompts, both models under the literal SimpleRL prompt. The first three rows carry only each model's rate on its own trajectories, because the class is too rare on them to decompose; the fourth carries none, because the class does not map onto that successor's tokenizer and no pass was run; the fifth carries the whole decomposition, and its change in propensity is positive. Rows are named by the predecessor, the first four paired with that predecessor's SimpleRL-Zoo successor and the last with MiMo-7B-RL-Zero; rates per 1,000 tokens with 95% intervals where the row stores one; *own, predecessor* and *own, successor* are each model teacher-forced on its own trajectories; the change in reach is the counted change less the change in propensity at the successor's contexts.

Predecessor	Own, predecessor	Own, successor	Counted change
Qwen2.5-7B	0.03	0.03	—
Llama-3.1-8B	0.01	0.00	—
DeepSeek-Math-7B	0.01	0.00	—
Mistral-7B-v0.1	—	—	—
MiMo-7B-Base	0.39	4.87	4.48 [4.17, 4.81]

Predecessor	At the successor's contexts	At the predecessor's contexts
Qwen2.5-7B	—	—
Llama-3.1-8B	—	—
DeepSeek-Math-7B	—	—
Mistral-7B-v0.1	—	—
MiMo-7B-Base	2.03 [1.87, 2.18]	0.30 [0.24, 0.36]

1,000 tokens, a ratio of 0.05 of the population's own counted change. The control class's counted change on the same four cells is 0.26 per 1,000 tokens, of the opposite sign to the counted class's fall.

Table 33: The connective decomposition of the Llama-3.1-8B SimpleRL transition on three prompt populations, the class fixed at the same 46 ids used on the Olmo-3 RL transition. Rates are per 1,000 tokens. Each counted change is negative and each signed ratio is taken over that negative counted change under the successor reference; on the human-written prompts (‡) the signed ratio exceeds one because the change in reach is positive at the point and its interval does not exclude zero, leaving the group sign uncertain (Appendix B.4). The two middle columns are the change in propensity under each reference.

Population	Counted change	Successor's contexts	Predecessor's contexts	$\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ (succ. ref.)
model-written	−1.06 [−1.18, −0.92]	−0.69 [−0.74, −0.64]	−0.69 [−0.73, −0.64]	0.66 [0.60, 0.74]
mathematics	−2.19 [−2.68, −1.67]	−1.54 [−1.67, −1.43]	−1.33 [−1.42, −1.23]	0.70 [0.58, 0.93]
human-written ‡	−0.77 [−1.10, −0.40]	−0.97 [−1.05, −0.90]	−0.77 [−0.83, −0.72]	1.27 [0.91, 2.34]

Table 34: Stop token ids for each transition's predecessor and successor, whether the two sides share the same set, and where scored positions follow a template token, the arithmetic worst-case bound on the change in propensity from those positions alone.

Transition	Predecessor stop ids	Successor stop ids	Same?
Qwen2.5-7B → Qwen2.5-7B-Instruct	151643	151643, 151645	no
Llama-3.1-8B → Llama-3.1-8B-Instruct (NousResearch mirrors)	128001	128001, 128008, 128009	no
OLMo-2-1124-7B → OLMo-2-1124-7B-Instruct	100257	100257	yes
Qwen2.5-1.5B → Qwen2.5-1.5B-Instruct	151643	151643, 151645	no
Qwen2.5-14B → Qwen2.5-14B-Instruct	151643	151643, 151645	no
Mistral-7B-v0.3 → Mistral-7B-Instruct-v0.3	2	2	yes
gemma-2-9b → gemma-2-9b-it (unsloth mirrors)	1	1	yes
OLMo-2-1124-13B → OLMo-2-1124-13B-Instruct	100257	100257	yes
Olmo-3-1025-7B → Olmo-3-7B-Instruct	100257	100257, 100265	no
Olmo-3-1025-7B → Olmo-3-7B-RL-Zero-Code	100257	100257	yes
MiMo-7B-Base → MiMo-7B-RL-Zero	151643	151643	yes
Llama-3.1-8B → SimpleRL-Zoo	128001	128001	yes
Olmo-3-1025-7B → RL-Zero-Mix	100257	100257	yes

gemma-2-9b → gemma-2-9b-it (unsloth mirrors): model-written: stored 1.34, worst-case [0.39, 2.30]; human-written: stored 0.93, worst-case [−1.42, 3.27] (sign can change in the worst case).
Mistral-7B-v0.3 → Mistral-7B-Instruct-v0.3: model-written: stored 0.26, worst-case [−1.73, 1.99] (sign can change in the worst case); human-written: stored 0.21, worst-case [−1.36, 1.58] (sign can change in the worst case).

H Response Termination

H.1 Stop Token Identifiers

Table 34 lists the stop token ids for the predecessor and successor of each transition. The successor carries extra stop ids on five transitions: Qwen (all three sizes), Llama-3.1-8B-Instruct and Olmo-3-7B-Instruct; the remaining transitions share the same ids. Separately, scored positions immediately after a template token occur only on Mistral-7B-v0.3 and Gemma-2-9B (both on both populations). The table gives the arithmetic worst-case bound on the change in propensity from those positions alone: only Gemma on model-written prompts keeps its sign under this bound.

H.2 The Stopping Event at the 384-Token Cap

Table 35 gives every transition and prompt population, and Figure 8 draws the same cells. The event is a single token id per transition, so no classifier enters the measurement. Every row sits under a 384-token cap that the Olmo-3 RL successor reaches on 0.97 [0.96, 0.98] of its model-written responses against the predecessor's 0.85 [0.84, 0.87], and on 0.98 [0.97, 0.99] of its mathematics ones; of the other successors trained from a base, Mistral's reaches the cap most often, at 0.95 [0.95, 0.96], so few of its responses end with a stop id at all and its calibration is the widest row. The table therefore divides the hazard, defined at every scored position, and not the mean length. The control class of comparable words moves in the same sense as the event on every row where both changes can be distinguished from zero, and by a larger amount, so the stopping event is specific in how its fall divides and not in the fall itself.

$\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ (the mean of the two references) of the Olmo-3 RL transition's fall is 0.16 [0.14, 0.18] on the model-written prompts, 0.14 [0.12, 0.16] on the screened ones, and 0.22 [0.20, 0.24] on the

Table 35: The stopping event, decomposed, under the 384-token cap. The second column is the counted change and the third the change in reach: the Olmo-3 RL transition's fall is almost all change in reach, Mistral's almost all change in propensity, and Llama's counted change is the one that rises. The lower block adds the change in propensity and the control class. Rates per 1,000 tokens with 95% intervals; the change in reach and change in propensity are the mean of the two references and sum to the counted change.

Transition, prompts	Counted change	Change in reach
Olmo-3 → RL-Zero-Code, model-written	−0.34 [−0.38, −0.29]	−0.28 [−0.32, −0.24]
screened	−0.44 [−0.49, −0.40]	−0.38 [−0.43, −0.34]
mathematics	−0.39 [−0.44, −0.35]	−0.31 [−0.35, −0.27]
Olmo-3 → RL-Zero-Mix, model-written	−0.31 [−0.35, −0.27]	−0.42 [−0.47, −0.38]
screened	−0.28 [−0.34, −0.23]	−0.44 [−0.50, −0.38]
mathematics	−0.42 [−0.46, −0.37]	−0.43 [−0.48, −0.39]
Qwen2.5-7B → SimpleRL-Zoo, model-written	−0.61 [−0.69, −0.53]	−0.29 [−0.37, −0.21]
screened	−0.66 [−0.75, −0.58]	−0.31 [−0.39, −0.23]
mathematics	−0.51 [−0.60, −0.42]	−0.28 [−0.36, −0.20]
Llama-3.1-8B → SimpleRL-Zoo, model-written	0.40 [0.31, 0.50]	0.31 [0.22, 0.40]
Mistral-7B-v0.1 → SimpleRL-Zoo, model-written	−0.65 [−0.71, −0.61]	−0.01 [−0.04, 0.03]

Transition, prompts	Change in propensity	Control class
Olmo-3 → RL-Zero-Code, model-written	−0.05 [−0.06, −0.05]	−2.94 [−3.29, −2.64]
screened	−0.06 [−0.07, −0.05]	−3.15 [−3.51, −2.82]
mathematics	−0.09 [−0.10, −0.08]	−5.26 [−5.93, −4.60]
Olmo-3 → RL-Zero-Mix, model-written	0.12 [0.10, 0.14]	−2.67 [−2.99, −2.35]
screened	0.16 [0.14, 0.18]	−2.81 [−3.17, −2.47]
mathematics	0.016 [0.002, 0.031]	−5.40 [−6.13, −4.70]
Qwen2.5-7B → SimpleRL-Zoo, model-written	−0.32 [−0.35, −0.29]	−5.33 [−6.70, −4.02]
screened	−0.35 [−0.38, −0.32]	0.60 [−1.13, 2.21]
mathematics	−0.23 [−0.26, −0.20]	−7.83 [−9.24, −6.46]
Llama-3.1-8B → SimpleRL-Zoo, model-written	0.09 [0.08, 0.11]	0.06 [−0.32, 0.44]
Mistral-7B-v0.1 → SimpleRL-Zoo, model-written	−0.65 [−0.69, −0.62]	−16.21 [−17.64, −14.66]

mathematics ones. Across the four successors trained from a base the same signed ratio is 0.52 [0.46, 0.60] on Qwen2.5-7B, 0.99 [0.94, 1.04] on Mistral-7B-v0.1, and 0.23 [0.18, 0.30] on Llama-3.1-8B. By the mean of the two references, the change in propensity is the larger component on Qwen2.5-7B's model-written cell and on Mistral-7B-v0.1, and the change in reach on every other cell of Figure 8 except Qwen2.5-7B's screened cell, whose shares the floor rule of Appendix C.4 withholds. Stopping and the connective class divide differently on the same successor: the paired difference in $\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ (the mean of the two references) is 0.21 [0.19, 0.23] on the model-written prompts and 0.20 [0.18, 0.22] on the screened ones, and on mathematics prompts the connective class's own counted change cannot be distinguished from zero, so no difference is defined there. The Olmo-3 RL transition's responses run 22 [19, 25] tokens longer and Llama's −14 [−19, −10]. On Olmo-3-7B-RL-Zero-Mix the change in propensity is positive on all three populations while the hazard falls, so there the change in reach accounts for more than all of the fall.

H.3 A Comparator Set at the 2,048-Token Cap

Every row of Table 35 sits under a cap most of the successor's responses reach, which makes the event partly a censoring artifact. We re-generated the Olmo-3 RL transition at 2,048 tokens, a generation cap most responses stop short of, and scored the hazard over the whole trajectory (Figure 9). The predecessor ends on its own stop token in 0.73 [0.71, 0.75] of its responses and the successor in 0.85 [0.82, 0.87]. The predecessor's median response is 636 [624, 648] tokens; the successor's is 1,239 [1,191, 1,279]. The hazard still falls, by −0.09 [−0.15, −0.03] per 1,000 tokens, out of four cells: 0.76 and 0.79 for the predecessor on its own and on the successor's trajectories, 0.65 and

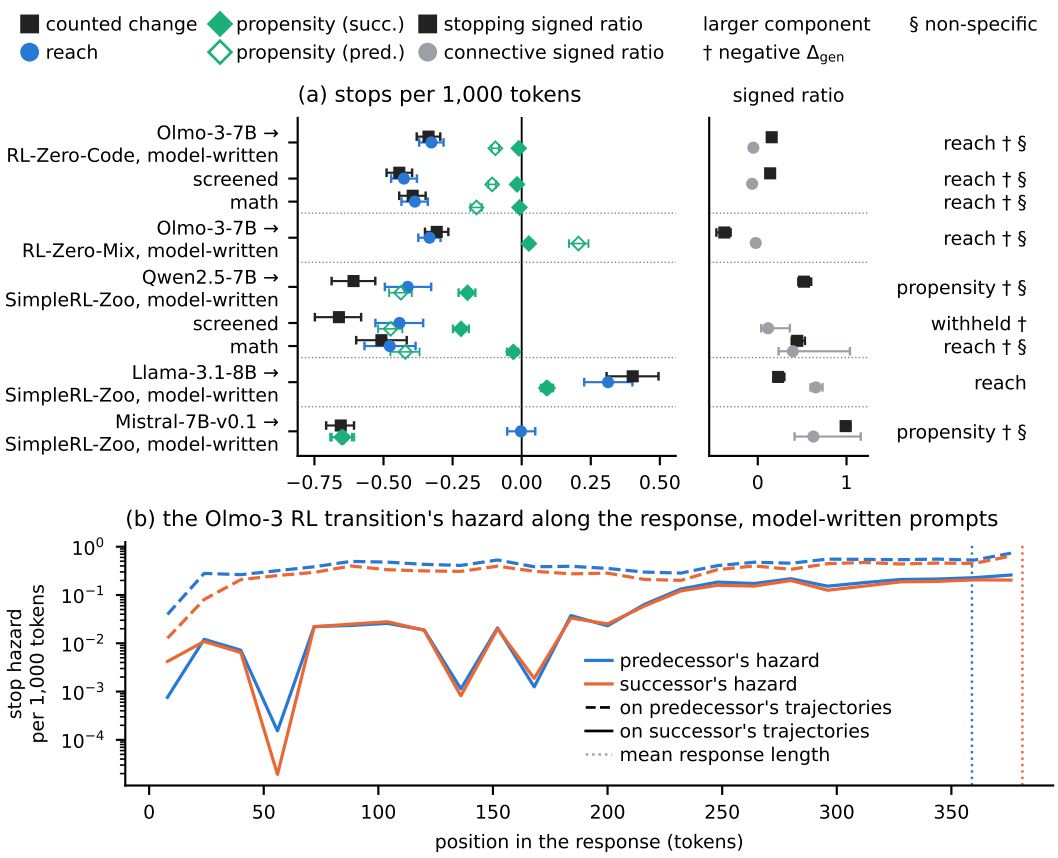


Figure 8: One event, four lineages, every cell at the 384-token cap; the Olmo-3 transition's own row is regenerated and evaluated below at 2,048 tokens. **(a)** Each row is a transition on one prompt population: the counted change in the rate at which a response ends, beside the change in reach and the change in propensity evaluated on the successor's trajectories (filled) and on the predecessor's (hollow). The strip at right is each row's $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ averaged over the two references, and the right-edge label names the larger component by that average; a signed ratio the floor rule of Appendix C.4 withholds is not drawn, and where it is the stopping event's, the label reads "withheld". **(b)** The Olmo-3 transition's four hazard curves by scored position: color is the model whose probability is evaluated (predecessor, successor) and line style the trajectories it is evaluated on (solid: the successor's; dashed: the predecessor's); dotted verticals mark each arm's mean response length; log scale. Rates per 1,000 tokens with 95% intervals.

0.67 for the successor on the predecessor's and on its own. The change in propensity now carries the whole of that fall: −0.12 [−0.13, −0.11] per 1,000 tokens evaluated on the successor's contexts, −0.10 [−0.11, −0.09] evaluated on the predecessor's. Both changes in reach have intervals containing zero, the successor-reference one at 0.04 [−0.03, 0.09] per 1,000 tokens, so the $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ of the fall (the mean of the two references) is 1.30 [0.77, 3.69]; $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ under the successor reference is 1.41 [0.81, 4.11], above one at the point. In Table 35 $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (mean of the two references) at 384 tokens is 0.16 [0.14, 0.18]. The distance between that signed ratio at 2,048 tokens and the one at 384 tokens, 1.15, exceeds the 0.30 at which the cap rather than the model would account for the earlier reading; restricting these same trajectories to their own first 384 tokens reproduces the capped signed ratio, so the two readings differ by the scoring window and not by the draw, the estimator or the prompts.

Table 36 scores six comparators beside the stop event on the same four passes: the connective class, the control class, a bare comma, "of", "the", and the whole 87-id class the other groups are drawn from. The stop event is the only group whose change in propensity carries the whole of its counted change. The comma's change and the change in "of" are carried by the change in reach; the control

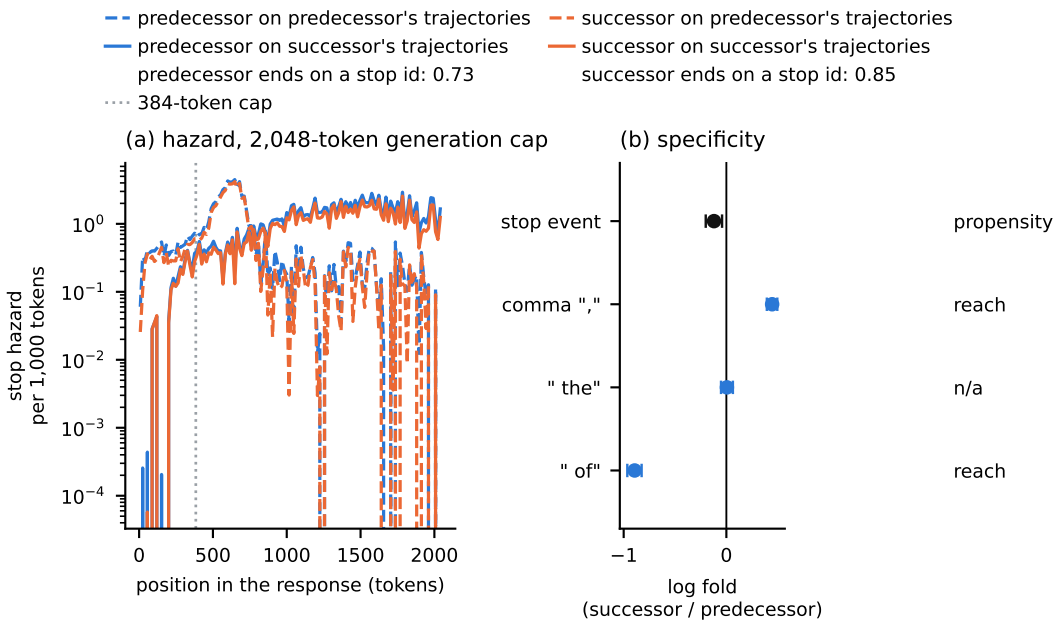


Figure 9: The Olmo-3 RL transition's stopping event at the 2,048-token cap: the curves separate well past the old cap, and the event is not the largest mover. **(a)** The four hazard curves by scored position, each model on its own trajectories and on the other's, with the 384-token cap marked and each arm's share of responses that stop before the new one in the legend. **(b)** The event's log fold beside three function tokens, and at the right how each of them divides. Log scale in (a); 95% intervals in (b).

class divides between the two components; "the" has a counted change that cannot be distinguished from zero and so no signed ratio; and neither can the 87-id class's, its members moving in opposite directions and canceling. Two of the three function tokens also move further in relative terms than the stop event (the comma's log fold is 0.45 [0.39, 0.50] and the fold of "of" is −0.89 [−0.96, −0.83] against the event's own −0.12 [−0.20, −0.04]), so the division of the change separates stopping from its neighbors, and its size does not.

H.4 The Stopping Event at the 8,192-Token Generation Cap

The division changes between the 384-token, 2,048-token and 8,192-token caps. Does the division survive a longer generation cap? We generated the same transition again at 8,192 tokens, teacher-forced both models on both arms' whole trajectories, and evaluated the same trajectories again over their first 2,048 positions; Table 37 sets the three readings beside one another. The transition is still censored at the longer cap: 452 of the predecessor's 1,800 responses and 80 of the successor's reach it. Over the whole trajectories the hazard rises by 0.27 [0.23, 0.31] per 1,000 tokens at $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (mean of the two references) of −0.29 [−0.35, −0.25], with the change in propensity below zero under both references (−0.12 [−0.13, −0.11] and −0.04 [−0.05, −0.04]); over the first 2,048 positions of those same trajectories the counted change is −0.07 [−0.13, −0.02] at $\Delta_{\text{prop}}/\Delta_{\text{gen}}$ (mean of the two references) of 1.51 [0.81, 5.64], which is the reading the 2,048-token draw gives.

What is stable across the two generation caps is the sign of the change in propensity, below zero and distinguishable from zero under both references at both caps. What moves is the change in reach, and with it the counted change: at 8,192 tokens the predecessor stops far more often at the successor's contexts (0.68 per 1,000 tokens) than at its own (0.29), because the responses that never stop dominate its own trajectories. In expected-length units the change is −845 [−1,037, −670] tokens, so the successor's responses are the shorter ones at this cap where at 2,048 they were the longer by 296 [248, 345]. The change in propensity is therefore stable in sign across the two generation caps, while the division of the counted change is not.

Table 36: Seven events on the Olmo-3 RL transition's own 2,048-token trajectories, decomposed the same way. Only the stop event's counted change is carried by the change in propensity; the connective class, the comma and "of" are carried by the change in reach. Rates per 1,000 tokens, which for the stop event is the hazard per scored position, with 95% intervals; the change in reach column is the predecessor's own conditionals at the successor's contexts (successor-reference); the two propensity columns are the same quantity evaluated on each model's contexts; the signed ratio is taken over the row's own counted change, under the mean of the two references, and is withheld where that counted change cannot be distinguished from zero. The 87-id class carries a mean change in propensity of -0.46 per 1,000 tokens in place of a signed ratio.

Event	Counted change	Change in reach
Response ends	-0.09 [-0.15, -0.03]	0.04 [-0.03, 0.09]
Connective class	1.27 [1.17, 1.37]	1.43
Control class	-1.55 [-2.27, -0.87]	-1.42
Comma	19.71 [17.72, 21.59]	20.69
"of"	-12.10 [-13.46, -10.85]	-11.35
"the"	0.22 [-2.08, 2.57]	-1.26
All 87 ids	-0.28 [-1.00, 0.39]	0.00

Event	At the successor's contexts	At the predecessor's contexts
Response ends	-0.12 [-0.13, -0.11]	-0.10 [-0.11, -0.09]
Connective class	-0.15 [-0.17, -0.14]	-0.02 [-0.03, -0.02]
Control class	-0.13 [-0.16, -0.10]	-0.62 [-0.68, -0.56]
Comma	-0.98 [-1.11, -0.86]	-0.17 [-0.20, -0.14]
"of"	-0.75 [-0.79, -0.71]	-0.34 [-0.37, -0.32]
"the"	1.48 [1.37, 1.59]	0.71 [0.66, 0.78]
All 87 ids	-0.28 [-0.31, -0.25]	-0.64 [-0.70, -0.58]

Event	$\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ (mean of the two references)	Larger component
Response ends	1.30 [0.77, 3.69]	propensity
Connective class	-0.07 [-0.08, -0.06]	reach
Control class	0.24 [0.17, 0.41]	reach
Comma	-0.03 [-0.03, -0.03]	reach
"of"	0.05 [0.04, 0.05]	reach
"the"	—	counted change not distinguishable from zero
All 87 ids	—	members cancel

At the 384-token cap 0.85 [0.84, 0.87] and 0.97 [0.96, 0.98] of predecessor and successor responses are truncated, and at 2,048 tokens 0.27 [0.25, 0.29] and 0.15 [0.13, 0.18] are. The division changes with the generation cap: at 2,048 tokens the signed ratio $\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ (succ. ref.) is 1.41 [0.81, 4.11], above one at the point, and at 8,192 tokens the counted change itself turns positive (0.27 [0.23, 0.31]) while the change in propensity stays negative (-0.12 [-0.13, -0.11]). The propensity share (Appendix B.4) is 0.77 [0.57, 0.98] of the gross movement at 2,048 tokens, where the components move in opposite directions and reach moves against the count (sign uncertain: the reach interval includes zero), and 0.23 [0.21, 0.25] of the gross movement at 8,192 tokens, where propensity moves against the count. Four other transitions divide the same event differently at the 384-token cap; the sign of the change in propensity is stable across generation caps and censoring assumptions (Table 37).

For the census, the longer cap is a sensitivity rather than a cleaner measurement: predecessor responses degenerate at its tail (Appendix F.3).

H.5 Degeneration as a Competing Terminal Event

Does the change in propensity remain negative under both references when degeneration is treated as a competing terminal event? A response that has entered repetition has not stopped, but it is not at risk of stopping either, and leaving its later positions in the denominator confounds the two. On the same 8,192-token trajectories we mark an onset at the first 256-token span (stride 64 tokens) whose unique-token ratio falls below 0.20 with no later span above 0.30, and take every position at or after it out of the risk set. The predecessor arm has an onset on 0.13 [0.11, 0.14] of its 1,800 trajectories against the successor's 0.02 [0.01, 0.03], at a median position of 192 tokens against the successor's 1,408; before position 2,048 the shares are 0.12 [0.11, 0.14] and 0.01 [0.01, 0.02], so the shorter

Table 37: The stopping event at three readings of the same transition. The first row is the 8,192-token trajectories over their whole length, the second the same trajectories over their first 2,048 positions, and the third the separate 2,048-token draw of Table 36; the second and third are the same generation cap in tokens but different draws. Hazard per 1,000 tokens with 95% intervals on every counted change and signed ratio; the four cells are the predecessor and the successor each scored on each arm's trajectories; the two propensity columns are the same quantity evaluated on each model's contexts, and *mean* is their average.

Reading	$b\|b$	$b\|t$	$t\|b$	$t\|t$	Counted change
Whole trajectories	0.29	0.68	0.25	0.56	0.27 [0.23, 0.31]
First 2,048 pos.	—	—	—	—	−0.07 [−0.13, −0.02]
The earlier draw	0.76	0.79	0.65	0.67	−0.09 [−0.15, −0.03]

Reading	At the successor's contexts	At the predecessor's contexts
Whole trajectories	−0.12 [−0.13, −0.11]	−0.04 [−0.05, −0.04]
First 2,048 pos.	−0.12	−0.10
The earlier draw	−0.12 [−0.13, −0.11]	−0.10 [−0.11, −0.09]

Reading	Mean change in propensity	$\Delta_{\mathrm{prop}}/\Delta_{\mathrm{gen}}$ (mean of the two references)	Larger component
Whole trajectories	−0.08 [−0.09, −0.07]	−0.29 [−0.35, −0.25]	reach
First 2,048 pos.	−0.11 [−0.12, −0.10]	1.51 [0.81, 5.64]	propensity
The earlier draw	−0.11	1.30 [0.77, 3.69]	propensity

window is not free of the event either. The detector marks only part of the responses that never stop; those responses are near-repetitive (mean unique-token ratio 0.03).

Table 38 sets the four readings side by side. The change in propensity is below zero under both references in all four; over the whole generation cap the counted change stays above zero and is carried by the change in reach after censoring, and over the first 2,048 positions the change in propensity is the larger component either way. A second detector, a repeated-8-gram rate above 0.50, gives the same sign and the same reading (0.16 [0.10, 0.22] per thousand at-risk positions, changes in propensity −0.13 [−0.14, −0.12] and −0.06 [−0.07, −0.06]). It marks an onset on 0.13 [0.12, 0.15] of the predecessor's trajectories, and where both detectors mark one they agree to within a stride on 177 of 198 (89%).

I Upstream Context Analyses

I.1 The Paragraph Exchange on the Olmo-3 Transition and on Qwen2.5-7B SimpleRL

On the Olmo-3 transition, each model writes a first paragraph and each then continues the other's (Figure 10). The four cells are sets of paragraphs rather than prompt-matched pairs, and the two prefix sets differ in size. Four replacements narrow the candidates. Shuffling its tokens collapses the effect. A paragraph written for a different problem still gives 0.93 [0.90, 0.97] of the own-problem rate. Stripping its last sentence leaves it unchanged. A length-matched paragraph of the predecessor's own content recovers 0.18 [0.16, 0.20] of it, and reweighting the exchange for length removes part of the gap and not most of it. The paragraph therefore supplies discourse register, with length part of it, rather than topic or the literal words.

The paragraph elicits a response the predecessor already had. Splitting the successor's change in reach by whether a word fills an existing scored position or a new one places 0.17 of it in creation of new positions and 0.83 within existing classes. Creation is the smaller part. The predecessor opens a further paragraph on 0.36 [0.34, 0.39] of its trajectories and the successor on 0.98 [0.98, 0.99].

The exchange on Qwen2.5-7B SimpleRL is scored by projection onto the predecessor-to-successor opener direction, fitted on a disjoint half of the prompts (Table 39). The primary control restricts the projection to the control class, which moves 0.02 of the shift in the predecessor-continuing direction and 0.02 of it in the successor-continuing direction, each below the quarter of the counted class's movement at or above which the control would count as moving with it. The share-based control fails in both directions (0.97 in the predecessor-continuing direction and 2.89 in the successor-continuing

Table 38: The stopping event at 8,192 tokens with degeneration onset treated as a competing terminal event, beside the same trajectories with every position left at risk. The changes in propensity are below zero in all four readings; the counted change and its larger component move with the generation cap and not with the censoring. The censored rows are stops per thousand at-risk positions and the rows with every position at risk are stops per 1,000 tokens, two different denominators (Appendix C.4 (ii)); 95% intervals where the block behind the row stores one; the two propensity rows are the same quantity evaluated on each model's contexts, and *removed* is the share of that arm's scored positions that leave the risk set. The second block is the onset detector itself.

Quantity, over 8,192 tokens	censored at the onset	every position at risk
Counted change	0.17 [0.11, 0.23]	0.27 [0.23, 0.31]
At the successor's contexts	−0.13 [−0.14, −0.12]	−0.12 [−0.13, −0.11]
At the predecessor's contexts	−0.06 [−0.07, −0.05]	−0.04 [−0.05, −0.04]
Larger component	contexts	reach
Removed, predecessor	0.35	—
Removed, successor	0.08	—

Quantity, first 2,048 positions	censored at the onset	every position at risk
Counted change	−0.24 [−0.30, −0.17]	−0.07 [−0.13, −0.02]
At the successor's contexts	−0.12 [−0.13, −0.11]	−0.12
At the predecessor's contexts	−0.12 [−0.14, −0.11]	−0.10
Larger component	propensity	propensity
Removed, predecessor	0.20	—
Removed, successor	0.01	—

Arm	Responses with an onset	Onset before 2,048	Median onset	Repeated-8-gram rule
Predecessor	0.13 [0.11, 0.14]	0.12 [0.11, 0.14]	192	0.13 [0.12, 0.15]
Successor	0.02 [0.01, 0.03]	0.01 [0.01, 0.02]	1,408	0.03 [0.02, 0.03]

Table 39: The paragraph exchange on Qwen2.5-7B SimpleRL, predecessor and successor each continuing the other's first paragraphs. The two off-diagonal rows carry it: handed the successor's paragraph the predecessor moves most of the way to the successor's opener distribution, while handed the predecessor's the successor keeps most of its own. s is the projection of a cell's opener distribution onto the predecessor-to-successor difference direction, estimated on a disjoint half of the prompts; R-share is the share of continuations opening with one of the 20 most-risen words; n counts the continuations that carried a first word. Both s and R-share carry a per-cell 95% interval from a 10,000-draw prefix-paired bootstrap that holds the difference direction fixed.

Continues	Prefixes	s	R-share	n words / rows
predecessor	predecessor's	0.001 [0.000, 0.002]	0.31 [0.28, 0.34]	962 / 1,022
predecessor	successor's	0.005 [0.004, 0.006]	0.36 [0.33, 0.38]	1,264 / 1,380
successor	predecessor's	0.005 [0.004, 0.006]	0.39 [0.36, 0.42]	975 / 1,022
successor	successor's	0.01 [0.01, 0.01]	0.39 [0.36, 0.42]	1,284 / 1,380

$\bar{m}$ (predecessor continuing) on s: 0.67 [0.45, 0.83].
m (successor continuing) on s: 0.31 [0.04, 0.49].

direction), so the supplied paragraph moves the continuation's general vocabulary, not the counted class alone.

I.2 The Paragraph Exchange on Two Census Transitions

Does the paragraph exchange transfer the behavior on two census transitions whose reach shares differ? Table 40 gives the full result on two transitions at the ends of the bare-rendering range: Qwen2.5-7B, whose change in reach carries most of its counted change under the mean of the two references (above one under the successor reference), and OLMo-2-7B, where it carries about half under the mean of the two references. Both are run on the even-numbered model-written prompts, each model continuing the other's first paragraph and its own, with the predecessor prompted bare and each successor under its own template. The outcome is a sampled count on the continuation alone: nothing here is teacher-forced (Appendix C.4 (ii)). A prefix is eligible when both models' trajectories for that prompt and draw have a text-defined first-paragraph break, covering 394 (15%)

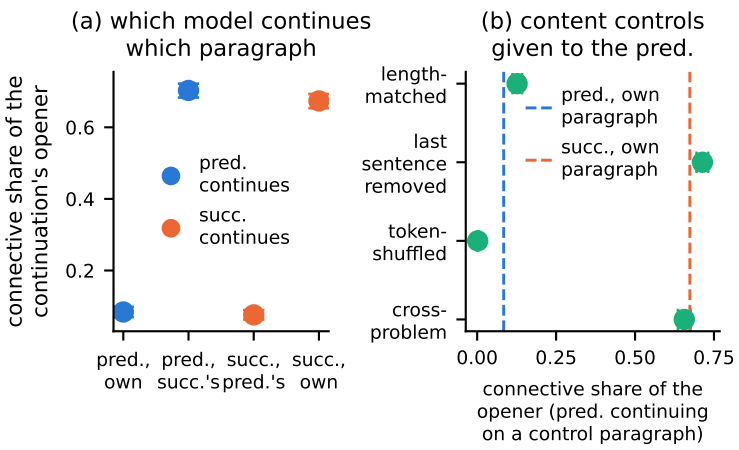


Figure 10: The paragraph exchange and four replacement paragraphs, 95% intervals. In **(a)** the first and third points (the predecessor's paragraph, continued by either model) are alike, and so are the second and fourth (the successor's paragraph): the share of continuations opening with a connective follows whose first paragraph is continued, not who continues it. **(b)** The predecessor continuing the four replacement paragraphs the panel names: shuffling the paragraph's tokens is the one replacement that removes the effect.

Table 40: The paragraph exchange on two census transitions, on the even-numbered model-written prompts. The first four rows are sampled rates of the published list per thousand continuation tokens; D is the diagonal gap; the two m and $\bar{m}$ rows are fractions of D. The control rows are the control class's movement as a fraction of the counted class's. 95% prefix-paired intervals; the contrast between the two transitions below the table is computed on the prompts both transitions share.

Quantity	Qwen2.5-7B → Instruct	OLMo-2-7B → Instruct
Predecessor on the predecessor's paragraph	3.77 [3.20, 4.42]	2.35 [1.75, 3.02]
Predecessor on the successor's	5.09 [4.54, 5.68]	3.37 [2.89, 3.93]
Successor on the predecessor's	4.58 [3.96, 5.21]	6.05 [5.38, 6.75]
Successor on its own	5.28 [4.67, 5.96]	5.62 [5.01, 6.26]
Diagonal gap D	1.51 [0.57, 2.37]	3.28 [2.35, 4.17]
$\bar{m}$ (predecessor continuing)	0.87 [0.62, 1.11]	0.31 [0.06, 0.47]
m (successor continuing)	0.47 [−0.30, 0.75]	−0.13 [−0.56, 0.14]
Control movement, predecessor-continuing	0.61	0.13
Control movement, successor-continuing	0.29	1.34
Eligible prefixes	394	269
Coverage of the predecessor's 2,700 trajectories	394 (14.6%)	269 (10.0%)

of one predecessor's 2,700 trajectories on the even-numbered prompts and 269 (10%) of the other's. The control class moves with the counted class on three of the four transition-by-direction cells, so the supplied paragraph moves the continuation's vocabulary, not the published list in particular.

The difference between the two transitions' $\bar{m}$ is 0.56 [0.27, 0.91] over the 140 prompts of the two transitions, resampled together so that the interval is the contrast's own.

I.3 The Plan-Announcement Stratification

The Olmo-3 RL transition's connective rise is associated with a large shift in first-paragraph structure: the successor writes a plan-announcing paragraph far more often than the predecessor (Table 41).

Each response is assigned to a plan-announcing or non-plan stratum by a fixed surface rule on its first paragraph, with "let me first" removed so that no stratum label depends on a token the connective

Table 41: The connective rise on the model-written prompts, split by whether the first paragraph announces a plan, for Olmo-3-7B-RL-Zero-Code and for Olmo-3-7B-RL-Zero-Mix, trained from the same predecessor. The upper block is each arm's token share in the plan-announcing stratum; the middle block is the counted change within each stratum with the change in propensity under both references; the lower block is the three-component decomposition of the aggregate rise under each reference and as their mean. Rates per 1,000 tokens with 95% intervals; shares are of the aggregate counted change in the same column; the plan stratum's own counted change does not distinguish from zero on either transition.

Quantity	RL-Zero-Code	RL-Zero-Mix
Predecessor, plan-stratum token share	0.04 [0.03, 0.04]	0.04 [0.03, 0.04]
Successor, plan-stratum token share	0.98 [0.97, 0.99]	0.72 [0.70, 0.75]
Plan stratum, counted change	−0.25 [−0.57, 0.05]	−0.22 [−0.54, 0.10]
at the successor's contexts	−0.25 [−0.27, −0.22]	−0.19 [−0.21, −0.16]
at the predecessor's contexts	−0.14 [−0.22, −0.06]	−0.10 [−0.17, −0.03]
Non-plan stratum, counted change	1.42 [0.93, 2.07]	2.72 [2.54, 2.93]
at the successor's contexts	−0.24 [−0.44, −0.10]	−0.12 [−0.16, −0.08]
at the predecessor's contexts	−0.02 [−0.03, −0.02]	0.05 [0.03, 0.06]
Aggregate counted change	2.71 [2.60, 2.82]	2.73 [2.63, 2.83]
Stratum-prevalence share, predecessor's rates	2.94 [2.65, 3.24]	2.14 [1.91, 2.37]
as a share	1.08 [0.97, 1.20]	0.78 [0.70, 0.87]
Stratum-prevalence share, successor's rates	1.35 [0.75, 1.82]	0.11 [−0.03, 0.24]
as a share	0.50 [0.27, 0.67]	0.04 [−0.01, 0.09]
Stratum-prevalence share, mean	2.14 [1.79, 2.42]	1.12 [0.99, 1.26]
as a share	0.79 [0.66, 0.89]	0.41 [0.36, 0.46]
Within-non-plan, successor's rates	1.37 [0.89, 2.00]	2.63 [2.45, 2.82]

class counts. The removal moved no response between strata. The aggregate rise decomposes, weighted by token share, into a stratum-prevalence share (the shift in how often a plan is announced, times the predecessor's rate difference between strata) and within-stratum components, under both references and their mean. The predecessor's plan-stratum token share is 0.04 [0.03, 0.04] and the successor's is 0.98 [0.97, 0.99]. The stratum-prevalence share is 0.79 [0.66, 0.89] on Olmo-3-7B-RL-Zero-Code, while Olmo-3-7B-RL-Zero-Mix, from the same predecessor, carries 0.41 [0.36, 0.46].

The two successors agree on everything except where the rise sits. Both raise the plan-announcing stratum's weight and both price the class below their predecessor inside each stratum. The within-plan change in propensity is −0.25 [−0.27, −0.22] at the successor's contexts and −0.14 [−0.22, −0.06] at the predecessor's. The plan stratum's own counted change does not distinguish from zero, so no share of that stratum's counted change is defined. The split is not specific to the counted class: the predecessor's own rate gap between the strata is 3.73 against 0.63 per 1,000 tokens for the connectives and runs the other way for the control class, whose aggregate counted change is −2.94 [−3.28, −2.63] with a stratum-prevalence share of −2.29 [−3.62, −1.39] in rate units. The rule separates two kinds of text wholesale, not the connective class alone.

Earlier occurrences and paragraph openings. Because each context contains the successor's own earlier tokens, reach includes contexts shaped by the successor's earlier choices, including earlier instances of the behavior: the first instance in a response enters the change in propensity at its position, and later ones are reached through it. The rise is concentrated at paragraph openings, located by each paragraph's first word. The successor produces 18.22 [17.95, 18.48] openings per 1,000 tokens against the predecessor's 6.09 [5.79, 6.40], a ratio of 3.0. At openings the change in propensity is −0.25 [−0.27, −0.22], entirely in the probability per opening. The counted change at openings decomposes into a slot-count part of 1.33 [1.27, 1.39] and a per-slot part of 1.78 [1.70, 1.86] per 1,000 tokens, 3.11 [3.01, 3.21] per 1,000 tokens in all, or 1.15 [1.12, 1.17] of the full counted change; other positions contribute −0.40 [−0.46, −0.33] per 1,000 tokens. The compound span event, a paragraph break followed by a connective, was not decomposed, because the stored per-position passes do not include the paragraph-break probability.

I.4 The Targeted First-Paragraph Structure Edit

The targeted plan-structure edit moves the successor's connective rate but is not specific to the connective class.

The plan-announcement stratification of Appendix I.3 demonstrated an association between first-paragraph structure and downstream connective use. The paragraph exchange of Appendix I.1 moved the connective class *and* the control class. This experiment manipulates *only* the plan structure of the first paragraph, in both directions, and asks whether the effect is specific to the connective class.

Design. Two sets of prefixes are cut from the Olmo-3 RL transition's own stored model-written trajectories, nothing generated to produce them. The insertion set consists of 89 predecessor first paragraphs without the plan feature, out of 117 eligible. The removal set consists of 64 successor first paragraphs with the plan feature. For each prefix a matched version is produced by a deterministic rewriter: the insertion set gains exactly one plan-announcing sentence and the removal set loses its plan-announcing span, with the rest of the paragraph byte-identical. Each prefix and version is continued 4 times per model, at 256 new tokens, temperature 0.7 and top-p 0.95. The connective difference $\Delta_{\text{conn}}(\text{model}) = r(\text{model} \mid \text{plan}) - r(\text{model} \mid \text{no-plan})$ is computed in connective ids per 1,000 tokens of continuation, and Δ_{ctrl} likewise on the control class. The specificity ratio $\rho = |\Delta_{\text{ctrl}}|/|\Delta_{\text{conn}}|$ tests whether the manipulation moves the connective class without moving the control class. Intervals are 95% from a 2,000-draw bootstrap paired by prefix across the two versions.

The pooled connective effect. On the successor, the plan version raises the connective rate by 0.50 [0.20, 0.82] per 1,000 tokens of continuation pooled over both sets (153 prefixes), with the insertion set at 0.61 [0.24, 1.02] and the removal set at 0.38 [−0.14, 0.91]. Almost all of the movement sits at opener positions: $\Delta_{\text{opener}}(\text{successor})$ pooled is 0.53 [0.23, 0.85]. The predecessor shows a larger pooled effect, 1.39 [0.93, 1.88], carried entirely by the removal set (2.24 [1.50, 2.93]) with the insertion set flat (0.09 [−0.32, 0.48]).

The control-class movement and specificity. The control class moves as much as the connective class, downward: $\Delta_{\text{ctrl}}(\text{successor})$ pooled is −0.64 [−1.08, −0.25], opposite in sign to Δ_{conn}. The specificity ratio is $\rho = 1.27$ [0.56, 2.86], well above the 0.25 below which the manipulation would be specific to the connective class. The predecessor's ratio is also above that bar, $\rho = 1.29$ [0.71, 2.15]. Changing the upstream plan structure moves the downstream vocabulary, not the connective class alone.

Survivors. 153 of 317 prefixes survive pooled, with 89 from the insertion set and 64 from the removal set, each above 60, the minimum at which a set is reported on its own.

The blind audit. Two independent readers labeled 40 rows blind to each other. On the insertion set all 20 rows were read as structure-only edits (1.00); on the removal set 13 of 20 were read as removing more than plan structure. Those 13 spans were deliberation openers ("Okay, let's see," "Okay, let's tackle this") that the stratification rule counts as plan announcements but the readers did not. The two readers' inter-reader agreement is $\kappa = 1.00$ on all 40 shared rows. Both readers are models.

Eight cells. Table 42 gives the connective and control rates in every cell.

Upstream first-paragraph structure affects downstream behavior: changing only the plan announcement moves the connective rate and the control rate together, in opposite directions. The manipulation is not specific to the connective class; the evidence is consistent with the paragraph exchange's finding and points to broad context sensitivity. Editing the first-paragraph structure of an upstream context moves the successor's connective rate, but the control class moves at least as far in the opposite direction, measured over the continuation only, so the edit is not specific to this class.

Table 42: The eight cells of the first-paragraph structure edit, each a rate per 1,000 tokens of continuation with a 95% prefix-paired interval. The insertion set inserted a plan sentence into predecessor paragraphs; the removal set removed the plan span from successor paragraphs. The connective class rises under the plan version and the control class falls; neither effect is specific to the connective class.

Cell	Connective class	Control class
Insertion set, plan, successor	3.08 [2.47, 3.74]	2.81 [2.12, 3.54]
Insertion set, no plan, successor	2.47 [1.87, 3.23]	4.08 [3.23, 5.01]
Insertion set, plan, predecessor	1.29 [0.49, 2.44]	3.80 [2.71, 4.95]
Insertion set, no plan, predecessor	1.20 [0.58, 2.04]	5.06 [3.75, 6.43]
Removal set, plan, successor	4.46 [3.91, 5.00]	1.75 [1.36, 2.16]
Removal set, no plan, successor	4.08 [3.61, 4.55]	1.56 [1.10, 2.11]
Removal set, plan, predecessor	5.06 [4.44, 5.64]	2.10 [1.60, 2.62]
Removal set, no plan, predecessor	2.81 [2.20, 3.52]	4.22 [3.09, 5.53]

J Intervention Methods

J.1 The Two Interventions, Defined

Context intervention. The context intervention replaces the upstream context while leaving the continuation model unchanged. The successor continues from a first paragraph the predecessor generated, and the event is counted over the scoring window after the supplied paragraph. For response termination, the predecessor's prefix at a fixed cut point replaces the successor's own prefix, and the successor continues from there.

Local intervention. The local intervention keeps the upstream context and replaces only the target event's total probability. At every generated position with prefix h, the successor's next-token distribution becomes $\tilde{P}(x \mid h)$ of Equation (2), so the successor's relative distribution inside the event and outside it is preserved, and the total event probability is set to the predecessor's. Equivalently, the non-event mass is scaled by $(1 - P_B(E \mid h))/(1 - P_T(E \mid h))$, with a floor of 10^{-8}. For response termination the event is a single token, so the local intervention is its singleton case. Both interventions act on the next-token distribution after the *decoding transform* (temperature 0.7, truncated to the top 0.95 of the mass, renormalized), the layer generation samples from and the layer the decomposition is defined on.

The correction fraction. For an intervened arm with rate y_I, we write

$$C \;=\; 1 \;-\; \frac{|y_I - y_B|}{|y_T - y_B|}, \tag{7}$$

where y_B and y_T are the unmodified predecessor and successor rates on the same rows and the same scoring window. $C = 1$ is the predecessor's own value, $C = 0$ the unmodified successor's, and a negative C an arm further from the predecessor than the successor already was. The folded correction relates to the signed movement $m = (y_T - y_I)/(y_T - y_B)$ by $C = 1 - |1 - m|$: C folds an overshoot ($m > 1$, the intervened arm past the predecessor) into a shortfall, so it counts an overshoot as a shortfall.

Correction denominators. Each m or $\bar{m}$ in Figure 3 divides by the gap Table 47 lists for its arm. For the two Olmo-3 connective arms this is one common-window gap shared by both arms; Table 44 gives both arms against their own rows' rates and on $\mathcal{E}$. Every other arm uses the gap of its own rows and window: the diagonal separation D for the two paragraph exchanges, $y_T - y_B$ for the local arms and the termination transplant. The intervened rates are the same under every gap; only the denominator changes.

Denominators of the rates. Paragraph-exchange cells are per thousand *continuation* tokens (the supplied paragraph is neither scored nor in the denominator). The connective arms on Olmo-3 are per thousand tokens after the first paragraph break, the other local arms per thousand response tokens, and the stopping arms per thousand scored positions. The decomposition rows of Section 4 are per thousand scored tokens of the whole response.

J.2 Eligible Subsets and Scoring Windows

A (prompt, trajectory) unit is eligible for the connective context intervention when the predecessor trajectory stored for it opens a paragraph break whose preceding paragraph is prose of at least 30 words, is not a list item and does not end in a colon, whose following paragraph begins with a word, and which leaves room inside the generation cap to continue from. That rule selects 380 of the 1,800 primary units, covering 272 of the 600 prompts.

The rule is stricter than carrying a paragraph break at all: 598 predecessor trajectories carry a break and 219 of those fall outside $\mathcal{E}$, most of them because the paragraph before the break is a list or a colon-ended lead-in.

The scoring window for connective arms begins after the response's first paragraph break and runs to the end of the response or the generation cap, whichever comes first. Each arm's trajectories therefore contribute to the window only if they contain a paragraph break; the context-intervention arm, whose first paragraph is supplied, contributes fewer trajectories than the arms that write their own.

For the Qwen2.5-7B paragraph exchange, the eligible subset contains 394 prefixes of 2,700 even-index trajectories over 94 prompts, with 4 continuations per prefix and 256 new tokens. For the Llama-3.1-8B transition, 9,364 predecessor prefixes and 10,078 successor prefixes over 4,804 prompts are eligible, again with 4 continuations per prefix and 256 new tokens.

For the stopping prefix transplant, three cut points (256, 512 and 768 tokens into each predecessor trajectory) define three per-cut arms. The scored population is every trajectory whose predecessor prefix reaches the cut point; the stopping hazard is counted over all positions after the cut at which the response could end, and a trajectory reaching the generation cap is censored there.

J.3 Reference-Set Corrections and Undefined Regions

The three Olmo-3 connective corrections. Table 44 gives both corrections under all three reference sets. The context-intervention correction is the same on the eligible subset as it is against its own arm's rates, 0.50 [0.28, 0.70] against 0.50 [0.28, 0.71], and smaller on the common window, 0.42 [0.28, 0.55], where the predecessor's rate is pulled down by trajectories that never open a second paragraph. The local-intervention correction moves the other way: on the eligible subset it widens the predecessor–successor gap by about a quarter of itself, -0.25 [-0.46, -0.09] against -0.13 [-0.19, -0.08] on the common window. The main text uses the common-window value, which Figure 3 plots.

Undefined positions. Where the successor's whole mass after the decoding transform lies inside the event, the local intervention is not defined: the intervened distribution stays on the event and is renormalized, so its event mass is 1 and not $P_B(E \mid h)$, by construction and for any implementation. These positions are counted and their exact contribution to the rate is reported.

On the Qwen2.5-7B transition, 518 positions are undefined, with an exact excess of 0.04 per 1,000 tokens; removing that excess exactly would place the local-intervention correction at -0.05.

On the Llama-3.1-8B transition, 1,165 positions are undefined, carrying 40.91 events = 0.03 per 1,000 tokens; removing them would place the correction at 0.68.

The threshold of 0.10 per 1,000 tokens bounds the region rather than certifying it; above it the undefined positions would materially change the correction.

The induced change in reach on each local-intervention arm. Table 43 gives the identity of Equation (3) on each arm: the fixed-trajectory account ($-\Delta_{\text{prop}}$), the induced change in reach ($r_B(R_I) - r_B(R_T)$), their sum, the observed change $y_I - y_T$, and the residual, in each arm's own rate units and scoring window.

J.4 Structural and Intervention-Validation Controls

Null arm. The null arm re-weights the successor's event mass by a factor of one and must therefore reproduce the unmodified successor. On connectives the null arm's rate deviates 0.03 per 1,000

Table 43: The identity of Equation (3) on each local-intervention arm. $-\Delta_{\text{prop}}$ is the fixed-trajectory account, the induced change in reach is the change in reach that the local intervention itself induces, and the residual is the observed change minus their sum. The Olmo-3 connective row reports connective openers per thousand tokens after the first paragraph break, the other local arms per thousand response tokens, and the stopping arms per thousand scored positions. 95% bootstrap intervals.

Arm	$-\Delta_{\text{prop}}$	Induced	Sum	Observed	Residual
Olmo-3 conn. openers	0.30 [0.27, 0.34]	0.01 [−0.11, 0.12]	0.31 [0.19, 0.43]	0.37 [0.23, 0.51]	0.06 [−0.02, 0.15]
Llama conn.	0.69 [0.64, 0.74]	0.67 [0.56, 0.78]	1.36 [1.23, 1.49]	1.35 [1.19, 1.52]	−0.01 [−0.08, 0.06]
Qwen list	0.20 [0.13, 0.27]	−0.05 [−0.20, 0.09]	0.15 [−0.01, 0.30]	0.11 [−0.09, 0.31]	−0.04 [−0.15, 0.08]
Termination	0.12 [0.11, 0.13]	−0.13 [−0.16, −0.10]	−0.01 [−0.04, 0.02]	−0.01 [−0.03, 0.01]	0.00 [−0.02, 0.01]

tokens from the unmodified successor's, setting the scale of an arm's own sampling noise at 0.29. On stopping the deviation is −0.01 against a noise scale of 0.05.

Placebo arm. Each placebo arm re-weights a class the local intervention should not affect. The control-class placebo on connectives moves the connective rate by 0.16, less than the null arm's own sampling noise, so re-weighting the control class leaves the connective rate within noise. The comma placebo on stopping moves the stop hazard by 0.02, against a noise scale of 0.05. One placebo arm is not clean on response length: the control-class placebo's mean response runs 3 tokens from the unmodified successor's 381, outside the 2-token tolerance within which the placebo's length must match the unmodified arm's, although its repetition rate and unique-token ratio stay at the successor's values.

Intervention exactness. On connectives, the mass the locally intervened sampler used is 0.004 against the predecessor's 0.004 and the successor's 0.003 at the same positions. On stopping, over the full arm's 2,290,643 scored positions, the mass is 0.67 against the predecessor's 0.66 and the successor's 0.57, and the difference from the predecessor's sits entirely at positions where the successor's whole mass is on the stop token (Appendix J.3).

Table 45 follows four structural features of the response across the connective arms. List numerals, complementary discourse markers and plan announcements all move from the successor's value toward the predecessor's under the context intervention and stay at the successor's value under the local intervention. The paragraph-break rate is the exception: after the supplied first paragraph the successor resumes its own break structure, at 15.31 [14.49, 16.17] per 1,000 tokens against the predecessor's 4.12 [3.80, 4.46].

K Additional Intervention Results

K.1 Every Connective Arm on One Scoring Window

Table 48 gives the four intervention-by-behavior cells under each correction's own reference set, and Table 49 puts every connective arm on one scoring window.

K.2 Every Stopping Arm's Realized Hazard

Table 50 gives every stopping arm's realized hazard; the locally intervened arm sits with the null and placebo arms at the unmodified successor's value.

On mean response length, the locally intervened arm differs from the successor's by 5 [−14, 24] tokens, an interval that includes zero, against a successor–predecessor gap of 288 [244, 332] tokens. On the probability of stopping before the cap the difference is −0.01 [−0.03, 0.01], again including zero, against a gap of 0.12 [0.09, 0.15].

K.3 The Stopping Prefix Transplant at Three Cut Points

Table 51 gives each model's hazard when continuing from the other's prefix at three cut points; at the two earlier cuts the predecessor–successor gap runs through or close to zero.

Table 44: The two connective corrections under three reference sets. The context-intervention correction is unchanged on the eligible subset and smaller on one common window, where one pair of comparator rates is shared over every trajectory; the local-intervention correction is negative under all three. Connective class per 1,000 tokens over the scoring window after the first paragraph break; C and the contrast carry 95% bootstrap intervals and the three rates are point estimates. The first column evaluates each intervention against its own arm's rows; the second shares one pair of comparator rates over every trajectory; the third takes them on $\mathcal{E}$, the eligible subset.

Quantity	Each arm's own rates	One common window
Context intervention: the predecessor's first paragraph supplied		
Predecessor, y_B	1.89	1.20
Successor, y_T	3.73	3.96
Intervened, y_I	2.80	2.80
Correction C	0.50 [0.28, 0.71]	0.42 [0.28, 0.55]
Local intervention: the predecessor's connective mass imposed		
Predecessor, y_B	1.20	1.20
Successor, y_T	3.96	3.96
Intervened, y_I	4.33	4.33
Correction C	-0.13 [-0.20, -0.08]	-0.13 [-0.19, -0.08]
Contrast of the two	0.64 [0.41, 0.86]	0.55 [0.41, 0.68]

Quantity	The eligible subset $\mathcal{E}$
Context intervention: the predecessor's first paragraph supplied	
Predecessor, y_B	1.89
Successor, y_T	3.73
Intervened, y_I	2.80
Correction C	0.50 [0.28, 0.70]
Local intervention: the predecessor's connective mass imposed	
Predecessor, y_B	1.89
Successor, y_T	3.73
Intervened, y_I	4.19
Correction C	-0.25 [-0.46, -0.09]
Contrast of the two	0.76 [0.54, 1.02]

Table 45: Four structural features of the response across the connective arms. The first three move toward the predecessor under the context intervention and stay at the successor's value under the local intervention; the paragraph-break rate does neither. Rates per 1,000 tokens with 95% bootstrap intervals, except plan announcements, which are the share of responses whose opening announces a plan.

Feature	Predecessor	Successor
List numerals	13.57 [12.08, 15.26]	4.34 [3.69, 5.02]
Complementary markers	4.83 [4.48, 5.21]	1.87 [1.69, 2.07]
Plan announcements	0.33 [0.31, 0.35]	0.99 [0.98, 0.99]
Paragraph breaks	4.12 [3.80, 4.46]	16.02 [15.76, 16.29]

Feature	Context intervention	Local intervention
List numerals	6.22 [4.63, 8.13]	4.25 [3.66, 4.89]
Complementary markers	3.47 [2.89, 4.04]	1.83 [1.66, 2.02]
Plan announcements	0.62 [0.57, 0.68]	0.99 [0.98, 0.99]
Paragraph breaks	15.31 [14.49, 16.17]	16.09 [15.82, 16.36]

The prefix-transplant arm conditions on the trajectories alive at each cut: the successor-minus-predecessor length gap is 288 [244, 332] tokens over every trajectory and 212 [164, 258], 127 [80, 179] and -440 [-479, -399] restricted to the trajectories alive at each successive cut, so the latest cut conditions on a population where the successor is the shorter continuer.

K.4 Two Interventions on the Llama Transition

The local intervention moves the successor back toward the predecessor. Over 5,400 trajectories the locally intervened arm writes the connective class at 3.38 [3.20, 3.54] per 1,000 tokens, against the successor's 2.03 [1.88, 2.18] and the predecessor's 3.03 [2.88, 3.18] across a gap of -1.00 [-1.16, -0.85], for a local-intervention correction of 0.65 [0.43, 0.84] on the folded scale. The intervened arm's rate passes the predecessor's; the signed movement is 1.35 of the gap, and the folded correction

Table 46: Intervention validation. Every intervened-arm row is computed on the full arm (2,290,643 scored positions on stopping); the stop mass used exceeds the predecessor's only at positions where the successor's whole mass is on the stop token (Appendix J.3). The first two blocks give the mass each locally intervened arm used at the positions it visited, beside the predecessor's and the successor's at those same positions; the last row gives the predecessor's own stopping rate over the successor's trajectories, the quantity the second row gives over the intervened arm's trajectories. Stopping rows are masses per thousand scored positions and connective rows are mean probabilities per position, so the two blocks are not comparable without that factor.

Quantity	Value
Stop mass the locally intervened arm used	0.67
The predecessor's stop mass at those positions	0.66
The successor's stop mass at those positions	0.57
Connective mass the locally intervened sampler used	0.003689
The predecessor's connective mass at those positions	0.003680
The successor's connective mass at those positions	0.003431
The predecessor's stopping rate over the successor's trajectories	0.79

Table 47: Each intervention arm in Figure 3 with the gap it is normalized by (the two Olmo-3 connective arms share one common-window gap), the signed movement m, and the resampling unit. All rows use model-written prompts; the Llama exchange is a fresh draw (Appendix K.4). Both exchange gaps are the diagonal separation D, not $y_T - y_B$; the Llama mirror row gives $\bar{m}$. Rendering: Olmo-3 and termination use the Instruct template grafted onto both models (Appendix D.3); the Qwen exchange uses bare-rendered predecessor paragraphs; the Qwen locally intervened arm's gap mixes renderings (y_B from the bare census arm, Appendix K.5); Llama uses the literal SimpleRL prompt. 95% bootstrap intervals on both the gap and m except the Qwen exchange gap, which is prefix-paired.

Arm	Transition / event	Prompts (trajectories)	Window / units	Gap $y_T - y_B$	m [95%]	Resample
Olmo-3 context	Olmo-3 → RL-Zero-Code / conn. class after break	272 of 600 (380)	after 1st ¶ break; per 1k window tokens	2.76 [2.49, 3.01]	0.42 [0.28, 0.55]	prompts
Olmo-3 local	same	600 (1,800)	same	same	−0.13 [−0.19, −0.08]	same
Qwen context	Qwen2.5-7B → Instruct / published list	94 (394 pfx.)	continuation; per 1k cont. tokens	1.51 [0.57, 2.37]	0.47 [0.25, 0.73]	prompts
Qwen local	same	600 (1,800)	whole resp., 384 cap; per 1k resp. tokens	1.27 [1.06, 1.51]	−0.09 [−0.25, 0.07]	prompts
Llama ctx. (filled)	Llama-3.1-8B → SimpleRL-Zoo / conn. class	4,804	continuation; per 1k cont. tokens	−0.68 [−0.80, −0.56]	0.67 [0.58, 0.74]	joint prompt
Llama mirror (hollow)	same	same	same	same	$\bar{m}$ = 0.46 [0.34, 0.55]	same
Llama local	same	600 (5,400)	whole resp. ≤384; per 1k resp. tokens	−1.00 [−1.16, −0.85]	1.35 [1.16, 1.57]	prompts
Term. context (pooled)	Olmo-3 → RL-Zero-Code / response term.	600 (17,036 rows)	after each of three cuts (Appendix J.2), to 2,048 cap; per 1k scored pos.	0.26 [0.19, 0.33]	1.80 [1.51, 2.30]	prompts
Term. local	same	600 (1,800)	whole resp. to 2,048 cap; per 1k scored pos.	−0.08 [−0.13, −0.02]	−0.14 [−0.80, 0.14]	same

counts this overshoot as a shortfall. The local intervention is exact wherever it is defined: at every one of the 1,605,078 defined positions the predecessor and intervened masses agree, with 0 positions failing and the largest deviation being floating-point noise. The 1,165 undefined positions carry 40.91 events = 0.03 per 1,000 tokens; removing them exactly would place the correction at 0.68.

Table 48: The four intervention-by-behavior cells on the Olmo-3 RL transition. y_B and y_T are the unmodified predecessor and successor on that row's own comparison and y_I the intervened arm. Both connective rows count the same event, connective class per 1,000 tokens over the scoring window after the response's first paragraph break; what differs between them is the set of trajectories each row's reference rates are computed on, which is why y_B and y_T are not shared between them. Table 49 lists every connective arm's pooled rate on the same scoring window; the shared-reference comparator is the common-window column of Table 44. Connective rows at the 384-token cap; stopping rows at the 2,048-token cap, in stop events per thousand scored positions; 95% bootstrap intervals on C and on the predecessor–successor gap.

Behavior, intervention	y_B	y_T	y_I
Connectives, context	1.89	3.73	2.80
Connectives, local	1.20	3.96	4.33
Stopping, local	0.74	0.67	0.66
Stopping, context (pooled)	0.71	0.97	0.50

Behavior, intervention	$y_T - y_B$	C (folded)	m (signed)
Connectives, context	1.84 [1.40, 2.26]	0.50 [0.28, 0.71]	0.50
Connectives, local	2.76 [2.49, 3.01]	−0.13 [−0.20, −0.08]	−0.13
Stopping, local	−0.08 [−0.13, −0.02]	−0.14 [−0.80, 0.14]	−0.14
Stopping, context (pooled)	0.26 [0.19, 0.33]	0.20 [−0.30, 0.49]	1.80

Table 49: Every connective arm on one scoring window. The null and placebo arms sit within their own sampling noise of the unmodified successor's rate; the local intervention raises it and the context intervention lowers it. Connective class per 1,000 tokens over the scoring window after the response's first paragraph break, with 95% intervals; *Trajectories* counts the trajectories each arm contributes (all 9 trajectories per prompt for the unmodified and locally intervened arms; trajectories without a window add nothing to either count). These pooled rates differ from the correction comparator of Table 44, which uses the primary trajectories.

Arm	Rate per 1,000 tokens	Trajectories
Predecessor, unmodified	1.10 [0.96, 1.24]	5,400
Successor, unmodified	3.97 [3.86, 4.09]	5,400
Context intervention (predecessor's first paragraph)	2.80 [2.43, 3.17]	380
Local intervention (predecessor's connective mass)	4.29 [4.19, 4.40]	5,400
Null arm (re-weight by one)	3.99 [3.86, 4.12]	1,800
Placebo arm (control class re-weighted)	3.80 [3.66, 3.94]	1,800

The paragraph exchange on this transition is evaluated on 4,804 prompts, from a fresh draw of model-written prompts, on which some trajectory opens with a prose first paragraph: 9,364 predecessor and 10,078 successor first paragraphs, each continued four times by each model. The four cells are 2.60 (the predecessor on its own paragraph), 2.29 (the predecessor on the successor's), 2.38 (the successor on the predecessor's) and 1.92 (the successor on its own) per 1,000 tokens of continuation, for an exchange gap of -0.68 [-0.80, -0.56]. The successor-continuing movement is $m = 0.67$ [0.58, 0.74] and the predecessor-continuing movement $\bar{m} = 0.46$ [0.34, 0.55]; both move the continuation toward the paragraph's source. Their difference, 0.21 [0.10, 0.33], excludes zero: the paragraph's effect depends on which model continues it, so the two directions are reported separately and no single movement is given for the transition. Intervals are joint prompt-paired bootstrap intervals over the prompts. Continuing its own paragraph, the successor writes slightly fewer connectives than on its own trajectories after the same paragraph (-0.07 [-0.13, -0.01] per 1,000 tokens on continuations that open with a letter-headed paragraph; the predecessor's gap is -0.02 [-0.11, 0.06]). Attributing that deficit to the continuing model places m in 0.59–0.98 and $\bar{m}$ in 0.41–0.65; attributing it to the paragraph's source places them in 0.53–0.71 and 0.21–0.52. On the window the exchange acts on, the continuation after the first paragraph, the propensity share (Appendix B.4) of this transition's fall is 0.13 [0.07, 0.28], against 0.66 [0.60, 0.74] over whole responses on all trajectories and 0.46 [0.35, 0.65] over whole responses on the trajectories the window uses (Appendix K.6). m and $\bar{m}$ above are computed on the full prompt set; on the model-written prompts on which these shares are computed they are $m = 0.59$ [-0.16, 0.93] and $\bar{m} = 0.48$ [-0.47, 0.77], and both intervals include zero.

Table 50: Every stopping arm's realized hazard on the Olmo-3 RL transition at the 2,048-token cap. The locally intervened arm sits with the null and placebo arms at the unmodified successor's value, not at the predecessor's. Stop events per thousand scored positions with 95% intervals; stops is the number of response-ending events counted.

Arm	Hazard	Trajectories	Stops
Predecessor, unmodified	0.74 [0.69, 0.79]	1,800	1,310
Successor, unmodified	0.67 [0.64, 0.70]	1,800	1,523
Local intervention (predecessor's stop mass)	0.66 [0.62, 0.69]	1,800	1,504
Null arm (re-weight by one)	0.66 [0.63, 0.71]	600	504
Placebo arm (comma re-weighted)	0.66 [0.62, 0.70]	600	502

Table 51: The stopping prefix transplant at three cut points and pooled, at the 2,048-token cap: each model continues from its own prefix and from the other's. At the two earlier cuts the predecessor–successor gap the correction divides by runs through or close to zero, which is why their corrections read as ratios rather than magnitudes. Stop events per thousand scored positions with 95% intervals; the cuts are ordered by prefix length.

Prefix cut	Predecessor on its own (y_B)	Successor on the predecessor's (y_I)
Earliest	0.85 [0.78, 0.92]	0.51 [0.47, 0.54]
Middle	0.87 [0.78, 0.97]	0.62 [0.56, 0.67]
Latest	0.21 [0.17, 0.24]	0.25 [0.21, 0.28]
Pooled	0.71 [0.65, 0.77]	0.50 [0.47, 0.53]

Prefix cut	Predecessor on the successor's	Successor on its own (y_T)
Earliest	0.77 [0.73, 0.82]	0.81 [0.77, 0.86]
Middle	0.96 [0.91, 1.02]	0.98 [0.92, 1.05]
Latest	1.18 [1.10, 1.28]	1.24 [1.15, 1.34]
Pooled	0.93 [0.88, 0.99]	0.97 [0.91, 1.03]

Prefix cut	C (folded)	m (signed)
Earliest	−8.30 [−168.13, −2.37]	−8.30
Middle	−1.23 [−20.84, 0.08]	3.23
Latest	0.96 [0.93, 0.99]	0.96
Pooled	0.20 [−0.30, 0.49]	1.80

The control class moves under the local intervention: −0.47 per 1,000 tokens, a movement ratio of 0.35 relative to the counted class's 1.35. Under the exchange, the control class's change is 0.16 [−0.07, 0.40] per 1,000 tokens in the predecessor-continuing direction, an interval that includes zero, and −0.62 [−0.88, −0.35] in the successor-continuing direction, which excludes it; the movement ratios are 0.51 [0.03, 1.42] and 1.36 [0.79, 2.09] (Table 54). The two models do not differ on the control class when each continues its own paragraph: the control class's diagonal gap is 0.09 [−0.19, 0.37] per 1,000 tokens, against the connective class's −0.68 [−0.80, −0.56]. The control class's movement therefore has no direction toward either model, and none is given.

The null arm's rate sits 0.11 [−0.16, 0.39] from the unmodified successor's, an interval containing zero.

K.5 Two Interventions on a Census Transition (Qwen)

The locally intervened arm writes the published excess-vocabulary list at 4.84 [4.46, 5.25] per 1,000 tokens against the successor's 4.73 [4.36, 5.12] and the predecessor's 3.46 [3.17, 3.77], across a gap of 1.27 [1.06, 1.51], for a local-intervention correction of −0.09 [−0.25, 0.07], negative at the point with an interval covering zero. The predecessor's comparator rate is its bare-rendering census arm, while the local intervention renders both models under the successor's chat template, so the correction's gap mixes renderings. On the prompts shared with the exchange the same correction is −0.17 [−0.52, 0.13].

The successor continuing the predecessor's first paragraph returns a successor-continuing movement $m = 0.47$ [0.25, 0.73] over the 94 shared prompts, on the four exchange cells of 3.77, 5.09, 4.58 and 5.28 per 1,000 tokens of continuation (a different denominator from the locally intervened arm,

Appendix C.4 (ii)). The contrast, on 2,000 prompt-paired draws over the shared prompts, is 0.64 [0.23, 1.12]: its interval sits above zero.

The control class moves under the exchange: 0.61 in the predecessor-continuing direction, where the control class's change excludes zero, and 0.29 in the successor-continuing direction, where it does not; on the second census transition (OLMo-2-7B) the ratios are 0.13 and 1.34, with the control class's change unresolved in both. On the Olmo-3 RL transition list numerals and complementary markers move toward the predecessor alongside the counted class (Appendix J.4).

The local intervention is exact at every one of the 637,982 defined positions, with 0 failing and the largest deviation at 0.000. The null arm's rate sits -0.15 [-0.44, 0.15] from the unmodified successor's, an interval containing zero.

Under the bare rendering the change in propensity is -0.39 [-0.46, -0.32] per 1,000 tokens, the predecessor-reference change in propensity is positive (0.43 [0.35, 0.53] per 1,000 tokens), and the predecessor share at matched contexts is 1.08 [1.07, 1.10]. Over prefixes, m is 0.47 [-0.30, 0.75], whose interval includes zero; over the 94 prompts the contrast shares it is 0.47 [0.25, 0.73], which sits above zero.

K.6 The Split on the Window Each Exchange Acts On

Table 52 gives the decomposition of Section 3 on the window the paragraph exchange acts on, the continuation after a response's first paragraph, beside the whole-response decomposition, for every exchange row. The window is cut on the stored trajectories by each exchange's own first-paragraph rule and scored over the positions that follow, on the trajectories that have a first paragraph; the whole-response columns are computed over all trajectories. Intervals are each row's prompt-paired bootstrap on the same draws as its whole-response interval. The exchange counts sampled tokens in fresh continuations and hands over only the first paragraph, while the window decomposition sums teacher-forced class mass on stored trajectories and scores the successor on the predecessor's whole stored context (Appendix C.4 (ii)).

On Olmo-3, Qwen2.5-7B and Llama, reach carries the change on the window: the propensity share there is 0.13, 0.04 and 0.13, each interval below one half. On Llama this reverses the whole-response split, in which propensity carries 0.66 of the fall; the window covers the 300 predecessor and 332 successor trajectories on the model-written population that open with a prose paragraph. Computed over whole responses on the same eligible trajectories, the propensity share is 0.46 [0.35, 0.65], so part of the difference from 0.66 is the restriction to trajectories that open with a prose paragraph. Olmo-3's window is computed on the pairs eligible for its exchange, where supplying the predecessor's first paragraph returns $m = 0.50$ [0.28, 0.70]; over every primary pair with a paragraph break the window share is 0.09 [0.08, 0.10]. Qwen2.5-7B's windowed share rests on the predecessor evaluated under the chat-template rendering, not on the bare rendering the exchange used: the exchange's predecessor paragraphs come from bare-rendered trajectories, which carry no per-position class mass, so the predecessor's side of the window is cut on its chat-template trajectories at the exchange's prompt keys, while the successor's side is the exchange's own. Re-applying the exchange's rule to the chat-template trajectories gives a window share of 0.036 [0.003, 0.227]. OLMo-2-7B is not usable on this window: at the exchange's keys only 46 of the predecessor's chat-template trajectories have a first paragraph, and the window change, -0.79 [-5.03, 2.60] per 1,000 tokens, has an interval that includes zero, so no share is reported from it. On its secondary population, the exchange's rule re-applied to the chat-template trajectories, the window change is 2.30 [0.95, 3.69] per 1,000 tokens with a propensity share of 0.21 [0.01, 0.85], an interval too wide to place it on either side of one half.

K.7 The Boundary Fallback

Where the successor's nucleus excludes the event but the predecessor's does not, the local intervention fills the event mass by a fallback, which spreads that mass by the predecessor's relative distribution inside the event, so the successor's is not preserved there. Table 53 gives the share of positions and the share of restored mass handled by the fallback on each of the four rows that carry a local intervention, and the first-order estimate of m with and without it.

Table 52: The split over whole responses and on the window each paragraph exchange acts on (the continuation after the first paragraph). Δ_{gen} per 1,000 tokens and the propensity share s_{prop} (Appendix B.4), successor reference, with 95% bootstrap intervals. Groups: opposite-direction with propensity against the count on the Olmo-3 and Qwen2.5-7B rows, same-direction on the Llama and OLMo-2-7B rows; ‡: a component's interval includes zero. Olmo-3: the pairs eligible for its exchange; Qwen2.5-7B and OLMo-2-7B: the exchange's prompt keys, the predecessor's side under the chat template; Llama: the model-written population's trajectories with a prose first paragraph. OLMo-2-7B's window change includes zero, so no share is reported; its secondary population is in the text.

	Whole response		Exchange window	
Transition	Δ_{gen}	Propensity share	Δ_{gen}	Propensity share
Olmo-3 RL, connective class	2.71 [2.60, 2.82]	0.08 [0.07, 0.08]	1.91 [1.45, 2.38]	0.13 [0.10, 0.17]
Qwen2.5-7B → Instruct, published list	1.02 [0.79, 1.25]	0.14 [0.10, 0.18]	1.47 [0.64, 2.33]	0.038 [0.004, 0.221]‡
Llama-3.1-8B → SimpleRL-Zoo, connective class	−1.06 [−1.18, −0.92]	0.66 [0.60, 0.74]	−1.24 [−1.87, −0.65]	0.13 [0.07, 0.28]
OLMo-2-7B → Instruct, published list	3.84 [3.47, 4.19]	0.39 [0.35, 0.42]	−0.79 [−5.03, 2.60]	not usable

Table 53: The boundary fallback on each local-intervention row. At positions where the successor's nucleus excludes the event, the local intervention imposes the predecessor's mass by a fallback, which is exact. The local intervention departs from the predecessor's mass only where the successor's whole mass lies on the event (Appendix J.3). The share of positions and the share of restored mass handled by the fallback are small on the Olmo-3 rows and larger on Llama. m is the signed movement of the gap; the last column is a first-order estimate excluding the fallback positions. Shares with 95% bootstrap intervals.

Row	Positions	Share of positions	Share of restored mass	m	m (no fallback, 1st order)
Olmo-3 RL: connectives	576,023	0.1% [0.1, 0.1]%	2.0% [1.7, 2.3]%	−0.13	−0.10
Olmo-3 RL: response termination	2,290,643	0.017% [0.015, 0.019]%	2.5% [1.9, 3.0]%	−0.14	−0.36
Qwen2.5-7B: published list	638,500	0.5% [0.5, 0.6]%	4.3% [3.9, 4.7]%	−0.09	0.08
Llama-3.1-8B SimpleRL: connectives	1,606,243	0.4% [0.4, 0.5]%	13.1% [11.8, 14.4]%	1.35	0.90

K.8 Control-Class Movement Under Each Intervention

Table 54 gives the control class's movement under each intervention arm and direction, beside the counted class's movement, on every arm-by-direction cell.

L Additional Robustness and Cross-Family Analyses

L.1 The Reversal Scan on Each Transition's Own Words

The Olmo-3 RL transition's pattern (a word whose counted change is positive while its change in propensity is negative under both references) appears when each transition's vocabulary is evaluated on its own words rather than on the connective class. Selecting such words on one half of the prompts and decomposing the resulting class as one event on the other half gives a sign discrepancy on each of the three SimpleRL-Zoo transitions in the scan (Qwen2.5-7B, Llama-3.1-8B and Mistral-7B-v0.1) in both folds, six of six (Table 55). The classes are small, between 16 and 28 ids, and the classes rise, with counted changes between 1.27 and 6.70 per 1,000 tokens, while both changes in propensity sit below zero on every fold.

That sign structure transfers; the specific vocabulary does not. On Llama and Mistral a single function word dominates the selected class ("and" takes 0.79 of the Llama class on one fold and 0.54 on the other), and removing it leaves all four of those folds with the same sign structure, at counted changes between 1.12 [0.62, 1.62] and 2.60 [1.86, 3.36] per 1,000 tokens with both changes in propensity still below zero.

Table 54: Control-class movement under each intervention arm and direction, beside the counted class's movement. m is the signed fraction of the way to the predecessor's own rate on successor-continuing and local rows; on predecessor-continuing rows the column gives $\bar{m}$, the signed fraction of the way to the successor's own rate. The ratio is |control movement|/|counted movement|; the flag says whether the control class's change excludes zero; the last two columns give the direction of each class's movement. The ratio compares the two classes' movements in the arm's own rate units; 95% bootstrap intervals on the ratio. On the Llama exchange rows no control direction is given (—), because the two models' own control-class rates do not differ there (Appendix K.4).

	Continuing	Counted	Control/counted	Control change	Toward the other model	
Intervention	model	m or $\bar{m}$	movement	excludes zero	Counted	Control
Olmo-3 RL, connective class; control: the control class at paragraph openings						
context	successor	0.42	0.91 [0.56, 1.43]	yes	toward	toward
local	—	−0.13	0.10 [0.01, 0.38]	no	away	away
Qwen2.5-7B → Instruct, published excess-vocabulary list; control: the control class						
context	predecessor	0.87	0.61 [0.17, 1.61]	yes	toward	toward
context	successor	0.47	0.29 [0.02, 1.27]	no	toward	toward
local	—	−0.09	0.13 [0.02, 9.79]	no	away	toward
Llama-3.1-8B → SimpleRL-Zoo, connective class; control: the control class						
context	predecessor	0.46	0.51 [0.03, 1.42]	no	toward	—
context	successor	0.67	1.36 [0.79, 2.09]	yes	toward	—
local	—	1.35	0.35 [0.15, 0.54]	yes	toward	toward
Olmo-3 RL, response termination; control: placebo words						
context	successor	1.80	0.48 [0.10, 0.94]	yes	toward	toward
local	—	−0.14	3.22 [0.21, 65.25]	no	away	away
OLMo-2-7B → Instruct, published excess-vocabulary list; control: the control class						
context	predecessor	0.31	0.13 [0.02, 2.48]	no	toward	toward
context	successor	−0.13	1.34 [0.09, 18.72]	no	away	toward

Counting words rather than classes is the stricter comparison, and it is where the three transitions separate (Table 56). Against a null that resamples whole prompts and so keeps the words' dependence on one another, Llama-3.1-8B SimpleRL and Mistral-7B-v0.1 SimpleRL exceed their own 99th percentile (33 against 30; 37 against 29) and Qwen2.5-7B SimpleRL sits exactly on it (27 against 27).

Computed across 13 released transitions, the same per-word statistic is a minority of each transition's risen words and a smaller minority of their mass (Table 57). The fraction runs from 0.02 to 0.35 with a median of 0.13, and 4 of the 13 rows exceed their own null's 99th percentile.

L.2 A Held-Out Class on a Fourth Family

Does a class chosen by its counted rise alone on one half of the prompts show a rise with a negative change in propensity on the other half? On DeepSeek-Math-7B it does (change in propensity −0.22 [−0.29, −0.17] at the successor's contexts, −0.13 [−0.18, −0.08] at the predecessor's), but the discrepancy is carried by " and" (0.67 of the class) and reverses sign without it, and the per-word count of reversals (18) sits at its null's mean (17.77).

DeepSeek-Math-7B and the successor SimpleRL-Zoo trained from it were generated once on model-written prompts and split by prompt index into a discovery half of 300 prompts and a held-out half of 300. On the discovery half 47 of 735 word forms rise with an interval excluding zero, and the class is the 20 largest of them by counted change, with no change in propensity computed there (Table 58). The class is chosen on the discovery half only.

On the held-out prompts the held-out class rises and both of its changes in propensity sit below zero (Table 59), so the class-level discrepancy appears on a fourth family, on prompts that had no part in choosing the words and under both references. It does not survive the removal of one word. " and" is 0.67 of the class by expected count, and the remaining 19 forms still rise, by 0.83 [0.21, 1.43] per 1,000 tokens, with both changes in propensity above zero instead of below: 0.21 [0.18, 0.24] at the successor's contexts and 0.23 [0.20, 0.26] at the predecessor's.

Word by word the same prompts show no excess of reversals. 18 forms of the universe reverse, against a reference whose mean is 17.77 and whose ninety-ninth percentile is 33: the count sits at

Table 55: The cross-fitted class on each transition's own words, both folds. The class is selected on one half of the prompts and decomposed on the other; both changes in propensity are below zero on every fold, and removing the class's largest member leaves each one rising. Rates per 1,000 tokens with 95% bootstrap intervals; ids is the size of the selected class; the top word's share is of the class's expected count.

Transition, fold	ids	Counted change	At the successor's contexts
Qwen2.5-7B SimpleRL, A	17	1.27 [0.48, 2.00]	−0.41 [−0.49, −0.34]
fold B	16	1.45 [1.11, 1.79]	−0.24 [−0.30, −0.19]
Llama-3.1-8B SimpleRL, A	22	5.83 [4.73, 7.04]	−0.48 [−0.57, −0.38]
fold B	20	5.89 [4.48, 7.37]	−0.76 [−0.85, −0.67]
Mistral-7B-v0.1 SimpleRL, A	24	6.25 [5.08, 7.46]	−0.67 [−0.76, −0.56]
fold B	28	6.70 [5.48, 7.94]	−0.56 [−0.66, −0.47]
Olmo-3 RL, fold A	76	53.17 [50.79, 55.52]	−16.60 [−17.31, −15.94]
fold B	82	52.62 [50.23, 55.17]	−16.44 [−17.15, −15.77]

Transition, fold	ids	At the predecessor's contexts	Top word
Qwen2.5-7B SimpleRL, A	17	−0.28 [−0.36, −0.21]	" of"
fold B	16	−0.11 [−0.16, −0.06]	" into"
Llama-3.1-8B SimpleRL, A	22	−0.32 [−0.40, −0.23]	" and"
fold B	20	−0.72 [−0.80, −0.63]	" and"
Mistral-7B-v0.1 SimpleRL, A	24	−0.57 [−0.66, −0.49]	"and"
fold B	28	−0.50 [−0.58, −0.42]	"and"
Olmo-3 RL, fold A	76	−2.37 [−2.60, −2.16]	" is"
fold B	82	−2.35 [−2.55, −2.14]	" is"

Transition, fold	ids	Its share	Counted change without it
Qwen2.5-7B SimpleRL, A	17	0.70	1.11 [0.77, 1.42]
fold B	16	0.18	1.14 [0.81, 1.46]
Llama-3.1-8B SimpleRL, A	22	0.79	1.12 [0.62, 1.62]
fold B	20	0.54	1.62 [0.64, 2.59]
Mistral-7B-v0.1 SimpleRL, A	24	0.58	2.07 [1.33, 2.84]
fold B	28	0.64	2.60 [1.86, 3.36]
Olmo-3 RL, fold A	76	0.09	50.82 [48.51, 53.11]
fold B	82	0.09	50.50 [48.18, 52.85]

Table 56: Per-word counts against each transition's own reference, over each transition's own universe of frequent word-initial ids. A word counts as a reversal when its own counted change has an interval above zero and both of its changes in propensity are at or below zero at the point estimate. The reference resamples whole prompts, so it keeps the words' dependence on one another; it is Equation (9), and the interval reading of a reversal, with its own reference, is in Table 60. Counts of words; mass share is the reversing words' share of the summed absolute per-word change on that transition.

Transition	Words	Rises	Reversals	Fraction	Mass share	Null mean	Null p99
Qwen2.5-7B SimpleRL	670	116	27	0.23	0.05	15.56	27
Llama-3.1-8B SimpleRL	737	134	33	0.25	0.09	16.94	30
Mistral-7B-v0.1 SimpleRL	666	144	37	0.26	0.09	15.76	29
Olmo-3 RL	587	253	86	0.34	0.20	14.24	27

that reference's center. Between the predecessor's own two draws, where the true change is zero, 32 forms rise on these prompts against the transition's 48.

L.3 Dependence-Preserving Nulls, Validated

The per-word rate. Fix a transition and a word universe W. For prompt i and each of the four teacher-forced cells c, the pass stores word w's expected count a_{iw}^c and the prompt's scored-position count d_i^c, and the rate is

$$R_w^c \;=\; 1000\,\frac{\sum_i a_{iw}^c}{\sum_i d_i^c}. \tag{8}$$

Table 57: The per-word reversal fraction across 13 released transitions. Every row is computed over the word universe its own measurement pass carried: the class ids that pass scored, plus its most frequent word-initial ids. The 10 census rows come from the census pass, so their universe holds the published excess-vocabulary list's words; the three SimpleRL-Zoo rows come from the pass that carries this paper's connective class. The Olmo-3 RL transition sits in both, and its fraction differs between the two tables for that reason. The fraction is a count over the transition's own risen words; the mass share, computed for the four transitions of the scan above, is those words' share of the summed absolute per-word change and is the smaller number in every case.

Predecessor	Successor	Fraction	Mass share
Qwen2.5-7B	SimpleRL-Zoo	0.23	0.05
Llama-3.1-8B	SimpleRL-Zoo	0.25	0.09
Mistral-7B-v0.1	SimpleRL-Zoo	0.26	0.09
Qwen2.5-7B	Instruct	0.13	—
Qwen2.5-1.5B	Instruct	0.35	—
Qwen2.5-14B	Instruct	0.06	—
Llama-3.1-8B	Instruct	0.18	—
Mistral-7B-v0.3	Instruct	0.02	—
Gemma-2-9B	it	0.11	—
OLMo-2-7B	Instruct	0.08	—
OLMo-2-13B	Instruct	0.09	—
Olmo-3	Instruct	0.06	—
Olmo-3	RL-Zero-Code	0.29	—

Table 58: The 20 forms of the held-out class and their increases on the discovery half, largest first. The class is chosen on this half by counted change alone, and " and" leads it by a wide margin. Counted change per 1,000 tokens with 95% bootstrap intervals over the discovery prompts; a leading space is part of the form. The lower endpoint of " changes"'s interval is above zero but rounds to zero at three decimal places.

Form	Change per 1,000 tokens	Form	Change per 1,000 tokens
" and"	1.39 [0.64, 2.13]	" well"	0.13 [0.03, 0.25]
" as"	0.36 [0.08, 0.63]	" team"	0.12 [0.01, 0.28]
" such"	0.30 [0.17, 0.44]	"Therefore"	0.12 [0.07, 0.17]
" good"	0.21 [0.01, 0.44]	" set"	0.12 [0.01, 0.23]
"Let"	0.17 [0.06, 0.29]	" part"	0.11 [0.03, 0.21]
" John"	0.15 [0.06, 0.27]	" various"	0.11 [0.04, 0.19]
"a"	0.15 [0.02, 0.33]	" specific"	0.102 [0.003, 0.207]
" cost"	0.14 [0.02, 0.27]	" over"	0.10 [0.01, 0.20]
"So"	0.13 [0.01, 0.27]	" employees"	0.09 [0.02, 0.16]
" changes"	0.129 [0.000, 0.322]	" skills"	0.08 [0.02, 0.16]

Every term of the decomposition is a difference of two of these four rates, word by word. Uncertainty is a prompt bootstrap: draw a multiplicity vector $k^{(b)} \sim \mathrm{Multinomial}(n, 1/n)$ over the n prompts and recompute each rate as $1000 \sum_i k_i^{(b)} a_{iw}^c / \sum_i k_i^{(b)} d_i^c$. One draw is shared by every word and every cell, and that is what *dependence-preserving* means here: two words that rise on the same prompts have replicates that move together, so a count cannot be made large simply by evaluating a large universe.

The dependence-preserving null. The reference for a count of reversals recenters those same replicates on the value they are drawn around and asks how often the recentring alone produces one. Let $\bar{c}_w^{(b)}$ be word w's change-in-propensity replicate, $d_w^{(b)} = \bar{c}_w^{(b)} - \bar{c}_w$ its centered form, and s_w the sign of that word's own counted change. The count in replicate b is

$$N^{(b)} \;=\; \#\big\{\, w \in W \;:\; d_w^{(b)}\, s_w \;<\; -h_w \,\big\}, \tag{9}$$

the words whose centered change-in-propensity replicate clears their own half-width on the side opposite their observed change: a reversal that is not there. What the reference negates is that the

Table 59: The held-out class on the held-out prompts, and the per-word count on the same prompts. The class rises with both changes in propensity below zero; removing its largest member leaves a class that rises with both changes in propensity above zero, and the per-word count of reversals sits at the center of its reference. Rates per 1,000 tokens with 95% intervals over the held-out prompts; the reference resamples whole prompts.

Quantity	Value
Predecessor on its own trajectories	32.26
Predecessor on the successor's trajectories	34.28
Successor on the predecessor's trajectories	32.12
Successor on its own trajectories	34.06
Counted change	1.80 [0.80, 2.76]
Change in reach	2.02 [1.04, 2.98]
Change in propensity (successor's contexts)	−0.22 [−0.29, −0.17]
Change in propensity (predecessor's contexts)	−0.13 [−0.18, −0.08]
Forms whose counted change is resolved above zero	48
Forms that reverse	18
The reference's mean	17.77
Its 95th percentile	28
Its 99th percentile	33
Forms rising between the predecessor's two draws	32
" and", share of the class by expected count	0.67
The other 19 forms, counted change	0.83 [0.21, 1.43]
at the successor's contexts	0.21 [0.18, 0.24]
at the predecessor's contexts	0.23 [0.20, 0.26]

observed count is what dependent sampling noise produces over a universe of this size at these interval widths.

The exchangeability null. A different question, whether the two models can be told apart at all, is asked by exchanging their labels. For each prompt independently, with probability one half, swap that prompt's four cells; at the prompt level the swap negates that prompt's counted change and exchanges the two references. Recompute the per-word rates, terms, intervals and count on the swapped arrays under the same bootstrap. The counts over many swaps are the reference. What it negates is any account under which the observed count comes from the estimator, the universe, the clustering or the four-cell geometry alone, all of which survive the swap. The two references answer different questions and are not interchangeable.

The pipeline on a pair with no change. The whole pipeline (the universe, the selection of the 20 largest risers on the discovery half, and the held-out evaluation) was run on two draws of the same predecessor model, where the change is zero by construction. The held-out class's counted change is 0.30 [−0.35, 0.91] per 1,000 tokens, an interval containing zero, so the selection does not manufacture a class-level rise.

L.4 Sign Reversals, Word by Word

On both transitions of Figure 11, the count of individually resolved reversals exceeds the 99th percentile of its dependence-preserving null (Appendix L.3); Figure 12 gives the reach share and the signed ratio under the predecessor reference for the connective class on the transitions it lists, except where the floor rule of Appendix C.4 withholds them.

L.5 The Decomposition Under Greedy Decoding

Every decomposition in this paper is computed on trajectories drawn at a temperature of 0.7 from the top 0.95 of the mass, with the decoding transform mirroring that draw, so the sign of a change in propensity could in principle be a property of the transform rather than of the models. Computed instead on greedy trajectories, with the decoding transform replaced by the one-hot at the argmax,

Table 60: Per-word reversal counts against two references. The dependence-preserving reference asks whether a count exceeds what correlated sampling noise produces over a universe of this size; the exchangeability reference asks only whether the two models can be told apart. The Olmo-3 RL transition's count exceeds both; DeepSeek-Math-7B's held-out count exceeds the second and sits at the center of the first. Counts of words; *point* is the reading of Table 56, a word whose counted change has an interval above zero and whose two changes in propensity are both at or below zero at the point estimate, and *strict* is a different, non-nested reading, a word whose counted change and the mean of its two changes in propensity both have intervals excluding zero with opposite signs, so it also counts falls; *exceedance* is the share of swaps whose count reaches the observed one.

Object	Reading	Words	Count	Dep.-pres. p99	Exch. mean	Exch. p99
Olmo-3 RL, model-written	point	587	86	27	1.30	4
	strict	587	125	27	1.58	8
DeepSeek-Math-7B, held out	point	735	18	33	6.24	13
	strict	735	18	33	1.64	6

The exceedance share against the exchangeability reference is 0.00 for the Olmo-3 RL transition and 0.00 for the held-out half, under either reading.

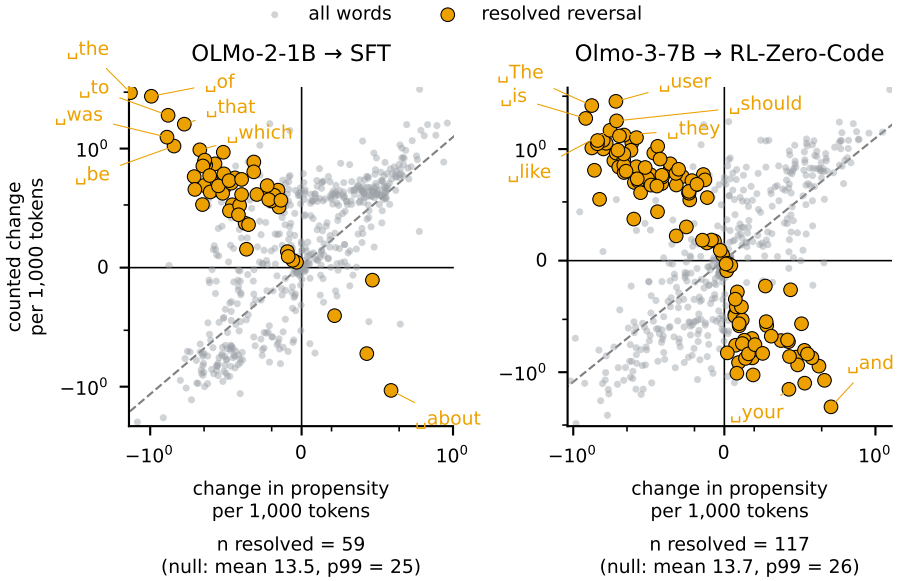


Figure 11: The sign of a word's counted change does not fix the sign of its change in propensity. One point per frequent word on the screened prompts: horizontally its change in propensity, the mean over the two references; vertically how much more often the successor writes it, both per 1,000 tokens. A point above the horizontal zero line and to the left of the vertical one has a positive counted change while its change in propensity is negative; the colored points are those whose two intervals both exclude zero with opposite signs, the eight largest by counted change labeled. Axes symmetric-log.

the Olmo-3 RL transition's connective class still rises. Its change in propensity stays below zero under the successor reference (-0.27 [-0.37, -0.17] per 1,000 tokens), and under the predecessor reference it is negative at the point with an interval that includes zero (-0.041 [-0.082, 0.005]). The rise in the published excess-vocabulary list is still carried by the change in reach on both instruction-tuned transitions (Table 61).

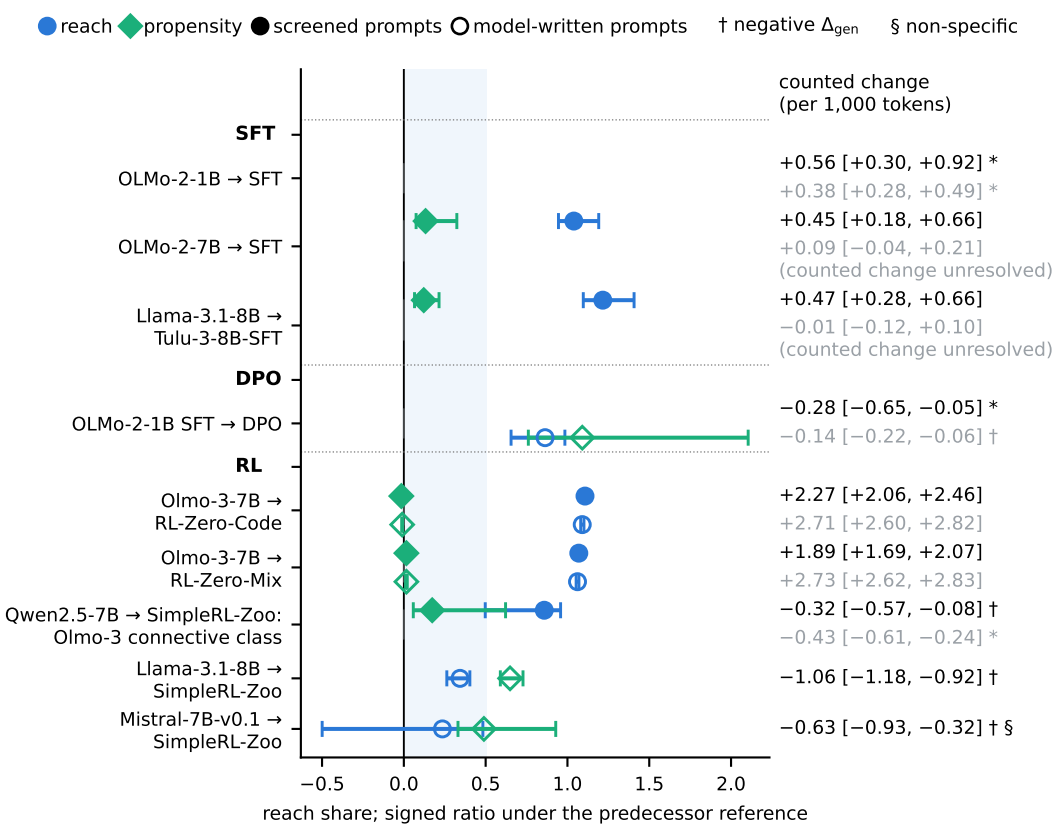


Figure 12: Which of the studied transitions the decomposition changes the reading on, measured by how far each circle sits from the middle: far right, the change in reach accounts for the row's whole counted change; a diamond below zero, the successor moved against the class at the predecessor's contexts; near the middle, a count would have been roughly right. Circles are the reach share, diamonds the signed ratio under the predecessor reference. Filled markers: screened prompts; hollow: model-written prompts. Counted changes per 1,000 tokens at the right edge; an asterisk marks a row whose shares the floor rule of Appendix C.4 withholds.

M Reproducibility and Experimental Protocol

M.1 Claim Index

Table 62 maps every main-text number to its event, transition, prompt population, rendering, generation cap and scoring window. A row whose generation cap or rendering differs from the default (384 tokens, one rendering) is marked with *; a row whose scoring window differs from the whole response is marked with †.

M.2 Artifact and Code Mapping; Data Provenance

Every result in this paper is computed from stored experiment artifacts, each a JSON file recording the design, the raw measurements and the bootstrap replicates. The artifacts, the scoring code that produces them, the test suite that checks them, and the figure-generation code are released alongside this paper.

Table 61: Three transitions under two decoding rules. Under greedy decoding the Olmo-3 RL transition's connective class rises; its change in propensity is below zero under the successor reference, and under the predecessor reference it is negative at the point with an interval that includes zero. Under the sampler the rest of the paper uses, both are below zero (Table 23). Both instruction-tuned transitions' rise in the published excess-vocabulary list stays majority change in reach. Rates per 1,000 tokens with 95% intervals; the two propensity rows are the same quantity computed on each model's contexts and the mean row is the mean of them; the reach share is of that block's own counted change. The instruction-tuned transitions are evaluated with their predecessor prompted bare. The lower block gives each transition's values under the sampler the rest of the paper uses; the last two rows are each arm's share of trajectories carrying a repeated block of twenty tokens.

Quantity	Olmo-3 → RL-Zero-Code connective class	Qwen2.5-7B → Instruct published list	OLMo-2-7B → Instruct published list
Greedy decoding			
Counted change per 1,000 tokens	2.64 [2.12, 3.01]	1.19 [0.75, 1.57]	3.69 [3.11, 4.26]
At the successor's contexts	−0.27 [−0.37, −0.17]	−0.48 [−0.70, −0.28]	1.62 [1.33, 1.93]
At the predecessor's contexts	−0.041 [−0.082, 0.005]	0.41 [0.18, 0.64]	1.49 [1.18, 1.81]
Change in propensity, mean	−0.15 [−0.21, −0.10]	−0.03 [−0.19, 0.12]	1.56 [1.33, 1.79]
Reach share	1.06 [1.04, 1.09]	1.03 [0.90, 1.20]	0.58 [0.51, 0.64]
The sampler the rest of the paper uses			
Counted change per 1,000 tokens	2.71 [2.60, 2.82]	1.19 [0.99, 1.41]	3.91 [3.53, 4.30]
At the successor's contexts	−0.25	−0.39	1.47
At the predecessor's contexts	−0.03	0.43	2.34
Reach share	1.05 [1.05, 1.06]	0.98 [0.93, 1.03]	0.51 [0.47, 0.55]
Looping trajectories, predecessor	5 of 600 (0.8%)	1 of 600 (0.2%)	17 of 600 (2.8%)
Looping trajectories, successor	7 of 600 (1.2%)	1 of 600 (0.2%)	0 of 600 (0.0%)

Table 62: What each main-text number is measured on. Every row corresponds to one claim or one family of claims in the main text; the column headings name the event, the model transition, the prompt population, the rendering and the generation cap. ⋆Rows whose generation cap or rendering differs from the default (384 tokens, one rendering). †Rows whose scoring window differs from the whole response.

Claim	Event	Transition	Population	Rendering	Cap	Window
Counted change (Section 4.3)	conn. class	Olmo-3 RL	model-written	template	384	whole response
Propensity (Section 4.3)	conn. class	Olmo-3 RL	model-written	template	384	whole response
Census breadth (Section 4.2)	published list	10 transitions	model-written	template	384	whole response
Census human (Section 4.2)	published list	10 transitions	human-written	template	384	whole response
MiMo ordinal (Section 4.3)	ordinal class	MiMo	mathematics	bare	384	whole response
Llama conn. (Section 4.3)	conn. class	Llama SimpleRL	model-written	literal	384	whole response
Llama conn., math (Section 4.3)	conn. class	Llama SimpleRL	mathematics	literal	384	whole response
Openings† (Section 4.3)	para. openings; conn. class	Olmo-3 RL	model-written	template	384	paragraph openings
Stopping⋆ (Sections 4.3, 5.2 and 6)	response term.	Olmo-3 RL	model-written	template	384; 2,048; 8,192	whole response
Census cap⋆ (Sections 4.4 and 7)	published list	4 transitions	model-written; human-written	bare	2,048	whole response
Context corr.† (Section 5.2)	conn. class	Olmo-3 RL	model-written	template	384	continuation after the first paragraph
Local corr.† (Section 5.2)	conn. class	Olmo-3 RL	model-written	template	384	continuation after the first paragraph
Qwen context⋆† (Section 5.2)	published list	Qwen2.5-7B Inst.	model-written	bare vs. template	384; 256 new tokens	continuation after the first paragraph
Qwen local⋆ (Section 5.2)	published list	Qwen2.5-7B Inst.	model-written	bare vs. template	384	whole response
Llama local (Section 5.2)	conn. class	Llama SimpleRL	model-written	literal	384	whole response
Llama exchange⋆† (Section 5.2)	conn. class	Llama SimpleRL	fresh model-written draw, 4,804 eligible prompts	literal	384; 256 new tokens	continuation after the first paragraph
Llama window split† (Section 5.2)	conn. class	Llama SimpleRL	model-written; prose-opening trajectories	literal	384	continuation after the first paragraph
Llama same-trajectory split (Section 5.2)	conn. class	Llama SimpleRL	model-written; prose-opening trajectories	literal	384	whole response
Other exchange windows† (Section 5.2)	conn. class; published list	Olmo-3 RL; Qwen2.5-7B Inst.	model-written	template	384	continuation after the first paragraph
Stopping local⋆ (Section 5.2)	response term.	Olmo-3 RL	model-written	template	2,048	whole response
Stopping context⋆† (Section 5.2)	response term.	Olmo-3 RL	model-written	template	2,048	after the cut (three cut points)
Control, Llama exchange⋆† (Sections 5.2 and 6)	control class vs. conn. class	Llama SimpleRL	fresh model-written draw, 4,804 eligible prompts	literal	384; 256 new tokens	continuation after the first paragraph
Control, other arms⋆† (Sections 5.2 and 6)	control class vs. counted class	Olmo-3 RL; Qwen2.5-7B Inst.; Llama SimpleRL	as the arm's row	as the arm's row	as the arm's row	as the arm's row

⋆Rows whose generation cap or rendering differs from the default (384 tokens, one rendering). †Rows whose scoring window differs from the whole response.