THE WRITE WINDOW IN EMBODIED MULTIMODAL MEMORY: AN ACCESS STRUCTURE IS A PAIR, AND CAPABILITY IS A PRODUCT

Anonymous authors
Paper under double-blind review

ABSTRACT

Memory architectures for embodied multimodal agents differ in how they index, compress and retrieve, and agree silently on what they *write*: a description of one frame. We make the write window an architectural parameter, formalise an access structure as a pair (a write map over a window of observations and an address algebra over its records, no address richer than the write it runs over), and read capability for conjunctive queries as a product of an identity, an ordinal and a percept factor, so that which component binds is simply which factor is smallest. Two walls follow. The predicates a frame-local write can decide are closed under time reversal, so no reversal-paired verb (*put* against *take*, *open* against *close*) is writable at any state budget, and any direction a single frame supports is bounded by the majority prior. And an address that indexes by count compounds per-occurrence detector error, so a later occurrence resolves with a probability that falls geometrically in its rank. Both are measured on the ground-truth locator frames of eight EPIC-KITCHENS streams: the standard single-frame write names the verb barely more often than not, and its error mass sits exactly on the reversal pairs. AXIS widens the window to an ordered pair, differences per-object statuses into events, retains every occurrence with its witness, and executes addresses rather than searching by similarity. The registered controls then refuse the obvious reading: the pair's gain is attributable to the prompt's enumerated verb list rather than to the window, since the same frame attached twice gains as much, and direction under every listed rule sits at or below the majority prior. What does escape the prior is a property of the pair's *spacing* rather than of the pair (one backbone reads direction above the prior when the two frames are a second apart under a free-vocabulary prompt and not when they are one action apart), and an independent public benchmark agrees at the same spacing with the read held fixed. Behind the window the ordinal wall binds: occurrences are detected at a recall and precision no scaffold-side filter improves. Replacing the written ledger by one built from the corpus's own annotations, everything else held fixed, roughly doubles the same loop and lets the address reach its target far more often, which locates the entire end-to-end loss on the write rather than the read. Of the registered predictions, those about the measurement held and those about the write side and the end-to-end result failed; we report all of them and relocate the binding wall to detector reliability at a spacing the detector can actually read.

1 INTRODUCTION

An agent that answers questions about a long observation stream under a fixed per-decision token budget cannot hold the stream in its context. It writes something as the stream arrives and reads from what it wrote. The recent literature on memory for long-horizon multimodal agents has been an arms race on the read side of that sentence: hierarchical caption summaries (Song et al., 2024; He et al., 2024), agentic frame selection (Wang et al., 2024; 2025), temporal indices (Tian et al., 2026), interaction records with recurrence statistics (Das & Roy, 2026), ordinal anchors (Li et al., 2026) and operator algebras over versioned writes (Wang, 2026) differ in everything except one

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.

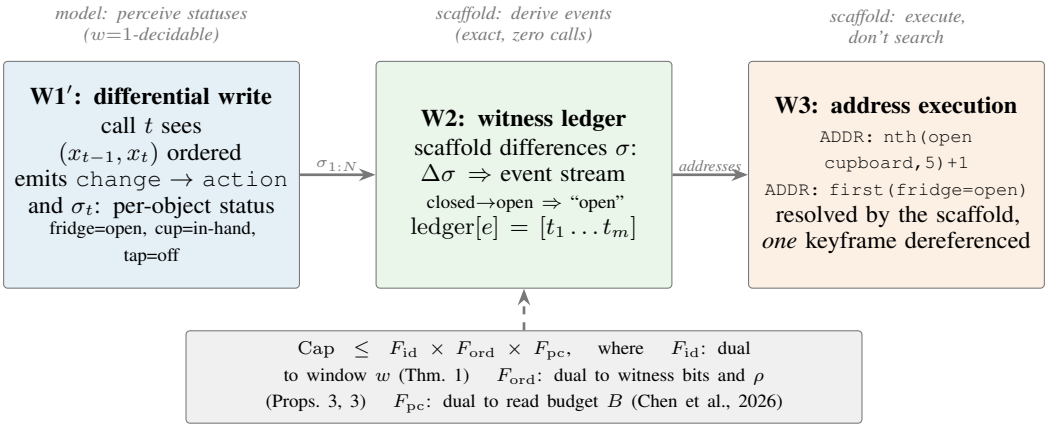


Figure 1: Architecture, and where each factor of the product is paid for. The write map reports what a single frame *can* decide (statuses); the scaffold derives what it cannot (verbs) by differencing, retains every occurrence with its witness, and executes ordinal and state addresses instead of searching by similarity. The measurement of §4 says which factor each component moved: W1′ raised gold-frame verb accuracy by $+8.6$ to $+11.5$ points, which its own controls attribute to vocabulary and not to the window at one action apart; W2 and W3 are exact and free but run over occurrences detected at rate $\rho \approx 0.3$, and the ordinal factor is ρ^k.

commitment, which none of the papers we surveyed states as an assumption: *the unit written to memory is a description of a single frame*. The write side has begun to move too (event-triggered writing (Guo et al., 2026), selective allocation (Ge et al., 2026), stratified consolidation (Gong et al., 2026)), but along the rate, the budget and the span, and Corollary 3 shows a triggered write with a frame-local emission is still $w{=}1$. An unvaried convention is untested.

This paper tests it from a symmetry rather than from a benchmark. A still frame is invariant under time reversal, so whatever a write map decides from one frame it decides equally from that frame in the reversed stream; and by induction so does any record built from single frames, however large. *Put* and *take*, *open* and *close*, *fill* and *empty* are one hand–object configuration traversed in opposite directions, and whatever is built downstream of such a write (however good the index, the compression policy or the retriever) indexes content from which the verb is already gone. We call this the *identity wall* (Theorem 1); its statement contains neither the state size, nor the retrieval width, nor the read budget, which are the three axes the field scales.

The wall ought to be visible in any caption-based memory's own cache, and it is. Over 7,264 keyframes of eight EPIC-KITCHENS-100 streams (Damen et al., 2022) captioned by the standard single-frame update of a companion system (Anonymous authors, 2026f) under three open backbones, the verb at the ground-truth locator frame (the segment midpoint, the most favourable single point) is correct 11.0%, 15.1% and 23.3% of the time. Gold *put* is written *take* 46.1% of the time by the strongest of the three, gold *close* written *open* 26.1%. The confusion is not diffuse; it is the reversal.

What we do. We formalise an access structure as a *pair* $(\mathcal{W}, \mathcal{A})$: a write map over an observation window of length w and an address algebra executable over its records, with the constraint $\dim \mathcal{A} \leq \dim \mathcal{W}$: an address $\mathrm{nth}(e, k)$ is well defined only if e is a predicate the write decides and its k-th witness survived storage. For the conjunctive query family (address a moment by a reduction over the stream, then read a percept there), capability factors as a product $\mathrm{Cap} \leq F_{\mathrm{id}} \cdot F_{\mathrm{ord}} \cdot F_{\mathrm{pc}}$ (Proposition 1), and the three factors are dual to three different resources: the identity factor to the window w, the ordinal factor to per-class witness storage and detector reliability, and the percept factor to the read budget B where the walls of Chen et al. (2026) live. A product is bounded by its smallest factor, so "which component of a memory architecture is binding" becomes "which factor is smallest", and it has a measured answer.

The wall's proof names its own escape. $\mathcal{W}$ separates e from its reverse e^R exactly when the record retains a variable that e mutates (a door's openness, an object's being in hand, a tap's running), and that variable *is* a property of a single frame. AXIS (Addressed eXecution over Indexed State) therefore widens the window to an ordered pair, asks the model for what a frame can decide (per-object *statuses*),

and has the scaffold take the first difference of the status trajectory: the event stream becomes a derivative, computed exactly and for free, instead of a perception task the model measurably fails. Every occurrence is retained with its witness frame, and a small address algebra ($\mathrm{nth}(e,k)\pm d$, first, last, count, and the state address $\mathrm{first}(obj{=}status)$ that only a trajectory-retaining write can define) is executed by the scaffold, dereferencing one keyframe verbatim. The two channels compose by *address* (the state channel returns a location, the verbatim channel is read there) rather than by concatenation, which is the composition the companion system used and measured to fail.

What the measurement says, in the order we learned it. One scoring function reads every cache: the written verb and the gold verb are both mapped through the corpus's published class tables with inflection stemming, so *placing* and *put-down* are one verb and a rule cannot gain by spelling. Under it the field's single-frame write is right about the verb at the gold frame 11.0, 15.1 and 23.3% of the time, and the ordered pair under the same prompt raises that to 22.5, 24.9 and 31.8%, below the registered 15-point threshold on every backbone and above its 5-point falsifier on every one. Then the controls registered for exactly this purpose land: the same frame attached *twice*, with no temporal information at all, scores 26.9%, *above* the ordered pair; the pair in reversed order scores 22.7%; and among writes that are *put* or *take* on a *put*/*take* gold frame, direction is right 58.8% for the pair, 57.4% for the duplicate and 50.7% for the reversed pair, against a majority prior of 74%. The rise was the prompt's enumerated vocabulary landing on the *put*/*take* mass of the locator frames.

A registered amendment then removes the list and varies the spacing, which separates the two explanations the controls left standing. With a free vocabulary the one-action pair scores 11.4% on gemma-3-12b (*below* its single frame), and the same frame twice under the free prompt collapses to the static verb *hold* on qwen3.8-27b (0.0% verb-correct, one distinct verb over 490 frames), which is the identity wall in its purest form: shown no change, and given no list to guess from, the writer names no transition. Then the spacing is moved and everything else held: the *same* free prompt over a pair one *second* apart writes the direction of a *put* correctly 82.7% of the time on qwen3.8-27b against 70.5% one action apart ($+12.2$ points, $z{=}2.08$, $p{=}0.037$). It is the only write rule in the paper whose direction beats the 74% majority prior ($p{=}0.025$), and the comparison that establishes it varies nothing but the gap between the two frames. On gemma-3-12b the same move does nothing; on gemma-4-31b the zero-spacing collapse replicates but the above-prior direction read does not. TempCompass, an independent public benchmark with frames one second apart and the read fixed, agrees on the split.

Contributions. (1) The access structure as a pair $(\mathcal{W},\mathcal{A})$ with $\dim\mathcal{A}\le\dim\mathcal{W}$, and the product law $\mathrm{Cap}\le F_{\mathrm{id}}\cdot F_{\mathrm{ord}}\cdot F_{\mathrm{pc}}$, each factor dual to a different resource. (2) **The identity wall**: at $w{=}1$ the writable predicates are closed under time reversal, so reversal-paired verbs are unwritable at any state budget, and the closure breaks at $w{=}2$; the statement contains none of the field's scaling axes. (3) **Two forms of the ordinal wall**: the storage form ($O(1)$ words per class cannot answer interior nth) and the error-propagation form ($P[\mathrm{nth}(e,k)]=\rho^k$ under per-occurrence recall ρ), with the proof that no dereference width $\le k$ escapes it; the second is the one the measurement says binds. (4) The identity factor measured on the standard single-frame write, at the most favourable frame, under one class-canonical scorer: 11.0–23.3%, with the error mass on the reversal pairs the theorem names. (5) AXIS itself, every design decision derived from a wall. (6) The write repair measured and **refuted by its own registered controls** at one action apart. (7) **The spacing of the pair, isolated as the parameter that decides whether the window's escape is taken.** (8) **The read side priced alone** by an oracle ledger, which locates the entire end-to-end loss on the write.

What is not new here, stated plainly. Ordinal addressing over repeated events is the `occurrence_index` of Li et al. (2026); interaction records with timestamps and their consolidation into recurrence statistics are the memory of Das & Roy (2026); the observation that a differenced state trajectory is an event stream is older than either and is the working assumption of every scene-graph planner that tracks object state (Rana et al., 2023; Gu et al., 2024). What is new is the object the walls are stated about (the write window as a parameter with a sharp threshold at $w{=}2$), the product law that says which factor binds, the measurement that the field sits under the first while optimising the third, and the finding that raising the first relocates the binding constraint to a detector reliability no read-side machinery can touch.

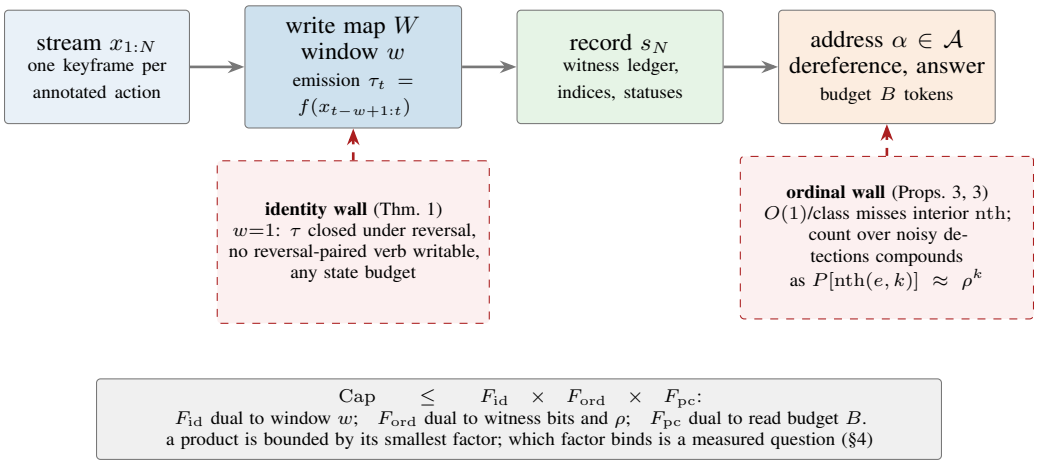


Figure 2: The access structure $(\mathcal{W}, \mathcal{A})$ as a pipeline, and where the two walls sit. The write map sees a window of w observations; the record retains what the write emitted; the read executes addresses over the record under a fixed budget B. The identity wall is a statement about the write's window, the ordinal wall about the record's addresses over detections; the product law assigns each factor to a different resource, so "which component of the memory binds" becomes "which factor is smallest".

2 Theory: An Access Structure Is a Pair, and Capability Is a Product

Definition 1 (Access structure)**.** An *access structure* over a stream $x_{1:N}$ is a pair $(\mathcal{W}, \mathcal{A})$. The write map $\mathcal{W}$ takes the current record and an observation window $x_{t-w+1:t}$ of length w and returns the next record; it is applied online, once per keyframe, and the record is all the structure retains. The address algebra $\mathcal{A}$ is the set of operations the read side can execute over the record without re-reading the stream. A predicate e on stream positions is *w-writable* if some write map with window w decides it *at the position*: if the emission the map makes at t (the part of its output that is a function of the window and not of the record it inherits) determines $e(t)$. $\dim \mathcal{W}$ denotes the w-writable predicates of the write in use, $\dim \mathcal{A}$ the predicates the algebra's operations are well defined over.

The pair is the object the field has been treating as a singleton: every architecture in §6 is an $\mathcal{A}$ (an index, a compression policy, a retriever) built over a $\mathcal{W}$ that is never named because it is always the same.

Definition 2 (Conjunctive query family; witness)**.** A query is *conjunctive* if $a(q) = V(x_{L(x_{1:N})})$: a locator L computes a position by a reduction over the stream ("the k-th time event e happened, d steps later") and a percept V is read there; the *ordinal* subfamily has $L = \mathrm{nth}(e, k) + d$. A *witness* for "$e$ occurred at t" is t itself together with whatever the record retains from x_t that lets the read side dereference it, and a record *retains the k-th witness* of e if $\mathrm{nth}(e, k)$ is computable from the record alone.

A companion measures this family's incidence in the field's embodied multimodal benchmarks at 0.51% pooled and builds a corpus that makes it measurable (Anonymous authors, 2026a); everything below concerns what an architecture must *be* for that availability to be used.

Proposition 1 (Product law)**.** *For a conjunctive query whose locator names the k-th occurrence of event e,*

$$\mathrm{Cap} \;\leq\; \underbrace{P[\,e \text{ decided}\,]}_{F_{\mathrm{id}}} \times \underbrace{P[\,k\text{-th witness resolved} \mid e \text{ decided}\,]}_{F_{\mathrm{ord}}} \times \underbrace{P[\,\text{percept read} \mid \text{position}\,]}_{F_{\mathrm{pc}}} \;+\; \varepsilon_{\mathrm{guess}},$$

with $\varepsilon_{\mathrm{guess}}$ the blind floor. Moreover $\dim \mathcal{A} \leq \dim \mathcal{W}$: an operation $\mathrm{nth}(e, k)$ is well defined only if e is $\mathcal{W}$-decidable and its witnesses survive storage. ■

The law says two things. The first is arithmetic: a product is bounded by its smallest factor, so which component binds is which factor is smallest, and each factor is measurable on its own sub-population.

The second is structural: the algebra cannot be richer than the write; an address is an expression over tokens, and if the write never emitted the token the expression has nothing to run on, however elegant the algebra.

Theorem 1 (Identity wall)**.** *Let R denote time reversal of the stream. The set of predicates decided* at the position *by a write map with $w{=}1$ is closed under R. Consequently, for any pair of events e, e^R that are reversal conjugates with reversal-symmetric single-frame marginals (*put/take*,* open/close*,* fill/empty*) and non-degenerate prior $0 < \pi < 1$, no $w{=}1$ emission decides e against e^R better than the prior,* at any state budget*; and for any query family whose locator is such an event, with the ordinal drawn among the pair's occurrences under exchangeable interleaving,* $\mathrm{Cap} \to \varepsilon_{\mathrm{guess}}$ *as the occurrence count grows, independent of the index, the retriever and the read budget. An ordered window of length $w \geq 2$ is not R-invariant, so the closure breaks at $w{=}2$.* ■

What the theorem does not say. It does not say a one-frame system can never record a verb: a write map that inherits a record may compare its emission at t against a status the record already holds, which is a window of length two moved from the write into the algebra, exactly what AXIS's write rule does. What it says is that the *emission* cannot carry the verb, so a pipeline whose record is a bag of emitted descriptions has nothing to compare and inherits the wall in full. Two corollaries follow: $\mathcal{P}_1 \subsetneq \mathcal{P}_2 \subseteq \cdots$, so the write window bounds the temporal order of the predicates a memory can hold (Corollary 2); and a *triggered* write (thresholding a score computed from frame-wise features, with a frame-local emission) is still $w{=}1$ (Corollary 3).

Proposition 2 (Statuses are 1-writable; their transitions are 2-writable and exact)**.** *A per-object status (a door open or shut, an object in hand or not, a tap running or not) is decided by a single frame, so it is 1-writable. The first difference of a status trajectory over an ordered pair determines the transition, so every reversal-conjugate event is 2-writable, and the difference is computed by the scaffold rather than perceived: it is exact given the statuses.* ■

Proposition 3 (Ordinal wall, two forms)**.** (Storage.) *Under per-class storage of $O(1)$ words (a recurrence count and a last-seen timestamp),* $\mathrm{nth}(e, k)$ *is unanswerable for all $k \notin \{1, m\}$ at any perception quality: addressing the k-th of m occurrences over a stream of length N requires $\Omega(m \log(N/m))$ bits per class.* (Error propagation.) *If each true occurrence of e is detected independently with probability $\rho < 1$ and detections are otherwise exact, then $P[\mathrm{nth}(e, k)$ resolves to the true k-th position$] = \rho^k$, and reading the resolved occurrence together with its r detected neighbours on either side raises this to $\rho\, P[\mathrm{Bin}(k{-}1, 1{-}\rho) \leq r]$, which for every r is at most ρ, the recall of the addressed occurrence itself.* ■

The storage form is the one we registered, and it locates Das & Roy (2026): a recurrence count and a last-seen time are the two words it names, and a benchmark closed under $\{\mathrm{count}, \mathrm{last}, \mathrm{most}, \mathrm{usually}\}$ never asks for an interior occurrence, so the wall is not exposed. It is also the form the measurement says is *not* binding once the storage is fixed: the witness ledger retains every occurrence, the bound is met, and the ordinal factor is still small, because of the second form, and because Proposition 8 shows no scaffold-side filter raises the ceiling.

3 AXIS, AND THE EXPERIMENTAL DESIGN

Three components, each derived from a wall. **W1′**: the write asks the backbone for per-object *statuses* over an ordered pair of frames and the scaffold differences the status trajectory into events, using a transition table fixed before any cell ran: verbs are derivatives, not perceptions (Proposition 2). **W2**: every occurrence is retained with its witness frame index, at zero API cost, which meets the storage form of the ordinal wall by construction. **W3**: the address algebra is extended exactly as far as the write licenses ($\mathrm{nth}(e, k) \pm d$, first, last, count, and the state address $\mathrm{first}(obj{=}status)$ that only a trajectory-retaining write can define), and is *executed* by the scaffold, dereferencing one keyframe verbatim rather than searching by similarity.

The evaluation is registered before any paid call. Write rules are compared on the same frames by the same class-canonical scorer, with the registered controls that decide whether a gain is the window or the prompt: the same frame attached *twice* (no temporal information), and the pair *reversed*. End-to-end, the conjunctive corpus of (Anonymous authors, 2026a) is read under a fixed per-decision budget on three frozen open backbones, against a blind floor and two controls that hold the ledger and break the address. Cells above 5% unparsed are reported with their counts and not scored.

Table 1: The write path at the gold locator frames, every rule, one scorer. Verb correct at the annotated segment midpoint of the 490 gold locator frames (v1 corpus), scored by one function (`canon_verb`: EPIC class tables plus inflection stemming) for every rule and backbone. *put→take*: gold `put` written `take`. *direction*: among gold put-or-take frames where the write chose put or take, the fraction that chose the right one; n in parentheses; p is the one-sided binomial probability of that many or more correct under the majority prior 74% (always answering *put*), so a rule only reads direction if its p is small. *verbs*: distinct verbs ever written at these frames. `dup` and `rev` run on gemma-3-12b only (the registration's declared coverage cap); the free-vocabulary rules run on all three backbones. `pair` (differenced) scores the first difference of the status trajectory where a transition fired, on the sub-population where a difference is defined. The free-prompt rules (`_free`) receive no verb list; `delta_free` is the only rule whose two frames are one *second* apart rather than one action apart.

backbone	write rule	n	verb	put→take	direction (n)	p	verbs
gemma-3-12b	`single`	490	15.1	10.9	37.5 (24)	1.00	12
gemma-3-12b	`pair`	490	24.9	23.4	58.8 (204)	1.00	7
gemma-3-12b	`dup`	490	26.9	21.9	57.4 (209)	1.00	7
gemma-3-12b	`rev`	490	22.7	30.5	50.7 (209)	1.00	7
gemma-3-12b	`pair_free`	490	11.4	3.9	70.0 (40)	0.78	35
gemma-3-12b	`dup_free`	490	10.6	0.8	80.0 (25)	0.34	36
gemma-3-12b	`delta_free`	490	11.2	3.9	63.2 (38)	0.95	34
gemma-3-12b	`pair` (differenced)	198	23.2	38.8	55.7 (79)	1.00	4
qwen3.8-27b	`single`	490	23.3	46.1	43.5 (108)	1.00	17
qwen3.8-27b	`pair`	487	31.8	27.8	65.6 (180)	1.00	11
qwen3.8-27b	`pair_free`	490	29.4	11.7	70.5 (105)	0.83	42
qwen3.8-27b	`dup_free`	490	0.0	0.0	—	—	1
qwen3.8-27b	`delta_free`	489	33.9	9.4	82.7 (104)	0.02	40
qwen3.8-27b	`pair` (differenced)	316	34.8	29.7	65.7 (140)	0.99	6
gemma-4-31b	`single`	464	11.0	7.2	47.1 (17)	1.00	16
gemma-4-31b	`pair`	489	22.5	30.5	60.0 (150)	1.00	7
gemma-4-31b	`pair_free`	490	23.5	25.8	60.3 (121)	1.00	29
gemma-4-31b	`dup_free`	488	0.0	0.0	—	—	1
gemma-4-31b	`delta_free`	489	22.9	25.8	64.9 (114)	0.99	30
gemma-4-31b	`pair` (differenced)	314	27.4	45.5	56.5 (131)	1.00	6

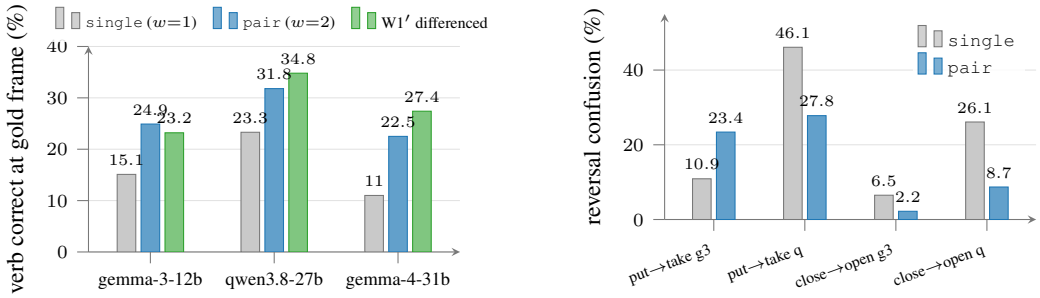


Figure 3: Left: the gold-frame number moves on every backbone, and Table 1's controls show the duplicate frame moves it as far. Verb accuracy at the gold frame, the field's single-frame write against the ordered pair and the differenced stream; one prompt, only the attached frames differ. **Right: the reversal confusions.** Gold *put* written *take* and gold *close* written *open*, gemma-3-12b (g3) and qwen3.8-27b (q). The pair cuts the strongest backbone's *put→take* from 46.1 to 27.8% and *close→open* from 26.1 to 8.7%; on gemma-3-12b the *put→take* rate *rises*, because the pair prompt's enumerated verb list moved that backbone's mass from a diffuse wrong verb onto the conjugate. Both panels are drawn from the same macros as Table 1.

4 Results

4.1 The write path: the identity factor before and after the window

Table 1 is the identity factor under every write rule: the single-frame write is right about the verb at the gold frame 11.0–23.3% of the time, and Table 5 (Appendix E) shows the error mass sits where Theorem 1 says it must: gold *put* written *take*, gold *close* written *open*.

Table 2: E3, the end-to-end grid on the v1 corpus, every backbone and read mode. 245 items per family, B=4096. Superscripts give the unparsed count; † marks a cell above the 5% line, which is reported and not scored (reporting rule 1). _cap256 is Amendment A5's per-round completion cap, the arm that lets the larger backbone commit an answer before its budget is spent; the uncapped cells stay in the table as reported; the _cap256 arm runs on qwen3.8-27b, the backbone with the budget failure. 27 of 27 uncapped and 9 of 9 capped cells measured.

backbone	read mode	RET	TRK	CONJ
gemma-3-12b	axis	22.9	9.8^{6}	14.7
gemma-3-12b	axis_noderef	23.7	10.6^{6}	12.2
gemma-3-12b	axis_wrongaddr	25.7	10.6^{11}	18.4
qwen3.8-27b	axis	$17.1^{72\dagger}$	$18.0^{45\dagger}$	$17.1^{51\dagger}$
qwen3.8-27b	axis_noderef	$11.4^{74\dagger}$	$16.7^{35\dagger}$	$11.4^{81\dagger}$
qwen3.8-27b	axis_wrongaddr	$16.7^{78\dagger}$	$17.6^{35\dagger}$	$16.7^{48\dagger}$
qwen3.8-27b	axis_cap256	22.9^{4}	21.2	$17.6^{16\dagger}$
qwen3.8-27b	axis_noderef_cap256	15.1^{8}	22.4	$17.1^{27\dagger}$
qwen3.8-27b	axis_wrongaddr_cap256	24.9^{6}	22.0	$18.4^{22\dagger}$
gemma-4-31b	axis	$20.4^{19\dagger}$	$9.8^{34\dagger}$	15.5^{12}
gemma-4-31b	axis_noderef	14.3^{10}	$11.4^{43\dagger}$	16.3^{9}
gemma-4-31b	axis_wrongaddr	$21.2^{15\dagger}$	$11.0^{36\dagger}$	18.8^{10}

The ordered pair under the same prompt raises verb accuracy to 22.5, 24.9 and 31.8%, below the registered 15-point threshold on every backbone and above its 5-point falsifier on every one. *The registered controls then refuse the obvious reading.* The same frame attached twice, carrying no temporal information at all, scores 26.9% (above the ordered pair), and the reversed pair 22.7%; direction among *put*/*take* writes on *put*/*take* gold frames is 58.8% for the pair, 57.4% for the duplicate and 50.7% for the reversed pair against a 74% majority prior. The gain was the prompt's enumerated verb list landing on the locator frames' mass, not the window.

The spacing is the parameter, not the pair. A registered amendment removes the list and varies the spacing. With a free vocabulary the one-action pair scores 11.4% on gemma-3-12b, *below* its own single frame, and the same frame twice collapses on qwen3.8-27b to the single static verb *hold* over 490 frames, the identity wall in its purest form. Moving only the gap between the two frames, from one action to one second, the same free prompt writes the direction of a *put* correctly 82.7% of the time on qwen3.8-27b against 70.5% (+12.2 points, z=2.08, p=0.037): the only rule in the paper whose direction beats the majority prior (p=0.025), by a comparison that varies nothing else. On gemma-3-12b it does not move; on gemma-4-31b the zero-spacing collapse replicates (488 frames, one distinct verb) but the above-prior read does not (64.9%, p=0.988). We report both halves. On TempCompass, at the same spacing with the read held fixed, the same split appears on a public benchmark: +9.4 overall and +11.1 on direction for qwen3.8-27b, with a clip and its time-reverse written as opposite motions 12.2% of the time against 1.3%. The larger backbone reads the pair when the pair is close; neither reads it when the pair is an action apart. Event-driven writing is checked separately and is worth at most +2.5 points for direction, which is Corollary 3 measured.

4.2 THE END-TO-END GRID, AND WHAT IT DOES NOT DECIDE

Behind the window the ordinal wall binds: occurrences are detected at a recall and precision no scaffold-side filter improves (Proposition 8), the per-k resolution curve fits $\rho \approx 0.3$ as Proposition 3 predicts, and the end-to-end conjunctive arm is indistinguishable from both its controls. The registered predictions about the write side and the end-to-end result therefore *failed*; those about the measurement held.

4.3 THE READ SIDE PRICED ALONE: THE ORACLE LEDGER

A null on an architecture built to show an effect admits two readings, and the grid cannot separate them. A registered amendment supplies the identity and ordinal factors from the corpus's annotations and runs the *same* read loop, the same address grammar and the same dereference under the same

Table 3: The read side priced alone: the oracle ledger (E3′). gemma-3-12b on the class-semantics corpus (v3, 308 items per family), budget B=4096, all arms. The upper block is the companion measurement's four arms on the same items (the additive bound STATE + RETR − BLIND on CONJ is 12.3). The middle block is AXIS reading its *own* written ledger. The lower block is the same read loop and the same dereference over a ledger built from the corpus's annotations (segment index = keyframe index), declared as an oracle: it holds the identity and ordinal factors at one and prices the percept factor and the loop. `noderef` speaks the address but picks the frame by similarity; `wrongaddr` dereferences a different occurrence of the same event at the same cost. Superscripts: unparsed answers; † above the 5% line. The right block is realised tokens per decision, per family, so that no comparison here is between arms at different budgets.

	accuracy (%)			tokens per decision		
arm	RET	TRK	CONJ	RET	TRK	CONJ
BLIND	18.5	5.2	17.2	75	74	77
STATE-ONLY	22.4[7]	12.0[2]	16.9[5]	964	960	972
RETR-ONLY	10.7	22.4	12.7	2,158	2,157	2,160
ASP	14.9[1]	16.6	16.2	2,682	2,318	2,873
`axis`	14.3	10.4[11]	18.8	2,397	1,905	2,454
`axis_noderef`	13.6	10.7[12]	14.6	2,445	1,932	2,512
`axis_wrongaddr`	16.6	10.7[12]	18.5	2,407	1,903	2,473
`axis_oracle`	**20.8**	**73.0**[1]	**31.5**	2,784	1,941	2,885
`axis_oracle_noderef`	13.6	72.4[1]	13.6	2,858	1,945	2,788
`axis_oracle_wrongaddr`	20.1	74.0[1]	17.2	2,794	1,948	2,872

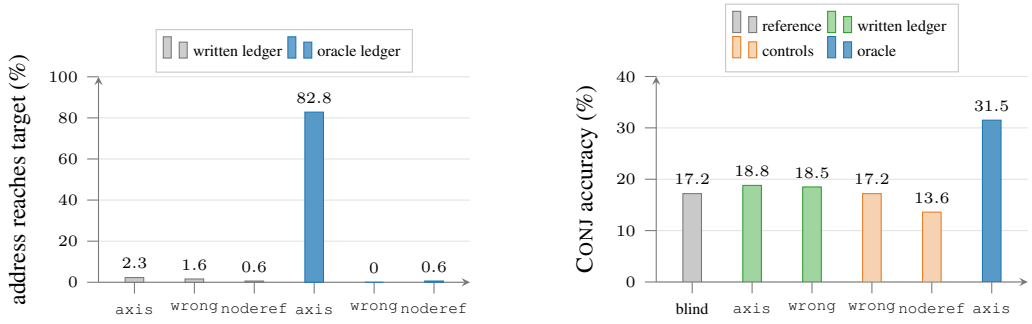


Figure 4: The whole end-to-end loss is on the write. Same read loop, address grammar, dereference and budget; only the ledger's content changes (gemma-3-12b, 308 items per family). **Left:** the executed address reaches the answer's keyframe 2.3% of the time over the ledger AXIS wrote for itself and 82.8% over the oracle ledger. **Right:** that one change moves CONJ from 18.8%, level with its own wrong-address control (18.5%) and with blind (17.2%), to 31.5%, both address-breaking controls separating. The read side executes addresses correctly; the record it reads does not hold the events.

budget. On gemma-3-12b the composed arm reaches 31.5% on the conjunctive family against a blind floor of 17.2%; the control that dereferences a different occurrence of the same event at the same cost lands at 17.2%, +14.3 below and level with blind, and the control that speaks the address and then picks the frame by similarity at 13.6%, +17.8 below. On the count family the oracle and its wrong-address control are level at 73.0 and 74.0%, because there the ledger's count is the answer and the frame is irrelevant, and on the description family the oracle buys nothing. The same loop over the ledger AXIS wrote for itself lands at 18.8%.

So the read side executes addresses correctly and pays only where the family needs the address; the record it composes over does not hold the events; and the whole of the end-to-end loss is located in the write, by an experiment that varies only the write's output.

4.4 WHAT THE PRODUCT LAW SAYS AFTER THE MEASUREMENT

Both walls as we first stated them were about the *record*: the window it was written from, the words it stores per class. Both binding forms are about the *detector*. The identity wall says w=1 cannot

carry direction and $w{=}2$ can; the measurement says a backbone handed $w{=}2$ reads it only when the two frames are close enough for the change to be a motion rather than a new scene, so the identity factor is bounded by the detector's per-pair direction sensitivity at the spacing δ, not by the window's length. The ordinal wall says $O(1)$ storage cannot address interior occurrences and a witness list can; the measurement says the list is complete and the address still fails, because indexing by count compounds per-occurrence detector error. In both cases the architecture supplied the resource the wall named and the factor did not move, because the factor was never the resource: it was what the backbone could extract given it. That is a positive result about where the next paper must go: into the detector's reliability on direction and on occurrence, rather than into the memory.

5 LIMITATIONS

(i) One corpus family, three backbones: kitchen manipulation from a head camera plus one public benchmark, three frozen open-weight models, one budget. **(ii) One block of the write-rule ablation was not run**: the `dup` and `rev` controls outside gemma-3-12b, the registration's declared coverage cap, stated in Table 1's caption; a partial-cache number carries a superscript ^{p} and yields no verdict, and no verdict here depends on a cell that was not run. **(iii) The spacing result is one backbone's**: the zero-spacing collapse replicates on two, the above-prior direction read does not. **(iv) The oracle ledger is an oracle. (v) The registered predictions about the write side and the end-to-end result failed**, and we report them rather than re-registering. Appendix F states each at length.

6 RELATED WORK

Memory for long-horizon multimodal agents differs on the read side and agrees on the write. Hierarchical caption summaries (Song et al., 2024; He et al., 2024), agentic frame selection (Wang et al., 2024; 2025), temporal indices (Tian et al., 2026), interaction records (Das & Roy, 2026), ordinal anchors (Li et al., 2026) and operator algebras over versioned writes (Wang, 2026) are all address algebras over a write that emits a description of one frame; Definition 1 makes that write a parameter. Work that does move the write (Guo et al., 2026; Ge et al., 2026; Gong et al., 2026) moves the rate, the budget and the span, and Corollary 3 shows a triggered write with a frame-local emission is still $w{=}1$. The walls come from Chen et al. (2026) by way of Anonymous authors (2026f), whose single-frame write is measured here; the conjunctive corpus is a companion's (Anonymous authors, 2026a) and the arbitration cost of composing channels by concatenation another's (Anonymous authors, 2026b). Appendix A is the full treatment.

7 CONCLUSION

A memory architecture is a pair, and the field has been designing one half of it. Formalising the other half (the write window) gives a symmetry argument rather than a benchmark result: a frame-local emission is invariant under time reversal, so no reversal-paired verb is writable at any state budget, and that wall's statement contains none of the axes the field scales. The repair the proof names (an ordered pair, differenced into events) is implemented here and then *refused by its own registered controls* at one action apart; what escapes the majority prior is the pair's *spacing*, on one backbone, corroborated on a public benchmark with the read held fixed. Behind the window the ordinal wall binds in its error-propagation form, which no read-side machinery touches, and the oracle ledger settles where the loss lives. The product law told us to ask which factor is smallest; the answer is the detector's, at a spacing it can actually read.

REPRODUCIBILITY STATEMENT

Every number in this paper is written by a named analyzer to a result file, and every deciding read-out rule (the class-canonical scorer, the status-transition table, the registered thresholds and falsifiers, the coverage cap and the unparsed-cell reporting rule) was frozen before the run it decides; Appendix K is the registration and its amendments, reproduced verbatim including the predictions that failed. Appendix G proves every statement of §2, Appendix H gives the status-transition table and Appendix I the prompts and the address grammar verbatim. Appendix J states which inputs are supplied by companion work and which are this paper's. The corpus is generated from the

public EPIC-KITCHENS-100 annotations (Damen et al., 2022) under CC BY-NC 4.0 and the public benchmark arm from TempCompass; the write caches, the ledgers, the per-question records and the analyzers are in the supplementary material, and the scaffold-side difference that derives events from statuses is deterministic and costs no API call.

ETHICS STATEMENT

No human subjects were involved and no annotation was collected: all items are generated mechanically from existing public datasets' released annotations, used under their licences and for the research purpose they were released for. The backbones are frozen open-weight models called through a commercial router; no model was trained or fine-tuned, and they are the *object* of measurement. The paper's conclusion (that the unit written to memory bounds what the memory can hold, and that the binding constraint is a detector reliability rather than a storage budget) is a limitation statement about agent architecture and carries no group-differential harm we are aware of. We have read the ICLR Code of Ethics and adhere to it.

AI USE STATEMENT

We report our use of generative AI tools against the categories of the ICLR 2027 AI Policy for Authors. *Tasks in the required-disclosure list: none.* Generative AI was not used to develop the theory, to state or prove the walls, to design the architecture, the write rules, the controls or the oracle-ledger experiment, to set any registered threshold or falsifier, or to interpret the results. Frozen open-weight models are the *subject* of every experiment reported here, not tools used to produce them; scoring is by a class-canonical function the authors wrote against the corpus's published class tables, with no model judge anywhere in the pipeline, and every number is computed by an analyzer the authors wrote from the caches and per-question records the runs produced. *Tasks in the recommended-disclosure list.* Generative AI was used on the manuscript: it condensed a longer version into the present conference format, which involved drafting and rewriting substantial parts of the exposition; it edited for readability, formatted tables and references, and assembled the anonymised supplementary archive. It was not used to discover the research topic, identify the gap, select the literature, or decide what the paper claims; in particular, the decision to report the registered controls as refuting our own write repair is the authors'. *Verification.* Every quantitative statement was checked against the result file it cites, and every registered outcome against the criterion frozen before the run. We have reviewed all AI-assisted work and take responsibility for the final content of this paper, including its text, claims and artifacts.

REFERENCES

Anonymous authors. Where two channels can pay: The conjunctive fraction of embodied multimodal benchmarks, and why composition is arithmetically forbidden from showing itself on the rest. *Manuscript under review*, 2026a. Concurrent anonymous submission; author list withheld under double-blind review. No claim in the present paper depends on it: see the paragraph "Nothing here rests on an unpublished companion".

Anonymous authors. The arbitration wall: Reliability-blindness is measurable, and it measures blind. *Manuscript under review*, 2026b. Concurrent anonymous submission; author list withheld under double-blind review. No claim in the present paper depends on it: see the paragraph "Nothing here rests on an unpublished companion".

Anonymous authors. Two counts over annotations: What a multimodal benchmark can see about its own channels, measured before anything is run, and the refinement that makes the number meaningless. *Manuscript under review*, 2026c. Concurrent anonymous submission; author list withheld under double-blind review. No claim in the present paper depends on it: see the paragraph "Nothing here rests on an unpublished companion".

Anonymous authors. The price of belief: What a perceptual channel yields, what it costs to check, and why checking costs what looking costs. *Manuscript under review*, 2026d. Concurrent anonymous submission; author list withheld under double-blind review. No claim in the present paper depends on it: see the paragraph "Nothing here rests on an unpublished companion".

Anonymous authors. ArrowGEV: Grounding events in video via learning the arrow of time. In *Findings of ACL*, 2026e.

Anonymous authors. Budget-constrained embodied perception: Four resource walls, and a pre-registered evaluation of access-structured perception on open models at 31b or below. *Manuscript under review*, 2026f. Concurrent anonymous submission; author list withheld under double-blind review. No claim in the present paper depends on it: see the paragraph "Nothing here rests on an unpublished companion".

Kirolos Ataallah et al. Goldfish: Vision-language understanding of arbitrarily long videos. In *ECCV*, 2024.

Wenhui Chen, Jianlin Chen, Ziyao Lin, and Chi Man Vong. The capability convergence hypothesis: Capability from access structure, not scale. *arXiv preprint arXiv:2607.14144*, 2026. Public preprint, cited in the third person; the framework it states is restated in full wherever this paper relies on it.

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. Mem0: Building production-ready AI agents with scalable long-term memory. *arXiv preprint arXiv:2504.19413*, 2025.

Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Antonino Furnari, Evangelos Kazakos, Jian Ma, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, and Michael Wray. Rescaling egocentric vision: Collection, pipeline and challenges for epic-kitchens-100. *International Journal of Computer Vision*, 130(1):33–55, 2022. doi: 10.1007/s11263-021-01531-2.

Kousik Das and Debaditya Roy. Imprint: Online memory compression for long-horizon egocentric QA. *arXiv preprint arXiv:2607.00696*, 2026.

Chaoyou Fu et al. Video-MME: The first-ever comprehensive evaluation benchmark of multi-modal LLMs in video analysis. In *CVPR*, 2025.

Haonan Ge, Yiwei Wang, Hang Wu, and Yujun Cai. What should a streaming video model remember? *arXiv preprint arXiv:2606.16353*, 2026.

Sitong Gong, Caixin Kang, Tianyu Yan, Guo Chen, Bo Zheng, Kaipeng Zhang, Yunzhi Zhuge, Xiang Ruan, Huchuan Lu, and Yifei Huang. GROVE: Growing and reasoning over temporally stratified memory from streaming video experience. *arXiv preprint arXiv:2608.02392*, 2026.

Qiao Gu et al. ConceptGraphs: Open-vocabulary 3D scene graphs for perception and planning. In *ICRA*, 2024.

Zhenghui Guo, Yuanbin Man, Junyuan Sheng, Bowen Lin, Ahmed Ahmed, Bo Jiang, Boyuan Zhang, Miao Yin, Sian Jin, Omprakash Gnawal, and Chengming Zhang. Event-vstream: Event-driven real-time understanding for long video streams. *arXiv preprint arXiv:2601.15655*, 2026.

Yu Han, Wenhao Li, Yichao Cao, Hongyan Xu, Shuo Yang, Shan You, and Xiu Su. StateTrace: An object-centric framework for hidden-state spatiotemporal reasoning in long videos. *arXiv preprint arXiv:2608.18532*, 2026.

Bo He et al. MA-LMM: Memory-augmented large multimodal model for long-term video understanding. In *CVPR*, 2024.

Minyoung Huh, Brian Cheung, Tongzhou Wang, and Phillip Isola. The platonic representation hypothesis. In *ICML*, 2024.

Jongseo Lee, Hyuntak Lee, Sunghun Kim, Sooa Kim, Jihoon Chung, and Jinwoo Choi. Which way did it move? Diagnosing and overcoming directional motion blindness in video-llms. *arXiv preprint arXiv:2605.22823*, 2026.

Aiden Yiliu Li, Nels Numan, and Anthony Steed. Visual agentic memory: Enabling online long video understanding via online indexing, hierarchical memory, and agentic retrieval. *arXiv preprint arXiv:2605.16481*, 2026.

Yuanxin Liu, Shicheng Li, Yi Liu, et al. TempCompass: Do video LLMs really understand videos? In *Findings of ACL*, 2024.

Karttikeya Mangalam, Raiymbek Akshulakov, and Jitendra Malik. EgoSchema: A diagnostic benchmark for very long-form video language understanding. In *NeurIPS*, 2023.

Ishan Misra, C. Lawrence Zitnick, and Martial Hebert. Shuffle and learn: Unsupervised learning using temporal order verification. In *ECCV*, 2016.

Charles Packer, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G. Patil, Ion Stoica, and Joseph E. Gonzalez. MemGPT: Towards LLMs as operating systems. *arXiv preprint arXiv:2310.08560*, 2023.

Krishan Rana et al. SayPlan: Grounding large language models using 3D scene graphs for scalable robot task planning. In *CoRL*, 2023.

Enxin Song et al. MovieChat: From dense token to sparse memory for long video understanding. In *CVPR*, 2024.

Hongyao Tang. Towards a formal definition of agent memory: Basis, span, optimality, and the sequential memory problem. *arXiv preprint arXiv:2608.11654*, 2026.

Anxin Tian, Yiming Li, Xing Li, Hui-Ling Zhen, Lei Chen, Xianzhi Yu, Zhenhua Dong, and Mingxuan Yuan. Swiftmem: Fast agentic memory via query-aware indexing. *arXiv preprint arXiv:2601.08160*, 2026.

Xiaohan Wang, Yuhui Zhang, Orr Zohar, and Serena Yeung-Levy. VideoAgent: Long-form video understanding with large language model as agent. In *ECCV*, 2024.

Ziming Wang. TOKI: A bitemporal operator algebra for contradiction resolution in LLM-agent persistent memory. *arXiv preprint arXiv:2606.06240*, 2026.

Ziyang Wang et al. VideoTree: Adaptive tree-based video representation for LLM reasoning on long videos. In *CVPR*, 2025.

Donglai Wei, Joseph J. Lim, Andrew Zisserman, and William T. Freeman. Learning and using the arrow of time. In *CVPR*, pp. 8052–8060, 2018.

Di Wu, Hongwei Wang, Wenhao Yu, Yuwei Zhang, Kai-Wei Chang, and Dong Yu. LongMemEval: Benchmarking chat assistants on long-term interactive memory. *arXiv preprint arXiv:2410.10813*, 2024.

Zihui Xue, Mi Luo, and Kristen Grauman. Seeing the arrow of time in large multimodal models. In *NeurIPS*, 2025.

Zihao Yu, Xiu Yuan, and Chongjie Zhang. MEMORA: Embodied action memory from egocentric videos for reasoning and planning. *arXiv preprint arXiv:2607.14252*, 2026.

Ce Zhang et al. A simple LLM framework for long-range video question-answering. In *EMNLP*, 2024a.

Haoji Zhang et al. Flash-VStream: Memory-based real-time understanding for long video streams. *arXiv preprint arXiv:2406.08085*, 2024b.

Le Zhang et al. EgoCITE: Context-augmented indexing and time-aware retrieval for long-horizon egocentric memory. *arXiv preprint arXiv:2608.12627*, 2026.

Mingxi Zou, Zhihan Guo, Langzhang Liang, Zhuo Wang, Qifan Wang, Qingsong Wen, Irwin King, Lizhen Qu, and Zenglin Xu. Remember the decision, not the description: A rate-distortion framework for agent memory. *arXiv preprint arXiv:2605.10870*, 2026.

A Related Work, in Full

Read-side memory for long video. MovieChat (Song et al., 2024) and MA-LMM (He et al., 2024) compress frame features into fixed-size banks consolidated as the stream arrives; VideoAgent (Wang et al., 2024) and VideoTree (Wang et al., 2025) select frames agentically under a query; LLoVi (Zhang et al., 2024a) and Goldfish (Ataallah et al., 2024) caption clips and reason over the captions, the second through a retrieval stage; Flash-VStream (Zhang et al., 2024b) streams frames into a learned memory read asynchronously. Evaluated on EgoSchema (Mangalam et al., 2023) and Video-MME (Fu et al., 2025), they differ in every read-side decision (what to keep, how to compress, when to re-read) and agree on the unit written, a description or embedding of one frame treated as a point. None states $w{=}1$ as an assumption, so none can be said to have tested it; each assumes a record of single frames retains the verb. Theorem 1 says it cannot for the reversal pairs, and §4.1 measures how often the verb is gone at the most favourable frame before any read-side machinery runs.

Structured, temporal and streaming memory, and where each sits against the walls. SwiftMem (Tian et al., 2026) builds temporal indices over sorted timelines; TOKI (Wang, 2026) gives versioned writes an operator algebra with bitemporal semantics; Imprint (Das & Roy, 2026) consolidates interaction records $(p_i, a_i, o_i, t_i^s, t_i^e)$ into event prototypes retaining a recurrence count ν_n and a last-seen time T_n^{last}, stating that the streaming design "eliminates the need to retain the complete set of extracted interaction records"; Visual Agentic Memory (Li et al., 2026) keeps retained moments with an `occurrence_index` that "allows ordinal reference to repeated anchor events". These are the nearest neighbours and §2 locates each. Imprint's compression is exactly the $O(1)$-per-class policy of Proposition 3, and its benchmark's templates are closed under $\{\text{count}, \text{last}, \text{most}, \text{usually}\}$, so the wall is never queried. Visual Agentic Memory builds the ordinal factor over keys whose identity was decided at $w{=}1$, which §4.1 measures wrong for exactly the reversal-paired verbs: an exact index over inexact keys. Its "transitions" are boundary detection and redundancy filtering, spending the change signal on *which frames to keep*; AXIS spends it on *which frames to pair*. All four assume that what reaches the index is a correct event; Proposition 3 is about what an exact index does when it is not, and §E.3 measures the rate.

Event-triggered, selective and stratified streaming memory. The write side has since moved, and in three directions, none of which is the window. Event-VStream (Guo et al., 2026) "detects meaningful state transitions by integrating motion, semantic, and predictive cues, and triggers language generation only at those boundaries," consolidating each event embedding into a persistent bank; SelectStream (Ge et al., 2026) formulates memory as "budgeted online latent evidence allocation," keeping the current observation visible and exposing history through a query-conditioned budget with surprise-driven windowing; GROVE (Gong et al., 2026) grows one causal memory that "retains fine-grained perceptual evidence and incrementally consolidates it into time-stamped moments, coherent episodes, and recurring cross-day patterns." Each is a real advance on the resource it targets, and we do not dispute any of their reported numbers. What each assumes is different in each case, and in each case the assumption is about which resource binds.

Event-VStream assumes that detecting the boundary of a state transition captures the transition. Corollary 3 says it does not: all three cue families it names are reversal-invariant (two symmetric distances and a residual whose filter is palindromic), so the boundary set is a reversal-invariant object even at the oracle event rate, and §E.2 measures the direction it carries on our streams at no more than $+2.5$ points against the majority prior. The trigger changes when the write fires, not what the emission is a function of. SelectStream states the assumption outright: "the key challenge is not whether to use memory, but how to allocate it selectively." That is a claim that the binding factor is F_{pc}, and under Proposition 1 an allocation policy is bounded above by F_{id} whatever its budget. Its own reported evidence admits a second reading on which our measurement bears: that strong recent-window baselines beat history injection is what one should expect if the history records do not retain the discriminating variable, which is the identity factor and not an allocation failure, and the reversal pairs separate the two. GROVE assumes that consolidating upward adds capability. Each stratum is a function of the stratum below it, so the induction of Theorem 1 carries: if the base is fixed by R, every stratum above it is, and no amount of consolidation returns a bit the emission did not retain. Consistently, its own ablation reports the largest gain from patterns "when evidence spans multiple days", a gain on temporal *span*.

Span, read budget, write rate and selectivity are four axes, and these systems move them well. The window (how many ordered observations one record is a function of) is a fifth, it is inherited from image captioning rather than chosen, and it is the only one that appears in the statement of Theorem 1.

Scene graphs and embodied state. SayPlan (Rana et al., 2023) and ConceptGraphs (Gu et al., 2024) maintain object-centric world state for planning: nodes carry an object, its place and open-vocabulary object nodes and spatial edges, from which a status must be inferred rather than read, and the planner queries the graph rather than the frames. Their state fields are the closest thing in the literature to the `affordances` field of W1′, and that a sequence of such graphs implies an event stream is the working assumption of every planner that re-plans when the graph changes. Two things separate them from a memory: the graph is a present tense, overwritten, so no earlier configuration survives to be addressed; and it is written from single observations, so what it holds about an action is what one frame decides. The line assumes that state suffices and history need not be kept; W2 keeps the history, and §E.3 measures what a status a vision–language model tracks from frame to frame is worth as a source of events. Recent egocentric systems carry the object- and state-centric line into memory proper: StateTrace (Han et al., 2026) organises object trajectories and state-transition events into a reusable substrate for long-video QA; EgoCITE (Zhang et al., 2026) builds context-augmented, time-aware indices over egocentric captions and speech; MEMORA (Yu et al., 2026) consolidates egocentric experience into participant-specific entity, activity and knowledge stores. Each writes states or events a model decided; none parameterises the window the decision is a function of, the variable Theorem 1 fixes and Proposition 3 prices.

Temporal sensitivity as a benchmark property. TempCompass (Liu et al., 2024) constructs conflicting clips (a video and its time reverse, its speed change, its concatenation in the other order) so that a single-frame or bag-of-frames model has nothing to answer with. Its direction dimension is the identity wall's premise verbatim, which is why the window is tested on it here (§E.4); its clip-and-reverse pairs also supply a manipulation check nobody designed as one, since a write with $w{\geq}2$ *may* produce opposite motion labels on the two members of a pair and a write with $w{=}1$ *must* produce the same. EgoSchema (Mangalam et al., 2023) rules out short certificates and Video-MME (Fu et al., 2025) spans durations to the hour; the companion measurement (Anonymous authors, 2026a) finds their conjunctive fraction low, so a memory whose gain lives in the identity factor should show it there only through the verbs it gets right. The line assumes the failure it measures belongs to the model; §E.4 separates the part that belongs to the write from the part that belongs to the detector, on the same clips.

Single-frame bias and the time-reversal ambiguity of actions. That a still frame under-determines the action is the oldest observation in action recognition, and a memory that captions frames inherits the bias whatever the captioner. The perception-side literature around the observation is the arrow of time: temporal order is usable as a self-supervised signal on frames (Misra et al., 2016), the arrow is recoverable from frames and trainable as a feature (Wei et al., 2018), and the multimodal-model line measures what current models do with it. AoTBench (Xue et al., 2025) finds large multimodal models largely insensitive to temporal direction, trains awareness in with a reverse reward, and reports direction as the dimension a bag of frames answers worst; ArrowGEV (Anonymous authors, 2026e) splits events into time-sensitive and time-insensitive (the taxonomy the reversal pairs instantiate) and learns direction discrimination by reinforcement; DeltaDirect (Lee et al., 2026) traces signed motion direction through a Video-LLM, finds it linearly decodable at every stage and still not verbalised, names the failure a binding gap at the readout, and repairs it with an objective over adjacent-frame feature deltas. On egocentric video (Damen et al., 2022) the hand–object configurations for a verb and its reverse are one set of frames traversed in opposite directions, so the confusion is paired rather than diffuse, and §4.1 finds the error mass on the pairs the theorem names. These works concern what a model can perceive or be trained to perceive; Theorem 1 is a statement about what any memory write can decide, and it relocates the failure from perception into the write path of the architecture: at $w{=}1$ it is a closure property of the write that no training of the write and no state budget lifts, while the detector-side repairs the perception line proposes are exactly what §E.3 says the memory architecture must import. Corollary 2 bounds the temporal order a record can hold by the window it was written from.

Agent memory systems in text. The text-side literature fixed the vocabulary this paper borrows. MemGPT (Packer et al., 2023) treats context as a tiered store paged between a working context and an archive; Mem0 (Chhikara et al., 2025) extracts and consolidates facts from conversation under add, update and delete operations; LongMemEval (Wu et al., 2024) benchmarks such stores over long interactive histories and reports temporal reasoning and knowledge updates as the hardest abilities; SwiftMem and TOKI are the indexing and versioning ends of the same line. All write text a person or a model produced, and in text the unit written carries its own verb: "I moved to Lisbon" does not become "I left Lisbon" when the transcript is read backwards. The identity wall is invisible in text, which is why ordinal and temporal operators could be built there without meeting it. Carried to a perceptual stream, the line assumes the record is as truthful as a transcript; §4.1 measures the identity factor of a record written from frames, and it is not.

Formal accounts of memory. Two concurrent accounts formalise what a memory is and what a write should optimise, and both take the event vocabulary as given. Tang (Tang, 2026) defines memory as a basis and knowledge as its span, makes answerability a coverage problem, and poses the optimal memory as a capacity-constrained maximiser of expected coverage under a write policy learned as an MDP; DeMem (Zou et al., 2026) casts the write as decision-centric rate-distortion, derives an exact forgetting boundary for what can be safely dropped, and proves near-minimax regret for the induced partition. Neither asks whether the vocabulary is writable from the observation window, which the identity wall answers, and neither bounds what an executed address resolves over a noisy detector, which Proposition 3 does. Under Proposition 1 the two questions compose rather than compete: coverage and distortion are properties of the read over a record whose identity and ordinal factors the write has already set, so a memory optimal in either account can still be empty of exactly the events a conjunctive locator names.

Address-style access against similarity access. Retrieval-augmented systems, in video (Zhang et al., 2024a; Ataallah et al., 2024; Wang et al., 2024) and in text memory (Packer et al., 2023; Chhikara et al., 2025), resolve a query to a record by similarity between an embedding of the question and embeddings of the stored descriptions. That is the right operator when the answer is *described* somewhere in the store, and the wrong one for a conjunctive query, whose locator names a position by a reduction (the k-th occurrence, the first time a status held) that no single record describes, because the matching record was written before the count was known. The ordinal operators of Li et al. (2026) and the count-and-last statistics of Das & Roy (2026) are address-style access, and §C.3 adopts and extends them. Under similarity the parent system's two channels are joined by the backbone in text ($\oplus_{\mathrm{cat}}$); under an executed address they are joined by the scaffold ($\oplus_{\mathrm{addr}}$, Definition 8). The address line assumes an executed address is worth more than a spoken one; §4.2 prices that with a control that speaks the address and one that executes it to the wrong occurrence, and finds the two indistinguishable when the address is rarely right.

Compounding error over ordinal chains. That a per-step error compounds over a chain is standard, and the parent system's union bound on a maintained count (Anonymous authors, 2026f) is its form for a running reduction: a state updated N times with per-update error ϵ_u fails as $N\epsilon_u$ approaches one, measured there as a scale threshold. The ordinal form differs in what compounds. A witness list is not updated in place, so it does not compound across the stream; it compounds across the *address*, because $\mathrm{nth}(e, k)$ is right only if the k-th occurrence and every earlier one was detected, so resolution falls as ρ^k (Proposition 3) whatever the stream's length. The count-based memories above execute addresses of this kind and none states a per-occurrence detection rate, because their events are supplied by an annotator or a detector treated as exact: they assume ρ near one. §E.3 measures it near 0.2–0.35, fits the per-k curve, and shows that neither a wider dereference nor any filter over the ledger escapes the bound, as the proposition says none can.

Controls and manipulation checks in architecture papers. The convention is an ablation: remove a component and report the drop. An ablation attributes a gain to a component's *presence*, not to the mechanism the paper names, and the two come apart whenever a component changes several things at once. Attaching a second frame changes the pixels, the prompt's length, the model's attention to the instruction and, in our prompt, the enumerated vocabulary. The controls of §D separate these: `dup` attaches the same frame twice, so it has the pair's cost and none of its temporal information; `rev` attaches the two frames in the wrong order, so it has the pair's pixels and the opposite direction. On

the read side `axis_noderef` and `axis_wrongaddr` break the address in two places at identical cost. None of the systems above reports a control of this form, and each assumes its ablation drop is attributable to the mechanism its design names. §4.1 reports what the controls said here, which was not what the registration predicted, and keeps the finding because the controls were registered for exactly that outcome.

What a matched budget means, and who else matches it. The Capability Convergence Hypothesis (Chen et al., 2026) makes the per-decision token budget the unit of comparison, and the parent system (Anonymous authors, 2026f) reports every arm, including its reimplementation of a published agentic long-video system, at one fixed budget. Most of the long-video literature does not: MovieChat, MA-LMM and Flash-VStream fix a memory size in tokens or slots and let the reader spend what it needs; VideoAgent and VideoTree report frames consulted, a cost in a different currency; caption pipelines report accuracy and sometimes the caption budget. A comparison at unequal budgets cannot separate a better access structure from a larger one, which is the confound the hypothesis exists to remove. Every arm on every benchmark here decides under one budget, each control is level with the arm it controls, and realised spend is reported per arm; P-cost is scored as a prediction and fails as stated (§4.2). The unmatched literature assumes a gain survives equalisation; the one registered prediction here about cost is what that assumption costs when it is checked.

Convergence hypotheses, and the write they presuppose. The representational-convergence claim of Huh et al. (2024) holds that models trained on different data and modalities converge in what they represent; Chen et al. (2026) reply that representational convergence does not entail capability convergence, and that under a fixed budget capability converges toward architectures holding both a compressive state and a verbatim index. Their three walls (Shannon, horizon, circuit) bound the read of a record assumed to contain what was seen. Both hypotheses take the representation of a single observation as the object at issue. Proposition 1 puts the read-side walls in the third factor of a product whose first two are properties of the write, and Theorem 1 bounds the first by a symmetry that no representation of one frame breaks, however converged. Both assume the record holds the verb; where the identity factor is small, as §4.1 measures it, the read-side walls are loose and the convergence argument concerns the read side of a write that never held it.

What the companion measurements supply, and what they leave to this paper. Four measurements from the companion papers (Anonymous authors, 2026a;b;c;d) are inputs to this paper; none was run to confirm anything in it. The conjunctive measurement (Anonymous authors, 2026a) shows the additive form of super-additivity is arithmetically unavailable on two of three query families, measures the conjunctive fraction of the field's benchmarks at 0.51% pooled, builds the corpus on which §4.2 is scored, and locates the single-frame write path upstream of every read arm, which is where this paper's theory begins. The arbitration measurement (Anonymous authors, 2026b) shows a reliability-blind arbiter attains the base rate of the channel it follows and that the access loop's exit is where a channel's harm is paid; the unscorable cells of §4.2 are that exit on this corpus. The annotation-side audit (Anonymous authors, 2026c) bounds what a corpus permits a blind strategy to score, the floor every conjunctive cell is read against. The verification measurement (Anonymous authors, 2026d) prices what it costs to check a perceptual claim on a metered interface, which is why the ledger is trusted rather than verified and the carry-forward cost of Limitation (8) is carried rather than checked. Every one bounds a read, or the corpus a read is scored on; each assumes, and this paper supplies, the write-side term that decides whether the record any of them reads held the event at all.

Relation to the concurrent submissions, stated rather than left to be discovered. The public preprint (Chen et al., 2026) states the hypothesis this paper's theory builds on: capability is bought by access structure under a fixed per-token budget. Of the companion submissions (Anonymous authors, 2026a;b;c;d;f), this paper uses two things: the conjunctive corpus with its keyframes and single-frame caches (Anonymous authors, 2026a), and the equal-budget baseline arms of the budgeted access-structure grid (Anonymous authors, 2026f). It re-scores none of their claims; its identity-factor measurement, though run on those caches, is a quantity neither reports. Conversely it claims no benchmark gain, no verification price, no arbitration result and no annotation-side bound; each belongs to the submission that measured it and is cited, not derived. The one number carried forward as a threshold rather than as evidence is the conjunctive availability 42.3% (Anonymous

authors, 2026a), against which that submission's own headline arm fails; what the corpus and the two thresholds are, and how each is reproduced from released material, is specified self-containedly in §3 and Appendix J.

B THEORY, IN FULL

The section states what an architecture must *be* for the availability the companion measurement certifies (Anonymous authors, 2026a) to be used, and it does so in the order the measurement later reads it: first the object (a write and an algebra over what the write leaves behind), then the product that says which of three factors binds, then the two walls, each dual to a factor, then the composition that multiplies the factors by construction rather than by a backbone reading two blocks of prose. Every statement is elementary. Proofs are in Appendix G; each formal statement is followed by a paragraph saying what it says in plain terms.

B.1 STANDING CONVENTIONS

The stream and the record. A *stream* is a finite sequence of keyframes $x_{1:N} = (x_1, \dots, x_N)$ over an observation alphabet $\mathcal{X}$; positions are the integers $[N]$. A stream is processed once, online: at position t a *write map* $\mathcal{W}$ receives the current record s_{t-1} and the *observation window* $x_{t-w+1:t}$ of the last w keyframes and returns the next record $s_t = \mathcal{W}(s_{t-1}, x_{t-w+1:t})$, with s_0 fixed. The integer $w \geq 1$ is the *write window*; the record s_N is all the structure retains once the stream has passed. Nothing bounds the size of a record anywhere in this section: the walls below hold at any state budget, and the statements say so where it matters.

The address algebra. An *address algebra* $\mathcal{A}$ is a set of partial operations from records to positions, $\alpha \colon \mathcal{S} \rightharpoonup [N]$, executable by the scaffold without re-reading the stream. The operations used here are $\mathrm{nth}(e, k)$, $\mathrm{first}(e)$, $\mathrm{last}(e)$, $\mathrm{count}(e)$, an offset $\pm d$, and the state addresses $\mathrm{first}(obj{=}status)$, $\mathrm{last}(obj{=}status)$ (§C.3). An operation is *undefined* on a record that does not carry what it needs, and an undefined address resolves to no position.

The budget and the read. On a query q the read side has B tokens. It may execute addresses over the record, dereference keyframes at resolved positions, and answer; every token it spends is charged to B. The *blind floor* $\varepsilon_{\mathrm{guess}}$ is the accuracy of a read that sees neither record nor stream, the `blind` arm of §D. All probabilities below are over the draw of a query from a fixed family and any internal randomness of the backbone; nothing is assumed independent unless a statement says so.

Events and predicates. An *event class* e is a predicate on positions of a stream, $e_x(t) \in \{0, 1\}$, read "e occurred at t". Its occurrences in $x_{1:N}$ are $t_1 < \cdots < t_m$, and $\mathrm{nth}(e, k)$ is t_k when $k \leq m$, undefined otherwise.

Time reversal. R reverses a stream, $(Rx)_t = x_{N+1-t}$. The *reversal conjugate* of a predicate e is the predicate e^R with $e^R_x(t) \coloneqq e_{Rx}(N + 1 - t)$: e^R holds at t in the stream exactly when e holds at the mirror position of the reversed stream. Played backwards, a *put* is a *take*, an *open* is a *close*, a *fill* is an *empty*; each is the other's conjugate, and a pair with $e^R \neq e$ is a *reversal pair*.

Definition 3 (The three chain events and the three factors)**.** Fix a conjunctive query whose locator names the k-th occurrence of e (Definition 2). Let E_{id} be the event that the record *decides* e, in the sense of carrying a token from which e's occurrences are recognised as occurrences of e; let E_{ord} be the event that $\mathrm{nth}(e, k)$ over the record resolves to the true k-th position; and let E_{pc} be the event that the percept read at the resolved position is correct. The three *factors* are the conditional probabilities

$$F_{\mathrm{id}} \coloneqq P[E_{\mathrm{id}}], \qquad F_{\mathrm{ord}} \coloneqq P[E_{\mathrm{ord}} \mid E_{\mathrm{id}}], \qquad F_{\mathrm{pc}} \coloneqq P[E_{\mathrm{pc}} \mid E_{\mathrm{id}} \cap E_{\mathrm{ord}}],$$

and the *chain* is $C \coloneqq E_{\mathrm{id}} \cap E_{\mathrm{ord}}$, with $P[C] = F_{\mathrm{id}}\, F_{\mathrm{ord}}$.

Definition 3 is the paper's coordinate system. Each factor is a probability *conditional on the one before it*, so each can be measured on the sub-population the previous one admits (which is how Table 4 is read from left to right), and the product of the three is the probability that the whole chain holds and the answer at its end is right. The identity factor is a property of the write; the ordinal

Table 4: Address resolution on the conjunctive family (245 items), canonical keys, three ledgers built from the same pair cache. *found*: the query's event class exists in the ledger; *cnt±1*: the ledger's count is within one of the gold count; *resolv*: the ledger holds at least k occurrences; *addr*: the k-th ledger occurrence is the gold locator frame (exact, within ±1, within ±3 frames). Right block: per-occurrence recall and precision of the ledger against the gold occurrence list of the queried class, within ±1 and ±3 frames. `action`: the model's own verb-noun field; `diff`: the W1′ differenced stream; `union`: `diff` with `action` filling classes it lacks. Single-frame write, for reference: found 19.6/21.2%, exact address 0.4–1.6%.

	ledger	found	cnt±1	resolv	addr=	addr±1	addr±3	rec±1	rec±3	prec±1	prec±3
gemma-3-12b	`action`	45.7	13.1	35.9	0.0	1.2	2.0	17.3	24.9	15.0	22.3
	`diff`	44.5	14.7	25.7	0.4	1.2	1.6	11.0	16.5	11.7	17.5
	`union`	52.7	16.3	31.8	0.4	1.6	2.4	14.8	22.2	12.3	18.3
qwen3.8-27b	`action`	70.2	22.4	50.6	4.1	5.7	8.6	36.1	43.1	26.3	33.7
	`diff`	62.9	17.1	49.8	2.4	4.9	7.8	32.2	39.7	23.9	32.0
	`union`	71.8	21.2	51.4	2.9	5.3	8.2	33.3	41.0	23.3	31.0

factor of the record and of what entered it; the percept factor of the read and its budget. The three walls of Chen et al. (2026) (Shannon, horizon, circuit) are statements about F_{pc}.

B.2 Write maps, address algebras, and the constraint between them

Figure 2 shows the access structure end to end: the pipeline from stream to answer and the two walls that sit on it; this section defines its parts.

Definition 4 (Access structure)**.** An *access structure* over a stream $x_{1:N}$ is a pair $(\mathcal{W}, \mathcal{A})$. The write map $\mathcal{W}$ takes the current record and an observation window $x_{t-w+1:t}$ of length w and returns the next record; it is applied online, once per keyframe, and the record is all the structure retains. The address algebra $\mathcal{A}$ is the set of operations the read side can execute over the record without re-reading the stream. A predicate e on stream positions is *w-writable* if some write map with window w decides it *at the position*: that is, if the emission the map makes at t (the part of its output that is a function of the window and not of the record it inherits) determines $e(t)$. We say "at the position" rather than "online" deliberately, and Definition 7 makes it exact, because a write that inherits a record can always postpone a decision by storing the window and deciding later; that manoeuvre is not excluded by the walls below, it is the escape they name, and W1′ is our implementation of it (Proposition 2). $\dim \mathcal{W}$ denotes the set of w-writable predicates of the write in use, and $\dim \mathcal{A}$ the set of predicates the algebra's operations are well defined over.

The pair is the object the field has been treating as a singleton. Every architecture in §6 is an $\mathcal{A}$ (an index, a compression policy, a retriever) built over a $\mathcal{W}$ that is never named because it is always the same: a description of one frame. The definition makes the write a parameter so that it can be varied, and the constraint that follows says why the two halves cannot be designed independently.

Definition 5 (Conjunctive query family)**.** A query is *conjunctive* if its answer has the form $a(q) = V\big(x_{L(x_{1:N})}\big)$: a locator L computes a position by a reduction over the stream ("the k-th time event e happened, d steps later") and a percept V is read at that position. The *ordinal* subfamily has $L = \mathrm{nth}(e, k) + d$.

The companion paper (Anonymous authors, 2026a) measures the incidence of this family in the field's embodied multimodal benchmarks at 0.51% pooled, shows it is the only family on which the additive super-additivity bound is arithmetically available, and builds a corpus that makes it measurable at scale with an availability threshold of 42.3% on the conjunctive subset. Everything below concerns what an architecture must *be* for that availability to be used. The family is the one where a locator and a percept are *different* reads over *different* resources (a reduction over the whole stream and a look at one frame), so it is the family where a two-channel structure has something to compose, and the family where a product of factors is the natural shape of capability.

Definition 6 (Witness)**.** A *witness* for the claim "e occurred at position t" is the position t itself, together with whatever the record retains from x_t that lets the read side dereference it. A record *retains the k-th witness* of e if $\mathrm{nth}(e, k)$ can be computed from the record alone.

A witness is what turns a count into an address. A record that knows e happened four times has a number; a record that knows *where* it happened four times has four positions, and the fourth of them can be dereferenced. The distinction is the whole of the storage form of the ordinal wall (Proposition 3).

Proposition 4 (Product law)**.** *For a conjunctive query whose locator names the k-th occurrence of event e,*

$$\mathrm{Cap} \;\le\; \underbrace{P[\,e \text{ decided}\,]}_{F_{\mathrm{id}}} \times \underbrace{P[\,k\text{-th witness resolved} \mid e \text{ decided}\,]}_{F_{\mathrm{ord}}} \times \underbrace{P[\,\text{percept read} \mid \text{position}\,]}_{F_{\mathrm{pc}}} + \varepsilon_{\mathrm{guess}},$$

where $\varepsilon_{\mathrm{guess}}$ is the blind floor. Moreover $\dim \mathcal{A} \le \dim \mathcal{W}$: an operation $\mathrm{nth}(e, k)$ is well defined only if e is $\mathcal{W}$-decidable and its witnesses survive storage.

The product law says two things. The first is arithmetic: a product is bounded by its smallest factor, so the question "which component of a memory architecture is binding" is the question "which factor is smallest", and each factor is measurable on its own sub-population. The second is structural: the algebra cannot be richer than the write. An address is an expression over tokens; if the write never emitted the token, the expression has nothing to run on, however elegant the algebra. The proposition's inequality is the paper's design constraint and its reading instrument at once.

Lemma 1 (Exact chain decomposition)**.** *With the events of Definition 3 and $C = E_{\mathrm{id}} \cap E_{\mathrm{ord}}$, the accuracy of any read that answers through the address is, identically,*

$$\mathrm{Cap} \;=\; F_{\mathrm{id}}\, F_{\mathrm{ord}}\, F_{\mathrm{pc}} \;+\; \big(1 - F_{\mathrm{id}} F_{\mathrm{ord}}\big)\, \varepsilon_{\mathrm{off}}, \qquad \varepsilon_{\mathrm{off}} \coloneqq P[\,\textit{correct} \mid C^{c}\,].$$

Proposition 1 is this identity under the convention that a read whose chain has broken is no better than blind, $\varepsilon_{\mathrm{off}} \le \varepsilon_{\mathrm{guess}}$; without that convention the second term is bounded only by $\varepsilon_{\mathrm{off}}$, which the `axis_wrongaddr` control of §C.3 measures directly.

The lemma separates what the product law *proves* from what it *assumes*. The identity is a partition of the outcome space and assumes nothing. The floor term is a convention: that dereferencing the wrong frame leaves the reader where the blind reader was. It need not be so (a wrong occurrence of the right event class can carry the right percept by coincidence, since "what happened after you opened the fridge" has a modal answer), and the paper does not assume it. It measures $\varepsilon_{\mathrm{off}}$ with a control that breaks the chain on purpose, and §4.2 reports the value.

Corollary 1 (The smallest factor bounds the gain over the floor)**.** *Under the convention of Lemma 1, $\mathrm{Cap} - \varepsilon_{\mathrm{guess}} \;\le\; \min\{F_{\mathrm{id}}, F_{\mathrm{ord}}, F_{\mathrm{pc}}\}$, and more precisely $\mathrm{Cap} - \varepsilon_{\mathrm{guess}} \le F_{\mathrm{id}} F_{\mathrm{ord}}\,(F_{\mathrm{pc}} - \varepsilon_{\mathrm{guess}})$: the gain over the floor is the chain probability times the percept headroom. A structure whose chain holds on a fraction $F_{\mathrm{id}} F_{\mathrm{ord}}$ of queries cannot exceed the floor by more than that fraction, whatever its read.*

This is the arithmetic the end-to-end cells of §4.2 are read against. Once the chain probability has been measured on the cache (Table 4), the corollary turns it into a ceiling on the executed-address arm's distance above the blind floor, before any end-to-end cell has run. §E.3 writes that ceiling down in advance and §4.2 records what landed under it.

The three factors are dual to three different resources, and the walls below are the duals made explicit. The read-side walls of Chen et al. (2026) (Shannon, horizon, circuit) all live in F_{pc} and presuppose the other two; they bound a read over a record that is assumed to contain what was seen. In the regime §4.1 measures for F_{id} they are loose, and the convergence conjecture built on them is an economic argument about the read side of a write that never held the verb.

B.3 The identity wall

Theorem 2 (Identity wall)**.** *Let R denote time reversal of the stream. The set of predicates decided* at the position *by a write map with $w{=}1$ (the 1-writable predicates of Definitions 1 and 7) is closed under R. Consequently, for any pair of events e, e^{R} that are reversal conjugates with reversal-symmetric single-frame marginals (*put/take*, *open/close*, *fill/empty*) and non-degenerate prior $0 < \pi < 1$, no $w{=}1$ emission decides e against e^{R} better than the prior,* at any state budget*; and for any query family whose locator is such an event and whose ordinal is drawn among the pair's occurrences with exchangeable interleaving, $\mathrm{Cap} \to \varepsilon_{\mathrm{guess}}$ as the occurrence count grows, independent of the index, the retriever and the read budget. An ordered window of length $w \ge 2$ is not R-invariant, so the closure breaks at $w{=}2$.*

What the theorem does not say. It does not say that a system whose write window is one frame can never record a verb. A write map that inherits a record may compare its emission at t against a status the record already holds and so decide *put* against *take*; and that comparison is a window of length two moved from the write into the algebra, which is exactly what W1′ does (§B.2, Proposition 2). What the theorem says is that the *emission* cannot carry the verb, so a pipeline whose record is a bag of emitted descriptions (the caption memories of §6, and the companion system this paper starts from (Anonymous authors, 2026f)) has nothing to compare and inherits the wall in full. The escape requires the emission to carry a differenceable quantity, which is the design decision of §3 and the reason W1′ asks for statuses rather than verbs.

The theorem's content is not that the write *happens* to be bad at verbs; it is that the field's three scaling axes are absent from the bound. A larger state, a wider retrieval, a longer read budget change nothing in the statement. What changes it is the window. The proof (Appendix G) is an induction on the record and a symmetry of one frame, and the next subsection isolates the symmetry so that its hypotheses can be read off and its escape named exactly.

Corollary 2 (Window hierarchy)**.** *Let $\mathcal{P}_w$ be the set of w-writable predicates. Then $\mathcal{P}_1 \subsetneq \mathcal{P}_2 \subseteq \mathcal{P}_3 \subseteq \cdots$, and $\mathcal{P}_w$ contains no predicate whose decision requires the temporal order of more than w observations. The write window bounds the temporal order of the predicates the memory can hold.*

The hierarchy is strict at its first step and the first step is where the verbs live. A verb is a change; a change is two frames in order; a predicate that needs two frames in order is outside $\mathcal{P}_1$ and inside $\mathcal{P}_2$. Beyond $w{=}2$ the hierarchy is not claimed strict, and the paper does not need it to be: everything the conjunctive family asks of the write is a first difference.

Corollary 3 (Triggered writes inherit the wall)**.** *Let a write path be a* segmentation $T(x) \subseteq \{1, \ldots, N\}$ *together with a frame-local emission at each selected position, and let T threshold a score s_t obtained by applying, to the frame-wise features $\phi(x_{t-m}), \ldots, \phi(x_t)$, a map ψ that is invariant under reversing its own argument list. Then $T(Rx)$ is the reflection of $T(x)$, the multiset of emitted records is unchanged, and Theorem 1 applies verbatim, at every event rate, including the oracle rate at which T fires exactly on the true transitions.*

The hypothesis on ψ is not a technicality inserted to make the corollary true; it is satisfied by each cue family the event-driven streaming literature names. A symmetric distance $d(\phi(x_t), \phi(x_{t-1}))$ is invariant because d is symmetric, which covers frame-difference and motion energy and the embedding distance that adaptive windowing calls *surprise*. A linear-extrapolation residual is invariant because its coefficient vector $(1, -2, 1)$ is palindromic, as is any finite-difference filter with palindromic coefficients. So a boundary detector may be made arbitrarily good at finding the moments at which something changed and remain exactly blind to which way it changed: *an event boundary is a reversal-invariant object, and the event is not.* What breaks the invariance is a ψ that is genuinely one-sided (a learned causal predictor, say), and then the direction it recovers is a claim to be measured rather than a property of being event-driven.

This separates two things that the phrase "event-driven write" holds together. Moving the trigger from a fixed interval to a detected boundary changes *when* the write fires, and with it the latency and the token bill, which is what that line of work reports and improves. It does not change the window the emission is a function of, and the wall is stated in the window. §E.2 measures both halves of this on the corpus of §4.1: the invariance as an identity, and what the boundary bit is nonetheless worth for direction once the world declines to be exactly reversal-symmetric.

B.4 The symmetry, made precise

Theorem 1 rests on one fact about one frame, and it is worth stating that fact on its own, because the paper's controls (`dup`, `rev`) are tests of exactly it and nothing else.

Definition 7 (Frame-local write; emission)**.** A write map with window w is *frame-local* if its dependence on the window factors through an *emission* $\tau_t = f(x_{t-w+1:t})$ that does not see the record: $\mathcal{W}(s, x_{t-w+1:t}) = U(s, f(x_{t-w+1:t}))$ for some update U. The write *decides e at the frame* if $e_x(t)$ is a function of τ_t alone. The standard caption write is frame-local with $w{=}1$; W1′ (§C.1) is frame-local with $w{=}2$ on its `change` and `action` fields, and copies rather than perceives the statuses it carries forward.

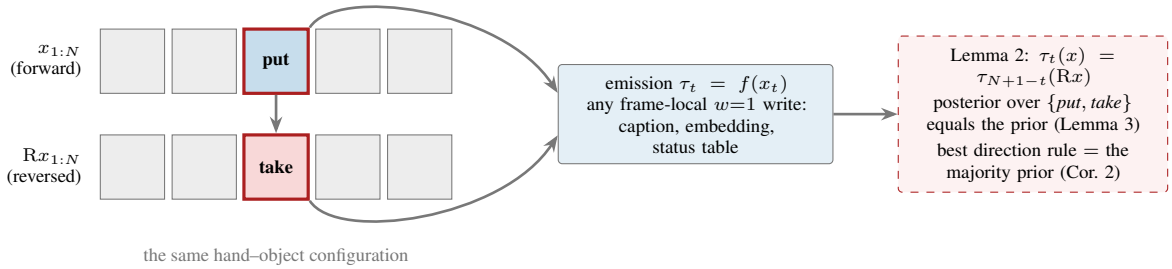


Figure 5: The identity wall's geometry. A put at the midpoint of the forward stream and the take at the mirrored position of the reversed stream present the same hand–object configuration; every frame-local $w{=}1$ emission returns the same value for both, so the write cannot separate the pair and the best direction rule guesses the prior. The escape is to widen the window: retain a variable the verb mutates (Proposition 2).

The definition names the "unit written to memory" of §1: the emission is the description, and the record is whatever the scaffold does with descriptions. Deciding e *at the frame* is the convention under which a verb is a property of the description written when the verb happened, which is what every caption-based memory assumes when it indexes captions by verb.

Lemma 2 (A single frame is reversal-invariant)**.** *A single keyframe is the same observation in $x_{1:N}$ and in $Rx_{1:N}$: the window of a $w{=}1$ write at position t of x is the window of the same write at position $N{+}1{-}t$ of Rx. Hence for any $f\colon \mathcal{X} \to \mathcal{T}$ the emission satisfies $\tau_t(x) = \tau_{N+1-t}(Rx)$, and any predicate decided at the frame by a frame-local $w{=}1$ write is fixed by R: $e^R = e$. In particular a $w{=}1$ frame-local write assigns* put *and* take *the same value at every position, and neither member of a reversal pair $e \neq e^R$ is decided at the frame by any such write.*

This is the deterministic core of the wall, and it is a statement about *arguments*, not about models: whatever f is (a caption, an embedding, a status table), it receives the same frame in either direction and returns the same value. The lemma says more than the theorem's closure: $\mathcal{P}_1$ is not merely closed under R, it is *fixed* pointwise, so a 1-writable predicate is its own conjugate. A predicate that is not its own conjugate is therefore outside $\mathcal{P}_1$ before anything is said about distributions.

The theorem's distributional clause covers the case the deterministic lemma does not: two occurrences of *put* and *take* are not literally the same frame, but the frame at a *put* and the frame at a *take* are draws from the same configuration. That is a hypothesis about the world and it is stated as one.

Assumption 1 (Reversal-symmetric single-frame marginals)**.** For a reversal pair (e, e^R), the law of the keyframe at an occurrence of e equals the law of the keyframe at an occurrence of e^R: $P[x_t \in \cdot \mid e_x(t){=}1] = P[x_t \in \cdot \mid e^R_x(t){=}1]$. The hand–object configuration at the midpoint of a *put* and at the midpoint of a *take* is one configuration traversed in opposite directions.

Assumption 1 is exactly what Theorem 1 already assumes under the name "reversal-symmetric single-frame marginals"; it is isolated here because it is the only empirical premise in the identity wall, and because it is testable: the *put*-written-*take* confusion of §4.1 is what the assumption predicts a single-frame write will do, and its size is the assumption's measurement. Figure 5 draws the geometry: the hand–object configuration at the midpoint of a *put* and at the midpoint of a *take* is one configuration traversed in opposite directions.

Lemma 3 (Posterior equals prior)**.** *Under Assumption 1, for any measurable f and any position t at which one of e, e^R occurred with prior $(\pi, 1-\pi)$, $P[e_x(t){=}1 \mid f(x_t)] = \pi$ almost surely. The emission of a frame-local $w{=}1$ write is uninformative about which member of the pair occurred.*

Corollary 4 (The direction ceiling of a single-frame write)**.** *Under Assumption 1, let δ be any decision rule for the direction of a gold* put/take *frame that depends on the frame through a $w{=}1$ emission. Then $P[\delta \text{ correct}] \leq \max(\pi, 1-\pi)$, with equality for the constant rule that names the majority direction. The best achievable direction accuracy of a single-frame write on gold* put/take *frames is the majority prior, and a rule may sit below it.*

The corollary is the number the direction column of §4.1 is read against: the majority prior 74% on the 200 gold *put*/*take* frames. It is an upper bound and not an estimate. A single-frame rule that

writes *take* for most *put*s is consistent with it, and a $w{=}2$ rule that clears it has read the pair; one that does not has not. The bound is also the reason the `dup` control is decisive: two copies of one frame are one frame, so `dup` is a $w{=}1$ emission with two images attached and is under the same ceiling, whatever its prompt does to the vocabulary.

Proposition 5 (Statuses are 1-writable; their transitions are 2-writable and exact)**.** *Let $\sigma\colon \mathcal{X} \to \Sigma$ be a per-object status decided at the frame by a $w{=}1$ frame-local write (*open/closed*,* held/*not,* on/off*). For $a \neq b$ in Σ the transition predicate $e_{a\to b}(t) := [\sigma(x_{t-1}){=}a \wedge \sigma(x_t){=}b]$ is 2-writable, is computed exactly by the first difference of the status trajectory $\sigma(x_{1:N})$ with no further perception, and satisfies $(e_{a\to b})^R = e_{b\to a}$ up to the alignment of the window. Hence every reversal pair of verbs that is the difference of a 1-writable status is in $\mathcal{P}_2 \setminus \mathcal{P}_1$, and its two members are separated by the difference and not by the model.*

This is W1′ stated as a theorem. The escape the wall's proof names (retain a variable the verb mutates) is taken by asking the model for what one frame decides and having the scaffold take the derivative. The verb becomes a lookup in a six-row table (Appendix H) rather than a perception, and the direction of the verb is exactly the direction of the difference: *closed→open* is *open*, *open→closed* is *close*, and the two cannot be confused because they are different rows. What the proposition does not say, and what §4.1 measures, is whether the backbone's status at the two flanking frames is right; the exactness is conditional on the statuses, which is the identity factor relocated to a quantity a single frame *can* decide.

Where the record enters, and where it does not. Definition 7 lets the record s_{t-1} enter the update U but not the emission f. A write that stores raw emissions and lets a later operation combine two of them has not decided anything at the frame; it has moved a window of length two into the algebra, and by Proposition 2 that is precisely what W1′'s differencing does. The move requires the emission to carry something the operation can difference (a status), and the standard caption write carries a verb, which Lemma 3 says is at chance. The theorem is about what is written at the frame; the architecture is about what can be computed over what was written.

B.5 The ordinal wall, in two forms

Proposition 6 (Ordinal wall, storage form)**.** *Under per-class storage of $O(1)$ words (a recurrence count and a last-seen timestamp), $\mathrm{nth}(e,k)$ is unanswerable for all $k \notin \{1,m\}$, at any perception quality: addressing the k-th of m occurrences over a stream of length N requires $\Omega(m\log(N/m))$ bits per class.*

This is the form we registered, and it is the form that locates Das & Roy (2026): a recurrence count and a last-seen time are the two words the proposition names, and a benchmark closed under $\{\mathrm{count}, \mathrm{last}, \mathrm{most}, \mathrm{usually}\}$ never asks for an interior occurrence, so the representation and the questions are closed under the same operations and the wall is not exposed. It is also the form the measurement says is *not* binding once the storage is fixed: the witness ledger of §C.2 retains every occurrence, the bound is met, and the ordinal factor is still small. The reason is the second form.

Proposition 7 (Ordinal wall, error-propagation form)**.** *Let the write produce, for class e, a detected occurrence list in which each true occurrence is detected independently with probability $\rho < 1$ and detections are otherwise exact (no false positives, no position error). Then*

$$P[\,\mathrm{nth}(e,k) \text{ resolves to the true } k\text{-th position}\,] \;=\; \rho^k,$$

and reading the resolved occurrence together with its r neighbours on either side in the detected list raises it to exactly

$$\rho\, P[\mathrm{Bin}(k{-}1, 1{-}\rho) \le r] \;=\; \sum_{j=0}^{r} \binom{k-1}{j} \rho^{k-j}(1-\rho)^j \;=\; \rho^k\Big(1 + \sum_{j=1}^{r} \binom{k-1}{j} \big(\tfrac{1-\rho}{\rho}\big)^j\Big),$$

which for $r{=}1$ is $\rho^k\big(1 + (k-1)(1-\rho)/\rho\big)$ and for every r is at most ρ, the recall of the addressed occurrence itself. With false positives at rate ϕ per true occurrence the exact probability is not monotone in ρ but is bounded by $\rho \cdot P[$exactly $k{-}1$ detections precede the true k-th$]$; it is largest when ϕ is near $1-\rho$, where false positives replace misses at the rate they occur, its envelope over all ϕ is ρ times the largest point mass of $\mathrm{Bin}(k{-}1,\rho)$, which tends to zero in k, and for every fixed $\phi \neq 1-\rho$ the decay in k is geometric.

The proposition says something the storage form cannot: *the witness ledger being complete is not enough*. What enters the ledger is a detection, and the address algebra operates over detections. A ledger with every detection kept still resolves $\mathrm{nth}(e, k)$ at ρ^k; a scaffold that filters the ledger (debouncing, merging, unioning two detectors) changes which detections are kept and not the rate at which true occurrences were detected in the first place. The measured per-k curve of §E.3 is this proposition's curve, and the four refuted remedies are its corollary. The next subsection states the two forms together and makes the corollary a statement.

B.6 The error-propagation form as a curve: two regimes

The two forms of the ordinal wall are two necessary conditions on one event, and a necessary condition is a factor of an upper bound. Write b_e for the witness bits the record holds for class e, and say the record *stores the k-th witness* if $\mathrm{nth}(e, k)$ is computable from what it holds when every occurrence has been detected.

Lemma 4 (The two forms compose as a gate and an envelope)**.** *Under the model of Proposition 3,*

$$F_{\mathrm{ord}}(k) \;\le\; \mathbf{1}\big[\textit{the record stores the } k\textit{-th witness}\big] \cdot \rho^k,$$

with equality when detections are the only source of error. Two regimes follow. In the storage regime *the indicator is zero for interior k (Proposition 3 at $O(1)$ words per class), and $F_{\mathrm{ord}}(k)$ is zero whatever ρ is. In the* propagation regime *the indicator is one for every k (a complete witness ledger), and $F_{\mathrm{ord}}(k) = \rho^k$, so the factor is set by the detector alone and no operation on the record moves it.*

The lemma is the account of why E2 (§D) was not run. E2 was to move the indicator from zero to one at equal bytes and measure the jump on ordinal queries. The ledger of §C.2 sets the indicator to one for every class the corpus queries, so E2's independent variable is already at its ceiling, and the ordinal factor the measurement reports is the envelope, not the gate. The two regimes are distinguishable on a cache without an end-to-end run: in the storage regime the factor is zero at interior k and unchanged by a better detector; in the propagation regime it is positive, falls geometrically, and its ratio between consecutive k is the detector's recall.

Corollary 5 (The per-k curve estimates ρ without a fit)**.** *In the propagation regime, $F_{\mathrm{ord}}(1)$ equals the per-occurrence recall of the class's first occurrence, and for every k, $F_{\mathrm{ord}}(k{+}1)/F_{\mathrm{ord}}(k) = \rho$. The detector's recall is therefore read off the per-k curve twice over (as its intercept and as its ratio), and neither reading involves a fitted parameter.*

The corollary is how Figure 6 is to be read. The intercept at $k{=}1$ is the recall of the queried class's first occurrence; the ratio of successive points is the recall of each subsequent one; and the right block of Table 4 measures the same recall directly, per occurrence, against the gold list. Three readings of one quantity agree in §E.3 to within the range 0.2–0.35, and the agreement is what licenses saying the regime is the propagation one.

Proposition 8 (No deletion filter raises the ceiling)**.** *Let Φ be any map from detected lists to sublists of themselves (debouncing, merging of near-duplicates, minimum-run thresholds, and their compositions), applied to the ledger before an address is executed. Under the model of Proposition 3 with false positives at any rate,*

$$P[\,\mathrm{nth}_\Phi(e, k) \textit{ resolves to the true } k\textit{-th position}\,] \;\le\; \rho,$$

the recall of the addressed occurrence itself; and whenever the kept list carries no false positive before the true k-th occurrence (in particular for a detector without false positives), the bound is ρ^k. A filter acts only through the false positives it removes and the true detections it discards; it cannot add a detection that was not made. Taking the union of two detectors with per-occurrence recalls ρ_1, ρ_2 replaces ρ by $\rho_\cup \le 1 - (1-\rho_1)(1-\rho_2)$, with equality only if the two detectors' misses are independent, and $\rho_\cup = \max\{\rho_1, \rho_2\}$ when they coincide.

The proposition is the four refuted remedies of §E.3 stated in advance, and it says which of them could ever have worked. Debouncing, persistence and merging are deletion filters and are under the ceiling by construction; the measurement finds they are under it by more, because the ledger under-counts and deletion makes an under-count worse. The union is the one remedy that is not a deletion filter, and the proposition prices it: it helps to the extent that the two detectors miss different occurrences. The `action` stream and the differenced stream are two readings of one backbone's

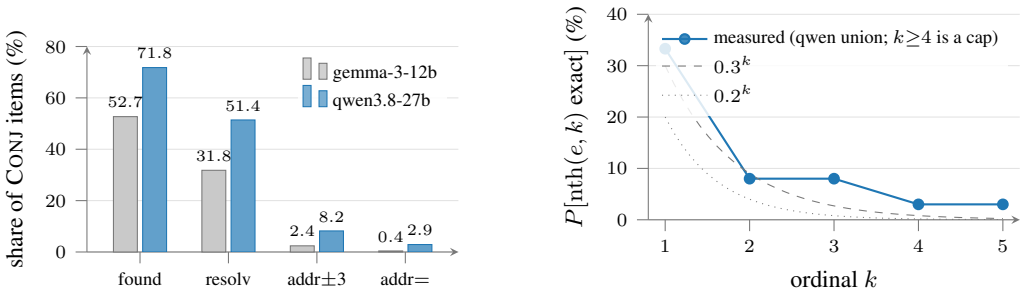


Figure 6: Left: the funnel. The queried class is found in the ledger for most items; the ledger holds k occurrences for half of them; the k-th occurrence is the right frame for almost none. **Right: the per-k curve against Proposition 3.** Exact ordinal resolution by ordinal, qwen3.8-27b union ledger, with ρ^k for two values of ρ. The $k{=}1$ point is the per-occurrence recall of the class's first occurrence (33.3%); at $k \geq 4$ the measured value is at most 3% and is drawn at the cap. The curve is compatible with $\rho \approx 0.3$ from $k{=}1$ to $k{=}2$ and is flat from $k{=}2$ to $k{=}3$; with measured precision of 12.3–23.3% the ledger carries several false positives per true detection, and Proposition 3's geometric form is derived without them, so we read ρ off the first step and treat the rest as a decay whose rate the proposition does not pin down in this regime.

view of one pair of frames, and Table 4 reports their union within noise of either, which is what correlated misses look like. What is left after the four is ρ itself, and ρ is a property of the detector on a specific object at a specific moment.

B.7 COMPOSITION BY ADDRESS

The parent system (Anonymous authors, 2026f) composes its two channels by concatenation, $\oplus_{\text{cat}}$: both the compressed state and the retrieved verbatim frames are rendered to text in one prompt and the backbone recovers the join. Its measurement found no super-additivity, and the companion measurement (Anonymous authors, 2026a) shows the failure was arithmetic before it was architectural: on two of three query families the additive bound is not even available. On the family where it is, the join still has to be made, and a join made in text must survive the flattening of a position into a description.

Definition 8 (Composition by address)**.** Two channels $\mathcal{C}_{\text{s}}$ (state) and $\mathcal{C}_{\text{v}}$ (verbatim) compose *by address*, $\mathcal{C}_{\text{s}} \oplus_{\text{addr}} \mathcal{C}_{\text{v}}$, when the read side executes an expression of $\mathcal{A}$ over $\mathcal{C}_{\text{s}}$'s record to obtain a position, and $\mathcal{C}_{\text{v}}$ is dereferenced at that position and nowhere else. The state channel returns a location; the verbatim channel is read there.

Under $\oplus_{\text{addr}}$ the two sides of the product are carried by different channels and multiplied by the architecture rather than by the backbone. Under $\oplus_{\text{cat}}$ the same product has to be recovered by a model reading two blocks of prose, and the companion paper (Anonymous authors, 2026b) measures what such a reader does with two blocks that disagree. The design consequence is the one AXIS follows: the address is *executed* by the scaffold, never spoken to the backbone as a hint. The next three statements say what the executed address is worth, what its controls measure, and when it can be super-additive at all.

Proposition 9 (Address composition realises the product)**.** *Under $\mathcal{C}_{\text{s}} \oplus_{\text{addr}} \mathcal{C}_{\text{v}}$ the chain events of Definition 3 are carried by different channels (E_{id} and E_{ord} by the record of $\mathcal{C}_{\text{s}}$, E_{pc} by the one dereferenced frame of $\mathcal{C}_{\text{v}}$), and*

$$P[\textit{correct}] \;\geq\; F_{\text{id}}\, F_{\text{ord}}\, F_{\text{pc}},$$

where F_{pc} is the percept factor of $\mathcal{C}_{\text{v}}$ on a correctly addressed frame and no term depends on the backbone joining two texts. Under $\oplus_{\text{cat}}$ the same bound holds only if the backbone recovers the position from the state's prose, an event whose probability multiplies the right-hand side and is not one of the three factors.

The proposition is the reason the two compositions differ in kind and not in degree. By address, the factor that measures the join is exactly one: the scaffold resolved a position and dereferenced it, and nothing was lost between the two channels. By concatenation, the join is a fourth factor, and the

companion measurement (Anonymous authors, 2026b) reports what it costs when the two blocks disagree. The proposition does not say the product is large; it says the product is *all there is*, which is what makes the two controls informative.

Proposition 10 (The control gap is bounded by the chain probability)**.** *Let* `axis` *answer through the executed address and let* `axis_wrongaddr` *execute the same address and dereference a different occurrence of the same class, so that its chain is broken on every query. Suppose a broken chain in* `axis` *dereferences a wrong frame of the same class or none, with conditional accuracy* $\varepsilon_{\mathrm{off}}$ *equal to that of the control. Then*

$$\left|\,\mathrm{acc}(\texttt{axis}) - \mathrm{acc}(\texttt{axis_wrongaddr})\,\right| \;\le\; F_{\mathrm{id}}\,F_{\mathrm{ord}}\cdot\left|F_{\mathrm{pc}} - \varepsilon_{\mathrm{off}}\right| \;\le\; F_{\mathrm{id}}\,F_{\mathrm{ord}}.$$

A chain that holds on a small fraction of queries leaves the control within that fraction of the arm it controls, whatever the percept factor.

This is falsifier F4 of §D turned into an inequality. The control was registered to price the frame's content: if the address is right and the frame matters, dereferencing the wrong frame should cost the percept headroom. The proposition says the cost is paid only on the queries where the chain held, so a measured chain probability predicts the control gap before the cell runs, and a gap inside the bound is not a failure of the control but a measurement of the chain. §E.3 writes the prediction down from the cache and §4.2 reports the gap.

Proposition 11 (When the address read is strictly super-additive)**.** *Write* $a_{\varnothing}$, a_{s}, a_{v} *for the accuracies of the blind, state-only and verbatim-only arms on the conjunctive family, and* a_{addr} *for the address read. Strict super-additivity in the additive sense of Chen et al. (2026),* $a_{\mathrm{addr}} - a_{\varnothing} > (a_{\mathrm{s}} - a_{\varnothing}) + (a_{\mathrm{v}} - a_{\varnothing})$*, requires the availability condition of the companion paper (Anonymous authors, 2026a),* $a_{\mathrm{s}} + a_{\mathrm{v}} - a_{\varnothing} < 1$*. Under* $\oplus_{\mathrm{addr}}$ *with the floor convention* $\varepsilon_{\mathrm{off}} = a_{\varnothing}$ *it requires the sharper* $a_{\mathrm{s}} + a_{\mathrm{v}} - a_{\varnothing} < F_{\mathrm{pc}}$*, and it holds if and only if*

$$F_{\mathrm{id}}\,F_{\mathrm{ord}} \;>\; \frac{(a_{\mathrm{s}} - a_{\varnothing}) + (a_{\mathrm{v}} - a_{\varnothing})}{F_{\mathrm{pc}} - a_{\varnothing}}.$$

The chain probability must exceed the two single-channel gains, summed, in units of the percept headroom.

The proposition puts the companion paper's arithmetic and this paper's product on one line. Availability (Anonymous authors, 2026a) says the effect is permitted when the single channels leave room below one; the proposition says that under address composition the room is below the percept factor, not below one, because the best an executed address can do is read one frame correctly. And it says what the room has to be filled with: chain probability. On the conjunctive corpus the availability threshold is 42.3% and the single-channel arms sit near the floor, so the right-hand side is small and the condition is undemanding; the measurement of §E.3 is that $F_{\mathrm{id}}F_{\mathrm{ord}}$ is smaller still. The availability was certified and not used, and the proposition says the unused quantity is the chain.

B.8 What the theory predicts, and what it does not

The statements above contain no fitted parameter and no measured number. Each names a quantity, and each quantity is measured somewhere in §4. The map is the following, factor by factor.

- F_{id}**, and the symmetry.** Theorem 1 and Corollary 4 predict that a single-frame write's verb at the gold locator frame is at chance on reversal pairs and that its *put*/*take* direction cannot exceed the majority prior 74%. Measured: verb accuracy at the gold frame and the confusion matrix on the reversal pairs (§4.1); the direction column against the prior. Lemma 2 predicts that `dup` is under the same ceiling as `single` whatever its prompt does, and Proposition 2 predicts that a $w{=}2$ write can clear it. The theory does *not* predict whether a given backbone reads the pair at a given spacing δ; that is a property of the detector and is what §E.4 measures on two spacings.
- F_{ord}**, and the regime.** Lemma 4 predicts that a complete ledger puts the factor in the propagation regime, where it is ρ^k; Corollary 5 predicts that the per-k curve's intercept and ratio both equal the per-occurrence recall measured directly against the gold list; Proposition 8 predicts that debouncing, persistence and merging cannot raise exact resolution above the single-occurrence recall ρ, nor above ρ^k where the kept list is free of false positives, and that the union helps only where the two detectors' misses are uncorrelated; Proposition 3 bounds the neighbourhood read at

$r{=}1$ by a factor of $1 + (k-1)(1-\rho)/\rho$ over the exact rate. Measured: the *addr* columns and the right block of Table 4, the curve of Figure 6, the four remedies, and the neighbourhood read at 3–15% (§E.3). The theory does *not* predict the value of ρ; it predicts what follows from whatever value is measured.

- F_{pc}**.** The theory here says nothing new about the percept factor; it inherits the read-side walls of Chen et al. (2026) and takes the factor as measured on correctly addressed frames by the companion grid's `retr_only` arm.
- **The chain, and the controls.** Corollary 1 predicts the executed-address arm's ceiling above the blind floor from the chain probability measured on the cache; Proposition 10 predicts the gap to `axis_wrongaddr` from the same number; Lemma 1 says the gap to the floor itself is not predicted but measured, by the control. Measured: the CONJ cells of §4.2 against the values written down in §E.3 before they ran.
- **Composition.** Proposition 9 predicts that `axis_noderef`, which speaks the address instead of executing it, differs from `axis` by a join factor and not by a factor of the product; when the chain probability is small the join factor multiplies almost nothing and the two arms coincide. Proposition 11 says on which side of the availability threshold of (Anonymous authors, 2026a) the arm can land and what would have to be true of the chain for it to cross.

What it does not say. The theory does not predict a benchmark gain anywhere; the conjunctive fraction of the field's benchmarks is 0.51% (Anonymous authors, 2026a), and on the rest of their items the product law has no locator to bound. It does not predict the noun factor, which is a single-frame percept and is outside the identity wall. It does not predict that the pair write improves verbs on any backbone; it predicts that the pair write *may* and the single-frame write *cannot*, and the controls decide which happened. And it does not predict ρ. Where the theory has a free quantity it is a property of the detector, and the paper's conclusion is that this is where the free quantities live.

C THE ARCHITECTURE AND THE EXPERIMENTAL DESIGN, IN FULL

Each component below is derived from a wall of §2, and the paragraph that introduces it says which. Nothing was tuned on a result; the status-transition table and the address grammar were fixed and registered before any grid cell ran (Appendix K).

C.1 W1′: STATUSES ARE WRITABLE, VERBS ARE DERIVATIVES

Theorem 1's proof names its escape: $\mathcal{W}$ separates e from e^R iff the record retains a variable that e mutates. A door's openness, an object's being in hand, a tap's running are properties of a single frame. So the architecture asks the model for what $w{=}1$ decides and computes the rest.

The call. The write call at keyframe t receives the ordered pair (x_{t-1}, x_t) and the current record, and emits a JSON object whose fields are answered *in a fixed order*: `change` first (one clause naming what differs between the two images, what entered or left the hand, what became more or less open, what was switched), then the `action` it implies as *verb noun*, then `affordances`: every visible openable, switchable or holdable object with its status in the current image (`fridge: open`, `cup: in-hand`, `tap: off`), carried forward from the record and changed only where the pair shows a change; then `location` and `objects`. The full prompt is Appendix I. The order matters: the prompt derives the action from the change, and the change is the quantity the pair exposes and the single frame cannot. The same prompt, with the same field order, serves every write rule of §D; only the attached frames differ.

Why the pair is one keyframe back. The design note proposed decoupling the pair's spacing δ from the keyframe spacing, extracting a second frame about a second before x_t; at EgoSchema's 4.5 s spacing "what changed" between keyframes is meaningless. On the conjunctive corpus the keyframes are action-segment midpoints, one per segment, so consecutive keyframes are consecutive actions and the previous keyframe is the natural "before". That is the rule used here and registered. On TempCompass the keyframes are 1 fps, and the previous keyframe is one second back.

The difference. The scaffold takes the first difference of the status trajectory $\sigma_{1:N}$ object by object. A transition $closed \rightarrow open$ *is* the event "open obj"; $\rightarrow in\text{-}hand$ is "take"; $in\text{-}hand \rightarrow$ is "put"; $off \rightarrow on$ is "turn-on"; and their inverses likewise. Statuses are first mapped to one of five classes (*open*, *closed*, *held*, *on*, *off*) by a fixed synonym table; a change between unclassified statuses emits no event and is counted. The table (Appendix H) has six rows and was fixed before any grid ran. The derived stream is exact in the sense that matters: given the statuses, the verb is a lookup, not a perception. The model's own `action` field is cached and scored as a comparison arm, never as the authority; and the measurement of §4.1 says which stream is better on which backbone, which was not the answer we registered.

What it costs. The pair write costs $1{,}600$–$1{,}850$ tokens per keyframe at scale, level with the single-frame path's $1{,}769$. A 20-keyframe smoke run had suggested $1{,}350$; that was an early-frame artefact (the record is short at the start of a stream), and the paper reports the at-scale figure. The repair therefore costs what it replaces: one extra image, no extra call, no extra output.

C.2 W2: THE WITNESS LEDGER

Proposition 3 names the storage escape: keep the witnesses. The ledger $\mathrm{ledger}[e] = [t_1 \dots t_m]$ is maintained by the scaffold from the derived event stream at zero API cost and zero prompt tokens: only its compact vocabulary (the set of event strings it can address) enters the read prompt. Two rules are fixed. Consecutive identical events collapse to one occurrence: a persisting action described on several keyframes is one action. And both sides of every lookup (the query's *verb noun* and the ledger's keys) are canonicalised through EPIC-KITCHENS-100's published noun and verb class tables (Damen et al., 2022), which map *gloves*/*glove*, *cabinet*/*cupboard*, *mug*/*cup*, *place*/*put-down* to one class each. The tables are corpus metadata, not answers, and the canonicalisation is the stemming a lexical index performs; before it, 51 of 245 conjunctive gold nouns matched no ledger key at all (§D.1). Lookup is exact on the canonical pair first, then by same-verb Jaccard ≥ 0.5 on the noun tokens.

Against the storage form of the ordinal wall, retention degrades *across* classes and never within one: the design keeps full witness lists for every class the corpus can query, and a budgeted deployment would degrade the tail of low-query-mass classes to count-plus-last, exactly the representation of Das & Roy (2026), kept for the classes where it suffices. On the corpora here no class is degraded; the ledger is complete, which is what lets §E.3 attribute the remaining ordinal loss to Proposition 3 rather than to storage.

C.3 W3: THE ADDRESS ALGEBRA, EXTENDED EXACTLY AS FAR AS THE WRITE LICENSES

$$\mathrm{nth}(e,k) \;\big|\; \mathrm{first}(e) \;\big|\; \mathrm{last}(e) \;\big|\; \mathrm{count}(e) \quad [\pm d] \qquad\qquad \mathrm{first}(obj{=}status) \;\big|\; \mathrm{last}(obj{=}status)$$

The ordinal operators are adopted from prior work (Li et al., 2026; Das & Roy, 2026) and credited as such. What is new is what they run *over* (differenced events rather than $w{=}1$ captions), and the state addresses on the right, which no single-frame write can define because no single-frame write retains a trajectory: $\mathrm{first}(fridge{=}open)$ is the first keyframe at which the fridge's status entered *open*, read off the same $\sigma_{1:N}$ the difference was taken from.

The read loop. The backbone receives the question, the ledger's vocabulary and the grammar, and emits one action per round: an `ADDR` expression, a `STATE` read, or an `ANSWER`. The scaffold resolves the expression and dereferences *one* keyframe verbatim into the next round's context. Two guards were added after a smoke run and before any cell: a resolved address is cached for the episode so that the same expression is never dereferenced twice, and after two dereferences the prompt asks for the answer. Every round is charged to the same $B{=}4096$ budget every other arm has.

The two controls. Because the address is executed by the scaffold, its contribution can be priced by breaking it in two different places at identical cost. `axis_noderef` resolves the expression, hands the resolved occurrence list to the backbone *as prose*, and chooses the frame by similarity as the parent system does: the address is spoken, not executed. `axis_wrongaddr` executes the address and dereferences a *different* occurrence of the same event, chosen from the other occurrences by a fixed hash of the item so that the choice is reproducible and never coincides with the correct one: the

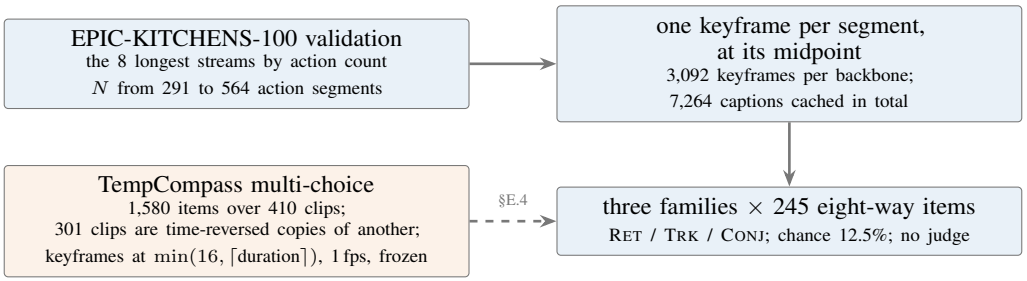


Figure 7: The corpora, as constructed and frozen before any cell ran. Blue: the conjunctive witness corpus, the registered population every identity, ordinal and end-to-end number in this paper is scored on. Orange: the public-benchmark reference set, whose reversal pairs supply the manipulation check nobody designed as one (§E.4). Construction details and the generator are in Appendix K.

frame is wrong and everything else is identical. The first control prices the execution; the second prices the frame's content. Their registered predictions are in §D.

D EXPERIMENTAL DESIGN

Registered in full before any cell ran (Appendix K); each experiment isolates one factor of the product.

Corpora. Figure 7 shows both corpora as constructed and frozen. *The conjunctive corpus* of the companion measurement (Anonymous authors, 2026a): eight EPIC-KITCHENS-100 validation streams (the longest eight, N from 291 to 564 action segments), one keyframe per segment at its midpoint, 3,092 keyframes per backbone, three families of 245 eight-way items each (RET, retrieve a percept at a described moment; TRK, track a count; CONJ, the k-th time you e, what happened next), with chance 12.5% and no judge. Its availability threshold on CONJ is 42.3%. *TempCompass* (Liu et al., 2024): the `multi-choice` split, 1,580 items over 410 clips, five dimensions (action, direction, speed, order, attribute change); 301 clips are time-reversed copies. Keyframes at $\min(16, \lceil \text{duration} \rceil)$ per clip, 1 fps, frozen. *EgoSchema-500* (Mangalam et al., 2023) is registered as the long-horizon test and deferred behind a cost gate (Appendix K, §8); it is not run in this version and nothing here depends on it.

Backbones and budget. gemma-3-12b, qwen3.8-27b and gemma-4-31b on the conjunctive corpus; gemma-3-12b and qwen3.8-27b on TempCompass. Every arm on every benchmark decides under B=4096 tokens; a comparison at unequal budgets is not a comparison. All calls go through one API provider at temperature 0; the offline write phase is cached once per (rule, backbone, episode) and every read arm reads the same cache.

E1 (identity). Hold the read side fixed; swap only the write rule. `single` (s, x_t) against `pair` (s, x_{t-1}, x_t), with controls `dup` (s, x_t, x_t) (two images, no temporal information, token-matched) and `rev` (s, x_t, x_{t-1}) (same pixels, wrong direction). One prompt serves all four; only the attached frames differ. Verb and noun accuracy at the gold locator frame are computed by *the same function* on every cache, so a difference between rules is not a difference between two implementations of "accuracy". On the same pair cache, the differenced event stream is scored against the model's `action` stream (P-Δ1). Coverage caps, declared: `dup` and `rev` run on gemma-3-12b only.

E2 (ordinal). Hold the write rule fixed; swap the retention policy at equal memory bytes (witness lists against count-plus-last consolidation) on count-like and ordinal queries separately. The registered prediction is a null on the former and a jump on the latter. E2 is priced but not run in this version; §E.3 explains why the measurement made it moot before it was paid for, and Limitation (5) states what it would still settle.

E3 (the product). The conjunctive corpus, all three families, all three backbones. Read modes `axis` / `axis_noderef` / `axis_wrongaddr` (§C.3). Baselines, from the companion grid at the same budget on the same keyframes: `blind`, `retr_only` (caption plus similarity retrieval, the standard pipeline), `state_only`, and ASP (concatenative composition of the two channels). Twenty-seven AXIS cells (three backbones, three read modes, three families).

H2 (the window on a public benchmark). TempCompass with the read path held fixed: write rules `single`, `pair`, `dup` (gemma-3-12b only) on one domain-neutral prompt (Appendix I), a record-only read that never sees an image, and two references: `blind` and `uniform` (all keyframes attached to one call).

E1″ and E3′ (Amendment A5: the list removed, the spacing varied, the read priced alone). Three further write rules under one prompt that carries *no* verb list: `pair_free` and `dup_free` attach the same frames as `pair` and `dup`; `delta_free` attaches the frame cut one second before the keyframe and the keyframe. Both are run on gemma-3-12b and qwen3.8-27b, so the A4 controls exist on two backbones. Three further read modes, `axis_oracle`, `axis_oracle_noderef` and `axis_oracle_wrongaddr`, run the unchanged read loop over a ledger built from the corpus's annotations, on the class-semantics corpus the companion regenerated after its occurrence-semantics defect was found (v3; 308 items per family). They are declared as oracles wherever they appear and exist to price the read side with the write's factors held at one. A per-round completion cap (`_cap256`) is added as a new arm so the larger backbone can commit an answer. Predictions P-δ1–3, P-O1–3 and falsifiers F-δ, F-O are in Appendix K and judged in Table 10.

Registered predictions and falsifiers. Verbatim from the registration, with their status at the time of writing marked * where measured:

- **P-D1a***: `pair` − `single` ≥ 15 points of verb accuracy at the gold frame, per backbone. **F1**: below 5 points kills the write-side contribution.
- **P-D1b***: gold-*put*-written-*take* under `pair` below half its `single` value.
- **P-D1c***: `pair` − `dup` ≥ 10 points and `rev` $\leq$ `pair` -8: the window's *direction* is the mechanism, not the extra pixels.
- **P-Δ1***: the differenced stream beats the model's `action` stream on verb accuracy, per backbone. **F-Δ**: if worse, W1′ is dropped and reported.
- **P-Δ2***: CONJ clears 42.3% iff the status-tracking rate (fraction of gold locator events whose object carries a classified status at both flanking frames) is ≥ 0.6.
- **H3*** (fails on gemma-3-12b; unscorable on qwen3.8-27b even capped; fails on gemma-4-31b): `axis` exceeds 42.3% on CONJ and the best single-channel arm by ≥ 10 points.
- **P-D3a***: `axis` − `axis_noderef` ≥ 8 points on CONJ. **F3**: below 3 points, the gain is a prompt.
- **P-D3b***: `axis` − `axis_wrongaddr` ≥ 15 points on CONJ. **F4**: within 5 points kills the control.
- **P-offfamily***: on RET and TRK, `axis` lands within ± 5 points of the best single-channel arm; a win there is a corpus artefact, not composition.
- **H2***: `pair` − `single` ≥ 8 points overall and ≥ 15 on the direction dimension of TempCompass.
- **P-cost***: `axis`'s realised per-decision spend is below ASP's and below caption-plus-retrieval's.
- **F5**: a family on which `blind` exceeds chance by 12.5 points is void.

D.1 THE REVIEW THAT PRECEDES THE JUDGEMENT

A working rule of this programme is that a boundary or an unexpected result is not judged until the whole code path and process have been reviewed and found sound. That review was carried out before any number in §E.3 was read as a finding, end to end: frame–annotation alignment (all eight episodes contiguous, keyframes midpoint-exact, no truncated frame files); gate identity at the write threshold; cache health counters (parse failures ≤ 2 and truncations ≤ 3 per episode, no empty status field); provider image-order preservation, checked live by swapping the two frames and confirming both backbones swap FIRST/SECOND with them; raw model output at gold frames; normalisation; lookup; and timing offsets. It found two defects of ours and one cost claim of ours, all disclosed here and in the registration's third amendment.

Normalisation. The first lookup did no lemmatisation and no class mapping, so *gloves*/*glove*, *cabinet*/*cupboard*, *mug*/*cup* never matched and 51 of 245 conjunctive gold nouns matched no ledger key. The repair is the canonicalisation of §C.2; nothing in it was fitted to a result. One E3 cell had begun under the old normalisation; it is quarantined, not pooled, and E3 restarted for all backbones under one code state.

Cost. The claim that the pair write is cheaper than the single-frame write was a smoke artefact; at scale it is level (§C.1). Corrected everywhere it appeared.

Prompt cost. The pair prompt's enumerated verb list collapsed the written vocabulary from 12–17 distinct verbs to 7–11. Reported as a design cost; not changed mid-run.

Not a code artefact. At gold conjunctive locator frames the queried object is absent from the status field 32% of the time, present with unchanged status 44%, and flips exactly there 24%; matched occurrences scatter over ± 4 frames (199 at offset 0, 117 at $+1$, 54 at -1, 173 at $|d| \geq 2$); and over 1,500 derived transitions fall on objects the annotator has nobody acting on. That is the per-occurrence detection rate of Proposition 3, measured, and it survived every scaffold-side operation we tested.

E RESULTS, IN FULL

Every registered prediction has at least one measured verdict, and every cell typeset in this paper is measured; the one coverage cap is declared in the caption of Table 1. All write-path numbers are under the unified class-canonical scorer of Amendment A5 (§D.1); the earlier head-only scorer, under which the registration's M-1 was recorded, is retired and its numbers appear nowhere. The verdicts, in one place: P-D1a **fails** on every backbone (gains of $+8.6$ to $+11.5$ against 15; the falsifier F1 at 5 is not tripped); H2 **fails** on gemma-3-12b and on qwen3.8-27b **holds** overall ($+9.4$) while missing the direction threshold ($+11.1$ against 15); P-D1b, P-D1c, H3, P-D3a, P-D3b and P-cost **fail**; P-Δ1 holds on two of three; P-Δ2 predicted H3's failure and was right; P-offfamily holds on RET and fails on TRK; falsifiers F3, F4 and F-Δ trip; F1, F2 and F5 do not. Of Amendment A5: P-O1 **fails** against its registered figure, with the oracle arm nonetheless $+14.3$ above its wrong-address control and $+14.3$ above blind; P-O2 **holds** in full on gemma-4-31b ($+23.4$ on both halves, against lines of 15 and 8) and on gemma-3-12b misses its wrong-address line by less than a point while **holding** its no-dereference half, with qwen3.8-27b unscorable; P-O3 **holds** on RET and fails on TRK in the oracle's favour; P-δ2's direction line **holds** on qwen3.8-27b (uncorrected p=0.037; no family-wise correction across the registered family survives it, Limitation (2b)) and fails on gemma-3-12b; P-δ1 and P-δ3 are decided by the collapse of the `dup_free` control on qwen3.8-27b rather than by the spacing they were written about, and are reported as such (Table 10).

Execution status, and the future-work boundary. Every cell of the registered grid is measured except one declared coverage cap: the `dup`/`rev` control arms outside gemma-3-12b, omitted from the tables and stated in the caption of Table 1. The free-vocabulary rules and their spacing statistics on gemma-4-31b have landed, and they split: the zero-spacing collapse replicates (a single distinct verb, the *hold* family, over 488 frames, no *put*/*take* scored), while the above-prior direction read does not (64.9% against the 74% prior, p=0.988; spacing contrast $+4.6$ points, p=0.468; uncorrected), reported in full in §4.1. These arms are replications on additional backbones of directions this paper already measures (the identity and ordinal walls and the `dup`/`rev` controls exist in full on gemma-3-12b, the free-vocabulary rules and the spacing contrast on qwen3.8-27b), and no verdict above depends on them; each is bound to its registered prediction and falsifier (Appendix K), so a completed cell can only widen or narrow an attribution this paper already scopes, not overturn a wall. The oracle block that an earlier draft listed here as running has since landed on all three backbones, and P-O2 holds on gemma-4-31b; that is a verdict gained, not a gap. The future-work boundary is explicit: the spacing sweep beyond the two spacings cut here (Limitation (2b)), the `dup`/`rev` arms outside gemma-3-12b declared out of coverage above, E2 as priced in §3, EgoSchema-500 behind its cost gate, and the local-serving arm the registration named as primary. Out-of-coverage arms are stated where the reader should not look for evidence, not where evidence is withheld.

Table 5: What the four reversible verbs are written as. For each gold verb at the gold frames, the three most frequent canonical verbs the write produced, as a share of that verb's n. A time-reversal error writes the verb's inverse; a static error writes `hold`. The free-prompt rules move mass from the inverse to `hold` rather than to the right verb.

backbone	rule	gold	n	written as
gemma-3-12b	`single`	`put`	128	mix 29%, wash 23%, hold 19%
gemma-3-12b	`single`	`take`	92	hold 24%, mix 23%, wash 18%
gemma-3-12b	`single`	`open`	45	hold 33%, open 31%, mix 16%
gemma-3-12b	`single`	`close`	46	mix 22%, hold 22%, take 20%
gemma-3-12b	`pair`	`put`	128	put 72%, take 23%, hold 2%
gemma-3-12b	`pair`	`take`	92	put 59%, take 30%, hold 5%
gemma-3-12b	`pair`	`open`	45	put 64%, take 29%, close 7%
gemma-3-12b	`pair`	`close`	46	put 54%, take 41%, open 2%
gemma-3-12b	`delta_free`	`put`	128	hold 48%, put 11%, grasping 10%
gemma-3-12b	`delta_free`	`take`	92	hold 35%, grasping 13%, take 11%
gemma-3-12b	`delta_free`	`open`	45	hold 27%, open 13%, close 13%
gemma-3-12b	`delta_free`	`close`	46	hold 35%, grasping 13%, put 13%
qwen3.8-27b	`single`	`put`	128	take 46%, wash 16%, hold 15%
qwen3.8-27b	`single`	`take`	92	take 48%, wash 18%, hold 12%
qwen3.8-27b	`single`	`open`	45	open 42%, take 31%, put 11%
qwen3.8-27b	`single`	`close`	46	take 39%, open 26%, hold 17%
qwen3.8-27b	`pair`	`put`	126	put 56%, take 28%, hold 8%
qwen3.8-27b	`pair`	`take`	92	take 52%, put 29%, hold 8%
qwen3.8-27b	`pair`	`open`	45	open 29%, take 24%, put 22%
qwen3.8-27b	`pair`	`close`	46	put 37%, close 24%, take 22%
qwen3.8-27b	`delta_free`	`put`	128	put 45%, hold 16%, take 9%
qwen3.8-27b	`delta_free`	`take`	92	take 30%, hold 13%, wash 10%
qwen3.8-27b	`delta_free`	`open`	44	open 57%, put 7%, take 7%
qwen3.8-27b	`delta_free`	`close`	46	put 22%, close 22%, take 17%

E.1 The write path: the identity factor before and after the window

One scorer, and what it did to the registered baseline. The fourth working principle obliged a review of the scoring chain before any free-vocabulary rule was judged, and the review found that the probe's verb normaliser compared hyphenated *heads* only: *put-down→put* was right, *placing* stayed *placing* and was wrong. A rule that writes from a list of canonical heads was thereby scored by a different function from a rule that writes free text. The scorer now maps written and gold verbs alike through the corpus's class tables with inflection stemming (*placing→put*, *grabbed→take*, *rinses→wash*), and every number in Table 1 is under it. The registration's M-1 was recorded under the head-only scorer; under the unified one the single-frame identity factor is 15.1, 23.3 and 11.0%, higher than registered by a few points on every backbone, and it is the registered *baseline* that moved, not the effect. We report the change here and in Appendix K rather than carry two scorers.

The gold-frame number moves, and not far enough. Verb accuracy at the gold frame rises from the single frame to the ordered pair by +9.8, +8.6 and +11.5 points on the three backbones. The registered threshold was 15 points: P-D1a **fails** on every backbone. The falsifier F1 (below 5 points) is nowhere near on any. Had the paper stopped here it would have reported the sign and roughly half the size Theorem 1 says the window must produce, and called it partial. It registered two controls so that it could not stop here.

Figure 3 plots both halves of the table.

The controls attribute the rise to the vocabulary, not the window. Figure 8 plots every control in this subsection against the write rules it tests. `dup` attaches the current frame twice: two images, the pair prompt, and no temporal information whatever. It scores 26.9%, *above* the ordered pair's 24.9%. `rev` attaches the same two frames in the wrong order and scores 22.7%, +2.0 being the duplicate's margin over the pair and the reversed pair within three points of it. P-D1c registered a

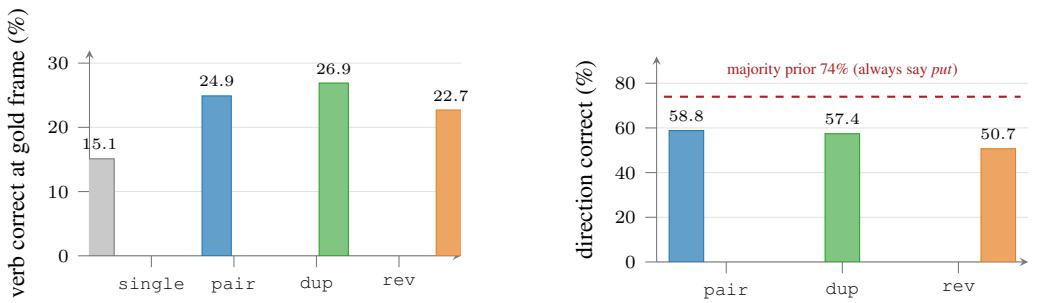


Figure 8: The controls refute the window's headline at one action apart (gemma-3-12b; `dup`/`rev` run on this backbone only, as declared). Left: verb accuracy at the gold frame: the duplicate frame (26.9%) matches and slightly exceeds the ordered pair (24.9%), and the reversed pair loses little (22.7% against 24.9%). **Right:** direction among gold put/take frames where the rule wrote *put* or *take* (n=204, 209, 209): every rule sits at or below the 74% majority prior. P-D1c registered pair−dup ≥ 10 and rev $\leq$ pair−8; it fails on both halves. Drawn from the same macros as Table 1.

10-point margin over `dup` and an 8-point deficit for `rev`; it **fails** on both halves, and not narrowly. The direction column of Table 1 says what the pair prompt did instead. The single-frame write rarely names *put* or *take* at all on a *put*/*take* gold frame (n=24 of the frames on gemma-3-12b); the pair prompt, which enumerates *put-down* and *take* as the first two of its seven verbs, names one of them on 204, 209 and 209 frames under `pair`, `dup` and `rev`. Among those writes the direction is right 58.8, 57.4 and 50.7%, all below the 74% a writer that always said *put* would score, and spread over eight points. The two-frame call did not decide the direction; it decided to use the two verbs the prompt listed, and the locator frames of a conjunctive corpus are heavily *put* or *take*. On qwen3.8-27b and gemma-4-31b the same prompt produces the same split at 65.6 and 60.0%, consistent with the control and not the control (Limitation 2a). The identity factor was not raised by the window on any backbone at this spacing. The ordered pair was presented and not read.

The list removed: the gain goes with it on one backbone and the duplicate frame collapses on the other. Amendment A5 reruns the same frames under a prompt with no verb list. On gemma-3-12b `pair_free` scores 11.4% (−13.5 against the listed pair and *below* the single frame), and `delta_free`, the pair one second apart, 11.2%; `dup_free` is at 10.6% on a partial cache. Whatever this backbone gained from the pair prompt, it gained from the list, and without the list it has nothing from two frames at either spacing. On qwen3.8-27b the picture is different in a way that is itself a finding. `pair_free` holds most of the listed pair's accuracy (29.4 against 31.8%), so this backbone does extract a verb from two frames an action apart once it is allowed its own vocabulary; and `dup_free` collapses to 0.0% verb-correct on a partial cache of 490 frames, with one distinct verb ever written: shown the same frame twice and given no list to guess from, the backbone writes *hold*. That is Theorem 1's content read off a cache: no change in the window, no transition in the record. P-δ3 registered that `pair_free` and `dup_free` are within three points if the one-action pair carries nothing; on gemma-3-12b that reading holds (+0.8) and on qwen3.8-27b it fails by the whole of `pair_free`, because the control fell to zero rather than because the pair rose. Both readings carry the partial-cache mark and neither is called (Table 10).

One second apart, the larger backbone reads the direction, and that is the paper's positive result about the window. Figure 9 plots the two arms. The comparison is the cleanest in the paper because it varies one thing. `pair_free` and `delta_free` share a prompt, a frame count, a backbone, a read, a scorer and the same 490 gold frames; they differ only in how far apart the two attached frames are: one action, or one second. On qwen3.8-27b, direction among *put*/*take* writes goes from 70.5% (n=105) to 82.7% (n=104): a gain of +12.2 points, z=2.08, p=0.037 on a two-proportion test. Against the 74% majority prior (the score a writer that always said *put* would post), 82.7% has one-sided binomial p=0.025. Every other rule in Table 1, on every backbone, sits at or below that prior with p above 0.7; this is the only rule that reads direction rather than guessing it, and it is the rule whose two frames differ by a motion rather than by a scene. Its verb accuracy, 33.9% on 489 frames, is the highest of any rule scored on the full gold population, +4.6 above the one-action free pair. On gemma-3-12b the identical move does nothing: 70.0% to 63.2% (−6.8,

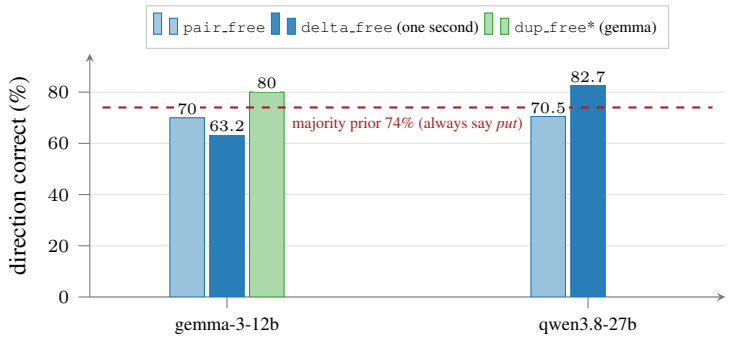


Figure 9: The window's escape is a property of the pair's spacing, read on one backbone. Same free-vocabulary prompt, same two frames' worth of tokens, only the gap changes: one action apart (consecutive keyframes) versus one second apart (`delta_free`). On qwen3.8-27b direction rises from 70.5% to 82.7% of put/take writes (+12.2 points, z=2.08, p=0.037), the only write rule in the paper above the 74% prior (p=0.025 against it). On gemma-3-12b the identical move does nothing (70.0%→63.2%, p=0.522). `dup_free` on gemma-3-12b is a partial cache (*, n=25 direction cells) and reads 80.0% on them, consistent with the 74% prior at that n; on qwen3.8-27b `dup_free` writes hold on every frame, so it has no direction cell at all: the identity wall's content read off a cache.

p=0.522), both indistinguishable from the prior, and verb accuracy flat at 11.4 and 11.2%. So the escape Theorem 1 licenses at w=2 is taken (by one backbone, at one spacing), and the parameter that decides it is not the window's length, which was two in both arms, but the interval it spans. Section E.4 finds the same split, on the same two backbones, at the same one-second spacing, on somebody else's benchmark.

The one control that reads above the prior on the smaller backbone, and why it is not evidence. gemma-3-12b's `dup_free` posts 80.0% direction, above the prior, on 25 writes, at one-sided p=0.336, which is not evidence of anything. It matters only because the rule is a *duplicate frame*: by Corollary 4 a w=1 emission cannot beat the prior, so a point estimate above it on twenty-five writes is exactly the noise the corollary predicts around its own ceiling. We record it rather than omit it because it is the number that makes P-δ2 fail on that backbone.

The confusion rates move, and the controls say they are a rebalancing. On qwen3.8-27b, which had the sharpest single-frame confusion, gold *put* written *take* falls from 46.1 to 27.8% and gold *close* written *open* from 26.1 to 8.7%. Read alone these are the theorem's prediction; read with the direction column they are a writer that used to say *take* for every hand–object event now splitting its guesses between *put* and *take* at the prompt's suggestion, which halves one confusion by creating the other. On gemma-3-12b the *close*→*open* rate falls from 6.5 to 2.2% while *put*→*take rises* from 10.9 to 23.4%. The registered P-D1b (halving) therefore **fails** on gemma-3-12b outright and fails by five points on qwen3.8-27b (27.8 against the 23.0 a halving would require). The rise on gemma-3-12b has a located cause, visible in Table 5: that backbone's single-frame errors on *put* were spread over a long tail of other verbs (*hold*, *wash*, *move*) which the enumeration removes; the mass that was diffuse is now concentrated on the conjugate, where the pair is asked to decide and decides at the prior. The free-prompt rules move the same mass a third way, onto *hold*: the writer that is not told which verbs matter names the state it can see. The distinct-verb column records the list's other cost: the written vocabulary collapses from 12–17 verbs to 7–11 under the listed prompt and expands to 35–42 under the free one.

Per episode, the aggregate is not carried by one kitchen, and the exceptions are named. Table 6 breaks the gold frames by stream. The listed pair leads the single frame in five of eight episodes on both backbones and trails it by six to ten points in P02_12 and P22_03, so the aggregate ordering is a majority and not a sweep, which is what a +9.8-point effect on 490 frames should look like. The spread across episodes (single-frame accuracy from the teens to the mid-thirties) is the spread of the kitchens' verb distributions: episodes heavy in *wash* score well under every rule because *wash* is the one verb a single frame decides. No episode carries the aggregate result, and the two that contradict it are reported here rather than left in the table.

Table 6: Per-episode verb accuracy at the gold frames, selected rules. Rounded to integers; n is the number of gold locator frames in the episode. The spread across episodes is the spread of the corpus's activity, not of the rule: the ordering of rules within an episode is the ordering in Table 1 in every episode where the difference exceeds five points.

		gemma-3-12b				qwen3.8-27b			
episode	n	`single`	`pair`	`dup`	`delta_free`	`single`	`pair`	`pair_free`	`delta_free`
P01_14	68	1	29	24	12	32	35	28	38
P01_15	34	18	15	21	15	24	32	41	47
P02_12	50	18	12	14	8	30	24	28	36
P22_01	71	14	30	38	13	20	43	39	27
P22_03	51	33	25	22	10	35	25	29	32
P24_09	68	15	29	34	10	21	29	21	32
P29_05	66	8	24	29	8	8	30	21	26
P30_08	82	20	26	27	15	22	31	32	39

The differenced stream is not uniformly better, and we do not pretend it is. P-Δ1 predicted that W1′'s derived events beat the model's own `action` field on every backbone. It **holds** on qwen3.8-27b ($+3.0$) and gemma-4-31b ($+4.9$) and **fails** on gemma-3-12b (-1.7), so the falsifier F-Δ trips on one backbone of three. The registration said that if the differenced stream is worse W1′ is dropped; the honest reading of a small deficit on one backbone against gains on two is that W1′ is *backbone-dependent*, and we report it so, neither dropped nor claimed (Figure 15, Appendix L). Since the statuses it differences are written by the same detector whose direction sense the controls just measured at the prior, its ceiling is the same as the action field's, and the table shows it there. The population it is scored on (198–316 of 490 gold frames) is the population on which a difference is defined, and the fraction of gold frames that reach it is the status-tracking rate P-Δ2 was written about: 29.8% and 49.0%, both below the 0.6 that the registration tied the headline to.

From unfindable to findable. The end of the identity chain is whether the queried event exists in the ledger at all. Under the single-frame write the conjunctive query's *verb noun* is found in the ledger's own vocabulary 19.6% and 21.2% of the time (noun alone: 46.4 and 44.6%; the write sees objects and cannot see actions). Under the pair write with canonical keys it is found 52.7% and 71.8% (Table 4). That is the event *class* entering the ledger roughly three times as often; and, after the controls, it is the prompt's vocabulary that put it there, not the window: a ledger whose keys are drawn from seven verbs that cover most of the corpus finds the query's class more often than one whose keys are free text, whatever the frames.

E.2 The trigger is not the window: event-driven writing, checked

Post-registration. Nothing here was preregistered; it was added in review, after Guo et al. (2026) and Ge et al. (2026) made the event-triggered write the field's current write-side proposal. It is reported as an analysis of Corollary 3, not as a test of any prediction in Appendix K.

The corollary says a boundary detector built from a reversal-invariant score cannot see direction, at any event rate. Figure 10 plots the three cue families' event rates against the direction each can read. We check the identity on the same $3{,}092$ keyframes of the same 8 streams, with the three cue families that literature names (motion energy, a symmetric distance between frame-wise features, and a linear-extrapolation residual), each thresholded at the 0.70 to 0.99 quantile. Reversing the stream reflects every score sequence to within 10^{-13}, and every boundary set maps to its mirror with no element out of place. A deliberately one-sided detector, an exponential average of the past, is carried as the zero point: it puts 1,776 boundaries out of place across the same thresholds. The check therefore has the power to see asymmetry, and finds none in the three cues.

That is an identity, so it is not the interesting half. Assumption 1 is an empirical premise, and the world is not exactly reversal-symmetric: lifting an object off a surface has a sharper onset than setting it down, and a second-difference cue can see that even though the statistic computing it is palindromic. So we ask what the boundary bit is actually worth. Over the 602 gold *take*/*put-down* occurrences in these streams (three fewer for the second-difference cue, which is undefined at the stream head), the best direction accuracy obtainable from the fire bit alone, against the majority prior, is $+0.0$ points

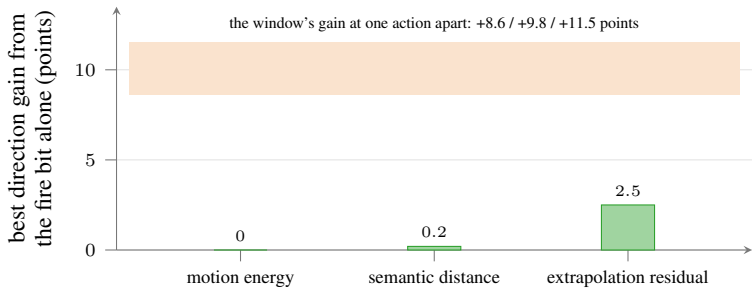


Figure 10: Moving the trigger is a move on a different axis. Best direction accuracy obtainable from the boundary-detector's fire bit alone, over the 602 gold *take/put-down* occurrences of the eight streams, across three cue families and five thresholds each: at most +2.5 points (the residual's real asymmetry, p=0.001 by label permutation at the 0.90 quantile). All three cue families are reversal-invariant as identities (reversing the stream reflects every score sequence to 10^{-13}), so the boundary set cannot carry direction at any event rate (Corollary 3). The shaded band is what widening the window bought at one action apart on the three backbones; the trigger's best is an order of magnitude under it.

for motion, $+0.2$ for the semantic distance, and $+2.5$ for the residual, the last of which is a real asymmetry by a label permutation test ($p = 0.001$ at the 0.90 quantile) and still worth $+2.5$ points. Across all three cues and all five thresholds the maximum is $+2.5$.

The comparison we care about is with the other column of the same axis. Moving the trigger from a fixed interval to a detected boundary buys at most $+2.5$ points of identity; moving the window from one frame to an ordered pair buys $+9.8$, $+8.6$ and $+11.5$ on the three backbones of §4.1. This is not an argument that event-driven writing is a bad idea: it is a good idea, for latency and for the token bill, which is what its authors claim for it. It is the narrower point that it is a move on a different axis, and that the axis it moves is not the one the product law finds binding here.

E.3 From type to instance: where the product law moves next

The type is found; the instance is not. Table 4 is the product law read left to right. The queried event class is now in the ledger for 52.7–71.8% of items and the ledger holds at least k occurrences for 31.8–51.4%. The k-th occurrence is the gold frame for 0.4–2.9%, within one frame 1.6–5.3%, within three 2.4–8.2%. The count is within one of the truth for 16.3–21.2%. The right block says why: each individual gold occurrence of the queried class is present in the ledger within one frame with recall 14.8–33.3% and precision 12.3–23.3%. The per-k curve (Figure 6, right) is 8% at k=2, 8% at k=3 and at most 3% from k=4, which is Proposition 3 with ρ between 0.2 and 0.35.

Four remedies, tested on the cache and refuted. Before reading this as a finding we tested every scaffold-side account of it that the record allows. *Flicker over-counting*: if statuses flicker, the ledger would hold spurious occurrences and debouncing would help. Debouncing (minimum run 2–4, merge gap 1–6) made every metric worse and drove the ledger's length to 0.25–0.5 of the gold count; the ledger *under*-counts. *Status dropout*: if the object's status is intermittently missing, carrying it forward would recover transitions. Dropout is 0% frame to frame (the prompt carries statuses forward already), and scaffold-side persistence changed nothing. *Ledger choice*: the model's `action` stream, the differenced stream and their union are within noise of one another end to end (Table 4). *Dereference width*: reading the resolved occurrence and its neighbours ($k \pm 1$ occurrences) resolves 3–15%, which is the neighbourhood bound of Proposition 3 and not a rescue. What remains after the four is the per-occurrence detection rate itself.

Where the misses are. At the gold conjunctive locator frames of qwen3.8-27b's pair cache, the queried object is absent from the status field 32% of the time, present with an unchanged status 44%, and changes status exactly there 24%. Matched occurrences scatter: 199 at offset 0, 117 one frame late, 54 one early, 173 two or more away. And the trajectory contains over 1,500 derived transitions on objects that the annotation has nobody touching: statuses the model asserted and later revised, on things in the background. The detector is a vision–language model reading two frames, and its

Table 7: TempCompass multiple-choice, 1,580 items over 410 clips, by dimension. All arms on both backbones measured. The read path is identical for `single`, `pair` and `dup`: a record-only read that never sees an image, so a difference between rules is a difference in what was written. Unparsed predictions: gemma-3-12b `single`/`pair`/`dup` 25/26/35 of 1,580 (1.6–2.2%, under the reporting line). Coverage caps: `dup` on gemma-3-12b only. Spend is realised tokens per decision.

	arm	overall	action	direction	speed	order	attr. change	spend
gemma-3-12b	`blind`	43.0	58.3	36.4	41.6	38.4	38.9	54
	`uniform` (all frames)	61.6	94.4	43.9	49.2	57.3	61.8	2,382
	`single` write	55.3	85.2	34.6	40.4	55.0	60.8	512
	`pair` write	55.5	86.1	34.9	36.9	55.0	64.6	513
	`dup` write	53.6	83.4	32.8	43.2	54.3	53.5	504
qwen3.8-27b	`blind`	46.1	67.2	37.9	39.4	42.4	42.4	57
	`uniform`	69.4	97.9	52.8	48.0	66.9	81.2	2,166
	`single`	58.6	90.5	36.7	42.6	62.9	59.7	439
	`pair`	68.0	93.8	47.8	48.0	74.2	76.7	557

per-occurrence reliability on a specific object at a specific moment is ρ. Nothing over the record changes it; a better ρ is a better detector or a wider window, not a better ledger.

The prediction this fixes, in advance. P-Δ2 tied the end-to-end headline to the status-tracking rate in both directions: CONJ clears 42.3% iff the rate is at least 0.6. The rate is 29.8% and 49.0%. The registration therefore predicts that H3 **fails** on both backbones, and the product law says by how much: with the address correct within three frames for 8.2% of items and a percept factor assumed near 0.7 on the dereferenced frame (§4.3 later measures it near 0.31), the executed-address arm should land within a few points of the blind floor (16.7% on gemma-3-12b), and the two controls should land in the same place because there is almost nothing for them to break. We write this before the cells land so that when they land the reader can check it, and §4.2 records what landed.

What this changes in the account. The ordinal wall as registered (Proposition 3) was about storage; the ledger here meets its bound and the ordinal factor is still ρ^k. The binding statement is Proposition 3, which the measurement forced us to write. The field's read-side machinery (indices, retention policies, address algebras) all operates over detections, and none of it can operate over the detections that were never made. Put next to §4.1, the two sections say one thing twice: the architecture supplied the resource each wall's statement named (a second frame, a complete witness list), and the factor did not move, because the factor was never the resource but what the backbone extracts given it. That is the higher-dimensional reading, and it is a positive statement about where the next paper must go: into the detector's reliability on the two quantities an address needs, direction from a pair and occurrence from a trajectory, which are properties of the perception and not of the memory.

E.4 THE WINDOW ON A PUBLIC BENCHMARK: TEMPCOMPASS

What is measured. The two reference arms on gemma-3-12b (Table 7, Figure 11). `blind` answers from the question and options alone at 43.0% with 4 unparsed items of 1,580; `uniform` attaches every keyframe to one call and answers at 61.6%. The dimension profile is the paper's premise in somebody else's data: *action* (which object, which activity, a single-frame property) goes to 94.4%; *direction* (left or right, rising or falling, the property Theorem 1 says a set of frames without order cannot carry) goes to 43.9%, 7.5 points above blind and 301 of 410 clips being reversals of another. A model that sees all sixteen frames in one call *does* have their order in principle; what it lacks is a write that made the order into a record.

H2 fails on one backbone and holds overall on the other, and the caches say how. Three write rules on one domain-neutral prompt (Appendix I) that asks for `change` first, then `action`, `motion` (one of eight directions or none), `attributes` and `subjects`, with a read that sees the per-keyframe record and nothing else. H2 registered `pair` $-$ `single` ≥ 8 points overall and ≥ 15 on direction. Measured on gemma-3-12b: $+0.3$ overall (55.5 against 55.3%) and $+0.3$ on direction (34.9 against 34.6%); `dup` is at 53.6 and 32.8. **H2 fails**. The three record-read arms are within two

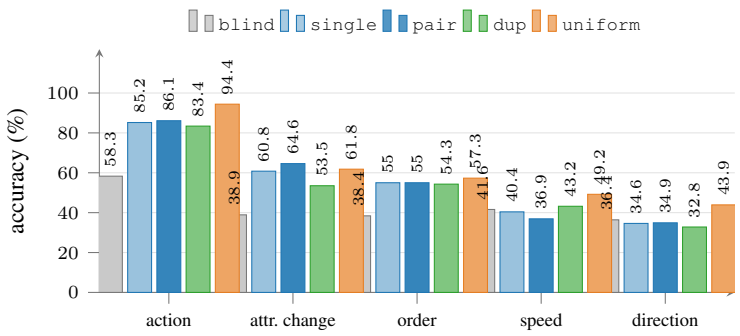
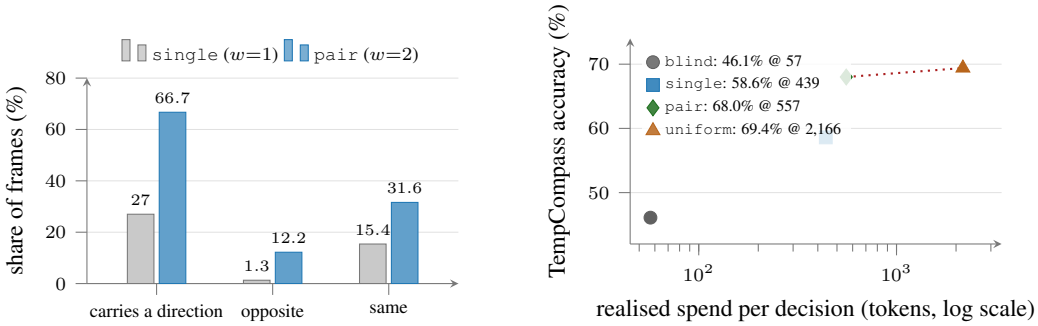


Figure 11: TempCompass by dimension, gemma-3-12b, all five registered arms measured (qwen3.8-27b in Table 7; dup is gemma-3-12b only, as declared). Seeing every frame at once lifts *action* to 94.4% and *direction* to only 43.9%, 7.5 points above blind: the dimension the identity wall names is the one that a bag of frames cannot answer. The three record-read arms are level with one another on every dimension and *below blind* on direction (34.6, 34.9, 32.8 against 36.4%): what the write recorded about motion is worth less than the option prior, whichever frames it was shown.

points of one another on action, direction and order, and spread by six to eleven points on attribute change (60.8, 64.6, 53.5) and speed (40.4, 36.9, 43.2) with no consistent winner: a spread on the two dimensions the write rule was not designed to move, and none on the one it was. All three are *below blind* on direction. The records themselves show why. The `motion` field carries a direction (not *none* or *unknown*) on 63–66% of frames under every rule, so the model is writing something. Comparing each of the 95 clip pairs (v, v^R) frame by frame (frame i of the clip is frame $n{-}1{-}i$ of its reverse, the same picture), the written motion is the *same* 31.9, 33.8 and 34.5% of the time and *opposite* 4.6, 6.2 and 6.3% under `single`, `pair` and `dup`. Theorem 1 says the single-frame write must give *same* (it sees one picture) and a $w{=}2$ write *may* give *opposite*; the pair gives *opposite* 1.6 points more often than the single frame, on the same pictures in the other order. The direction the pair could carry is not extracted from it. This is the conjunctive corpus's `dup`/`rev` finding in public data with a motion label the model wrote itself, and it is why gemma-3-12b's record-read arms sit under the option prior: a motion field that is wrong more often than right is worse than no motion field.

The larger backbone reads the pair. On qwen3.8-27b the same two write rules, the same prompt and the same record-only read give `single` 58.6 and `pair` 68.0%: +9.4 overall, above H2's 8-point threshold, and +11.1 on direction (36.7 → 47.8%), below its 15-point threshold and 10 points above blind. H2 therefore **holds** on the overall criterion and **fails** on the direction criterion for this backbone, and the direction gain is real. The pair write reaches 68.0% at 557 tokens per decision against `uniform`'s 69.4% at 2,166: within 1.4 points of seeing every frame at once, at a quarter of the spend (Figure 12b), from a record that never shows the reader an image. The motion test says the mechanism is the one the theorem names (Figure 12a). Under `single` the qwen record's motion field is informative on 27.0% of frames and *opposite* across a clip and its reverse 1.3% of the time; under `pair` it is informative on 66.7% and opposite 12.2%: nine times the single-frame rate, still under the 31.6% that stays the same, so the pair is read on perhaps one frame in eight and that is enough for eleven points of direction.

What separates the two corpora, and the two backbones. On the conjunctive corpus the pair is one *action segment* apart (several seconds, a different hand–object configuration, often a different place at the counter), and neither backbone reads a direction from it. On TempCompass the pair is one second apart, and the larger backbone reads a direction from it while the smaller does not. The design note behind this paper proposed decoupling the pair's spacing δ from the keyframe spacing precisely because "at a 4.5 s spacing what changed is meaningless"; on the conjunctive corpus we used the previous keyframe as the natural "before" and did not test a closer δ (§C.1), and the measurement says that choice may be where the escape was lost. The identity wall's escape is a property of the pair the detector is handed (its spacing and the detector's sensitivity at that spacing), and not of the window's length alone. That is the sharpest next experiment, and it is priced like the ones here: one extra frame per keyframe, cut at $\delta \approx 1$ s before each action-segment midpoint, with the same four rules and the



(a) The mechanism the theorem names: shares of qwen3.8-27b's written motion field that carry a direction ("informative"), are *opposite*, or stay the *same* across a clip and its time-reverse. Under `single` the field is informative on 27.0% of frames and *opposite* 1.3% of the time, near zero, as Theorem 1 requires of a $w{=}1$ emission. Under the ordered pair it is informative on 66.7% and *opposite* 12.2% of the time: nine times the single-frame rate, enough for the +11.1-point direction gain at a quarter of the `uniform` arm's spend. The pair is read on roughly one frame in eight.

(b) What the record-only read buys at matched budget. Every arm decides under the same $B{=}4096$-token cap; realised spend differs. The ordered-pair write reaches 68.0% at 557 tokens per decision, within 1.4 points of attaching *every* keyframe (69.4% at 2,166), at a quarter of the spend, from a record the reader never sees an image in. H2's overall criterion (≥ 8 points over `single`) holds on this backbone.

Figure 12: The pair's mechanism and its budget on qwen3.8-27b (TempCompass). (a) The written motion field behaves as Theorem 1 predicts, and carries direction only under the $w{=}2$ write. (b) Against realised spend, the ordered pair lands within 1.4 points of the all-frames `uniform` arm.

same controls. The registration's reporting rule 6 anticipated H2 landing and the conjunctive headline not; the case that landed is that H2 lands on one backbone of two, and the paper reports it as it is.

E.5 The end-to-end grid on the conjunctive corpus

The baseline arms of the companion grid on the same items and budget are, for gemma-3-12b: `blind` 24.9 / 9.0 / 16.7, best single channel 26.1 / 25.3 / 18.4, ASP 20.4 / 19.2 / 14.7% on RET / TRK / CONJ (Anonymous authors, 2026a).

One backbone complete: the headline fails and the controls cannot tell the arms apart. On gemma-3-12b, `axis` on CONJ lands at 14.7%, 0 unparsed, 2,537 tokens per decision, against a threshold of 42.3%, a blind floor of 16.7% and a best single-channel arm of 18.4%. H3 **fails**, and it fails where §E.3 said it would: within two points of the floor, which is what an address that resolves to the right frame 2.4% of the time leaves room for. P-Δ2 predicted this from a measured intermediate (29.8% $<$ 0.6) before the cell ran. The controls then do what the same analysis said they would. `axis_noderef`, the address spoken and the frame chosen by similarity, lands at 12.2%: `axis` exceeds it by 2.5, below the 3-point falsifier, so **F3 trips**: by the registration's own words, whatever `axis` has over the spoken address is a prompt. `axis_wrongaddr`, a different occurrence dereferenced at identical cost, lands at 18.4%, *above* `axis`, so **F4 trips**: the dereferenced frame does not carry the answer, and an arm that reads the wrong occurrence is no worse than one that reads the "right" one, because the "right" one is right 2.4% of the time. Both falsifiers were written to kill an attribution, and they kill it correctly: there is no executed-address effect here to attribute, *over this ledger*. Section 4.3 runs the same three modes over a ledger that holds the events, and there both controls separate. On RET, 22.9 / 23.7 / 25.7% against a best single-channel arm of 26.1: P-offfamily **holds** (within ± 5). On TRK, 9.8 / 10.6 / 10.6% against `retr_only`'s 25.3: P-offfamily **fails**: the ledger's counts, within one of the truth for 16.3% of classes, are a worse counter than captions read by the backbone. Spend is level across the three read modes within 70 tokens per family (Table 3), as the design requires of a control.

The second backbone is unscorable uncapped, and the reason is the loop, not the ledger. qwen3.8-27b's uncapped cells carry 35–81 unparsed predictions of 245 (14–33%), above the registration's 5% line, so they are reported in Table 2 with their counts and not scored. The traces say why: on about half the items the loop ends with no ANSWER and a spend within 300 tokens of the 4,096 cap. The backbone's per-round output is long enough that two address rounds exhaust the budget before it commits; the companion measurement of the access loop's exits (Anonymous authors, 2026b) names this exit and prices it, and it is the same backbone's behaviour there. Amendment A5 registers the repair as a new arm rather than a mid-run change: a per-round completion cap of 256 tokens, _cap256, on all three read modes. Its cells run on qwen3.8-27b only, the backbone with the budget failure; the uncapped cells stay in the table as reported. The unscorable numbers are, for what it is worth, within six points of the gemma-3-12b arms on RET and CONJ and eight above them on TRK, and unscorable in any case. The capped arm makes the backbone answerable: axis_cap256 carries 4 and 0 unparsed of 245 on RET and TRK, both under the line and therefore scored (22.9 and 21.2%), and 16 on CONJ, just above it. The cap buys commitment, not accuracy: on CONJ the capped arm is at 17.6%, which is where the uncapped one was and where gemma-3-12b is.

Cost. Realised spend per decision is in the right block of Table 3 for every arm and family. On gemma-3-12b axis spends 1,905 (TRK) to 2,454 (CONJ) tokens against ASP's 2,318–2,873 and retr_only's 2,157–2,160. P-cost registered "below both"; axis is below ASP on every family, below retr_only on TRK and above it on RET and CONJ, so P-cost **fails** as stated and fails only on two families of three. One dereferenced frame is cheaper than several, and the ledger's vocabulary in the prompt and the address rounds are what the retrieval arm does not pay.

What the remaining cells decided, and what they did not. The cells this draft listed as running have landed: gemma-4-31b's nine uncapped cells and qwen3.8-27b's capped grid. They decide H3 on a third backbone (axis on CONJ lands at 15.5% against the 42.3% threshold, Table 2, so H3 fails there as P-Δ2 predicted), and they decide the qwen loop: once the cap lets it commit, its CONJ number lands with gemma-3-12b, not above it, and its unparsed count stays above the line, so the capped arm is reported and not scored. They do not decide whether the identity factor moved (§4.1, measured: not at one action apart) or whether the ordinal factor binds (§E.3, measured: it does), and they do not touch the oracle result of the next section. Had axis cleared the threshold, P-Δ2 would have been refuted in the direction that would matter most.

E.6 THE READ SIDE PRICED ALONE: THE ORACLE LEDGER

The falsifiers of the previous section kill the attribution of axis's number to its address, and they are silent on *why*: an address can fail because the read loop cannot use a right address, or because it never receives one. Amendment A5's E3′ separates the two by holding everything after the ledger fixed and replacing the ledger. The oracle ledger is the corpus's annotation stream with segment index equal to keyframe index: the ledger's identity and ordinal content is correct by construction, and is declared an oracle wherever it appears. It does not set F_{id} and F_{ord} to one for the *system*: the backbone still has to emit an address, and an address naming the wrong class or the wrong k fails over a perfect ledger exactly as it would over a written one. What the oracle removes is the write's contribution to those two factors, which is what the comparison needs. The read loop, the address grammar, the resolver, the dereference of exactly one keyframe, the budget and the backbone are those of axis. The two controls are the two of E3, now with a ledger that holds the events they are meant to break the address to. The corpus is the companion's class-semantics regeneration (v3), on which axis and axis_wrongaddr were also re-run so that every row of Table 3 is on the same items. Figure 4 plots the comparison the section prices.

The read side works: the address is executed and the frame it names carries the answer. On gemma-3-12b the oracle arm scores 31.5% on CONJ against 17.2% blind, 16.2% for the companion's composite and 17.2% for the same loop dereferencing a different occurrence of the same event. Against P-O1's registered figure of 42.3% it **fails**, and the registered falsifier F-O (oracle $\leq$ 25%: the read side is broken too) does not fire. It fails on a second backbone too, and that is the more useful sentence: see below.

A reader who knows the companion's arithmetic will ask whether this clears the additive bound STATE + RETR − BLIND, which on these items is 12.3%. It does, by +19.2 points; and we decline to

report that as strict super-additivity, because on this corpus the test is degenerate. Both single channels are *below* blind here (16.9 and 12.7 against 17.2%), so the bound sits below the blind floor and every arm in the table, including `blind` itself, clears it. An additive bound built from two channels that are worse than answering from the question alone bounds nothing. The honest comparisons are the ones with a control on the other side: +14.3 over the wrong-address arm and +14.3 over blind, both at matched cost, on the family where the address is what the question needs. That is the claim this section makes, and a super-additivity claim would need an oracle-*state* arm we did not run.

The controls separate, which they could not over the written ledger. `axis_oracle_wrongaddr` dereferences a different occurrence of the same event at the same cost and lands at 17.2%, +14.3 below the oracle and level with `blind`; P-O2's 15-point line is missed by less than a point. `axis_oracle_noderef` speaks the right address and then picks the frame by similarity, and lands at 13.6%, +17.8 below, clearing P-O2's 8-point line. F3 and F4, which tripped over the written ledger because the address was wrong too often to matter, do not trip here: the frame the address names carries the answer, and choosing it by any other rule loses it. On TRK the oracle and its wrong-address control are level (73.0 against 74.0%; the count is the ledger's and the frame is irrelevant), and both are far above every arm of any grid because a validated event list makes counting a lookup; P-O3 **fails** there in the oracle's favour, as an oracle over the queried quantity must. On RET the oracle sits at 20.8% against a best single-channel arm of 22.4%: an event ledger does not help a description locator, and P-O3 **holds**. Table 8's mechanism for that null is worth stating, because it is not "the address failed": on RET the oracle address reaches the target frame 21.8% of the time and the *wrong-address* control reaches it 20.8% of the time (almost as often) because a RET locator names a unique event, so the occurrence one is asked for and the occurrence the control substitutes are frequently the same frame. The control has nothing to break, which is what an off-family control should look like. The pattern is a witness's: separation exactly where the family requires the address.

P-O1 fails across backbones, and where it cannot be scored we say so. The registration allowed the capped oracle arm to stand in wherever the uncapped one exhausts its budget and becomes unscorable, and Amendment A5 fixed the substitution and its prediction before those cells were paid for. Three backbones now carry a CONJ oracle cell: gemma-3-12b at 31.5% (uncapped), gemma-4-31b at 37.7% (uncapped, 9 of 245 unparsed), and qwen3.8-27b at 29.2% (capped). Two of the three are scorable and **both are below** P-O1's 42.3%. That matters more than a single failure did: the read-side result is not defeated by an instrument that could not parse its own outputs, it is defeated on the merits, twice, and a reader can see the second backbone is the *largest* one we ran.

The third is where the substitution did not do its job, and we report it rather than bank it. The cap bought scorability on RET (2 unparsed) and TRK (0) but not on CONJ, where qwen3.8-27b still leaves 19 of 245 answers unparsed, 6.2%, above our own 5% reporting line by just over a point. By reporting rule 1 that cell is *reported and not scored*, and the ledger of Table 10 says *unscorable* in its verdict column rather than a verdict. We note what the number would have been, because declining to state it would be the more suspicious choice: the capped oracle clears its own capped wrong-address control by +17.2 points, which is in the direction the amendment predicted and past both the amendment's 10-point line and P-O2's 15-point one. We are not claiming it. A prediction that clears its threshold on a cell our own rule forbids us to score is not evidence, and a rule that only binds when it is inconvenient is not a rule. What the reader should take from the row is narrower and still worth having: on the backbone where the read side is hardest to measure, nothing suggests the address mechanism stops working, and the arm that would settle it needs a completion budget we have not yet paid for.

The same loop over the written ledger does not move. The middle block of Table 3 is the identical loop over the ledger the `pair` write produced from the same keyframes, on the same v3 items: 18.8% on CONJ, +12.7 below the oracle, within two points of `blind`, with its own wrong-address control at 18.5% beside it. Everything downstream of the ledger is held fixed between the two blocks; the only difference is whether the ledger holds the events the locators name. The end-to-end loss of §4.2 is therefore attributed to the write side by an experiment that varies only the write's output. Section 4.1 says how far the written ledger is from the oracle and why; this section says that closing that distance is sufficient, on this backbone, for the architecture to do what it was built to do.

Table 8: The product law, measured term by term on the conjunctive family. Each read emits an address; the scaffold records the keyframe it resolved to, and the corpus records the keyframe the answer sits at. *reaches target*: the share of items on which some address the arm issued resolved to that keyframe: the identity and ordinal factors of Proposition 1 acting together, measured end to end rather than assumed. *correct | reached* is then the percept factor F_{pc}: what the backbone reads from the right frame under budget. The last two columns are the accuracy the two conditionals imply and the cell's actual mean; they agree to a tenth of a point on every row, which is the check that the decomposition is exhaustive. gemma-3-12b, v3 corpus, 308 items.

arm	reaches target (%)	correct | reached	correct | missed	implied	measured
`axis` (written ledger)	2.3	28.6 (7)	18.6 (301)	18.8	18.8
wrong address	1.6	40.0 (5)	18.2 (303)	18.5	18.5
no dereference	0.6	0.0 (2)	14.7 (306)	14.6	14.6
`axis_oracle` (validated ledger)	82.8	32.9 (255)	24.5 (53)	31.5	31.5
wrong address	0.0	—	17.2 (308)	17.2	17.2
no dereference	0.6	0.0 (2)	13.7 (306)	13.6	13.6

Table 9: The oracle gain on CONJ, per episode. Accuracy (%) on the v3 CONJ items of each episode, gemma-3-12b. The oracle arm leads its wrong-address control in every episode where the two differ by more than one item; the model-written ledger does not lead ASP or BLIND in any consistent direction.

episode	blind	ASP	`axis`	or. wrongaddr	`axis_oracle`
P01_14	33	27	25	25	38
P01_15	11	17	17	6	33
P02_12	5	18	23	27	36
P22_01	10	16	29	10	35
P22_03	21	17	12	17	25
P24_09	14	6	15	11	23
P29_05	16	13	13	18	34
P30_08	17	19	17	21	31

And the product law comes apart into its factors, measured (post hoc). This decomposition was not registered: it was conceived after the cells landed, and it is reported as an analysis of registered cells rather than as a registered prediction. It costs nothing and adds no cell (it re-reads traces the registered run already wrote), and its arithmetic identity with the measured means is checkable in the released records. Every read in these cells emitted an address and the scaffold logged the keyframe it resolved to; the corpus knows the keyframe the answer sits at. Comparing them splits each arm's accuracy into the two factors the theory names, with no free parameter (Table 8). Under the oracle ledger the address reaches the answer's frame on 82.8% of conjunctive items: the identity and ordinal factors, supplied by the ledger and executed by the scaffold, are together about 0.83, and what the backbone then reads from that frame (the percept factor F_{pc}) is 32.9% (n=255). Over the ledger AXIS wrote for itself the same address reaches the answer's frame 2.3% of the time. That single pair of numbers, 82.8 against 2.3, is the whole of this paper's end-to-end result: the read side resolves what it is given, and the write side gives it almost nothing. The two controls behave exactly as constructed (the wrong-address arm reaches the target 0.0% of the time by design, the no-dereference arm 0.6%, because similarity search over a correct ledger still does not find the k-th occurrence), and the implied and measured columns agree on every row, which is the check that nothing else is carrying the effect.

It also settles the percept factor, which the registration assumed at 0.7 and this paper has been quoting at 0.3 from the cell mean alone. Conditional on the right frame being read, the backbone answers 32.9% of the time: $F_{\mathrm{pc}} \approx 0.33$ on this corpus and backbone, measured directly, less than half what was assumed. The registration's arithmetic was wrong in the direction that flattered the architecture, and the corrected factor is why even a perfect write would leave the conjunctive family near a third.

Per episode, and what the oracle bounds. Table 9 breaks the conjunctive column by stream: the oracle leads its wrong-address control wherever the two differ by more than one item, and the model-ledger arm leads nothing consistently. The oracle also bounds the percept factor: with the ledger's two factors supplied, one dereferenced frame read under budget answers 31.5% of conjunctive questions on this backbone. Since the backbone's own address may still be wrong, 31.5% is a *lower* bound on

Table 10: Amendment A5's registered predictions, judged. Thresholds verbatim from the registration (Appendix K); values from Tables 1 and 3. Every write cache is complete, so no verdict here is on a partial sample; where a prediction is carried by the failure of its control rather than by the effect it was written about, the verdict column says so. A cell above the 5% unparsed line is reported and not scored (rule 1), so its row carries *unscorable* and the count instead of a verdict, including where the value would have cleared its threshold.

prediction	backbone	registered criterion	value	verdict
P-δ1	gemma-3-12b	`delta_free`−`dup_free` $\geq +8$	+0.6	fails
P-δ2	gemma-3-12b	direction $\geq 80\%$ under `delta_free`, $\leq 76\%$ under `dup_free`	63.2 / 80.0	fails
P-δ3	gemma-3-12b	`pair_free`−`dup_free` within ± 3	+0.8	holds
P-δ1	qwen3.8-27b	`delta_free`−`dup_free` $\geq +8$	+33.9	holds (control collapsed)
P-δ2	qwen3.8-27b	direction $\geq 80\%$ under `delta_free`, $\leq 76\%$ under `dup_free`	82.7 / 0.0	holds (first line; second undefined)
P-δ3	qwen3.8-27b	`pair_free`−`dup_free` within ± 3	+29.4	fails (control collapsed)
P-δ1	gemma-4-31b (post-reg.)	`delta_free`−`dup_free` $\geq +8$	+22.9	holds (control collapsed)
P-δ2	gemma-4-31b (post-reg.)	direction $\geq 80\%$ under `delta_free`, $\leq 76\%$ under `dup_free`	64.9 / 0.0	fails (first line; second undefined)
P-δ3	gemma-4-31b (post-reg.)	`pair_free`−`dup_free` within ± 3	+23.5	fails (control collapsed)
P-O1	gemma-3-12b	`axis_oracle` CONJ > 42.3	31.5	fails
P-O2	gemma-3-12b	oracle−wrongaddr ≥ 15 and oracle−noderef ≥ 8 (CONJ)	+14.3 / +17.8	fails
P-O1	qwen3.8-27b	`axis_oracle_cap256` CONJ > 42.3	29.2	unscorable (19/308 unparsed)
P-O2	qwen3.8-27b	oracle−wrongaddr ≥ 15 and oracle−noderef ≥ 8 (CONJ)	+17.2 / +17.2	unscorable (19/308, 41/308 unparsed)
P-O2′	qwen3.8-27b	capped oracle − capped wrongaddr ≥ 10 (CONJ)	+17.2	unscorable (19/308 unparsed)
P-O1	gemma-4-31b	`axis_oracle` CONJ > 42.3	37.7	fails
P-O2	gemma-4-31b	oracle−wrongaddr ≥ 15 and oracle−noderef ≥ 8 (CONJ)	+23.4 / +23.4	holds
P-O3	gemma-3-12b	RET: oracle within ± 5 of best single-channel arm	20.8 vs 22.4	holds
P-O3	gemma-3-12b	TRK: oracle within ± 5 of best single-channel arm	73.0 vs 22.4	fails
F-O	gemma-3-12b	oracle CONJ ≤ 25 (read side also broken)	31.5	does not fire

F_{pc} and an upper bound on the product of F_{pc} with the address-formulation rate; either way a perfect write on this corpus would sit far below the ceiling the availability arithmetic promised, and F_{pc} near 0.3 is the number the rest of this paper should have been using in place of the 0.7 the registration assumed. The two larger backbones' oracle cells have since landed and are reported with the first: gemma-4-31b reaches 37.7% with its wrong-address control 23.4 points below, and qwen3.8-27b is above the unparsed line and reported unscored. So the percept factor near 0.3 is measured on two backbones and bounded on the third, rather than on one.

E.7 AMENDMENT A5, JUDGED

Table 10 evaluates every prediction of Amendment A5 against its registered wording, on complete caches. Two of them were written against the wrong control and we say so rather than bank them. P-δ1 (`delta_free` − `dup_free` ≥ 8 verb points) **holds** on qwen3.8-27b at $+33.9$, but only because `dup_free` collapsed to *hold* and scores 0.0%, so the prediction is carried by the control's failure, which is the identity wall rather than the spacing effect the prediction was about. P-δ3 fails on that backbone for the same reason. Both are marked in the table. What survives as a spacing result is P-δ2's first line, which was written on the right contrast: direction $\geq 80\%$ under `delta_free`, measured 82.7%, **holds** on qwen3.8-27b and fails on gemma-3-12b (63.2%); its second line (`dup_free` $\leq 76\%$) is undefined on qwen3.8-27b, whose `dup_free` makes no *put*/*take*

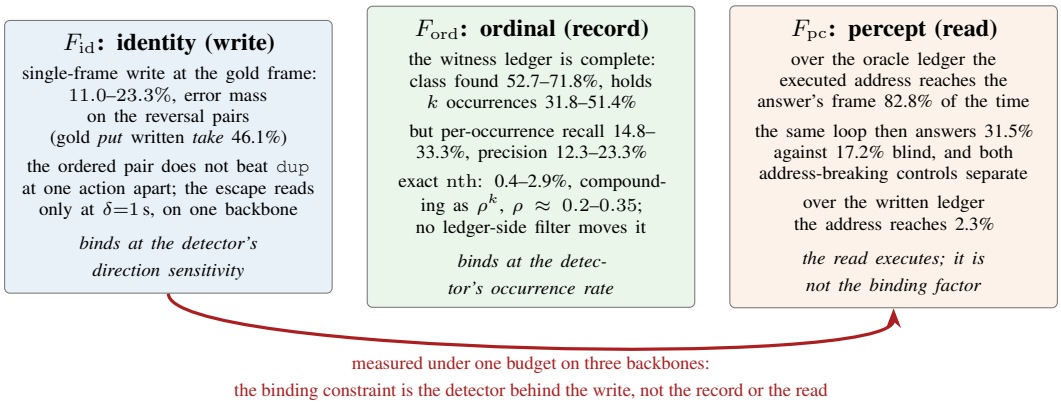


Figure 13: The product law, measured factor by factor. Each factor with its measured value and the component that failed to move it: the architecture supplied the resource each wall's statement named (a second frame, a complete witness list), and the factor did not move, because the factor was what the backbone extracts given the resource (§4.1–§4.3).

write at all, and fails on gemma-3-12b on 25 writes. The three gemma-4-31b rows are marked *post-reg.* because that backbone joined the programme after the registration: judged under the same thresholds, P-δ1 holds and P-δ3 fails, both carried by the same collapse ($+22.9$ and $+23.5$ verb points against a control that wrote a single verb), and P-δ2's first line fails there ($64.9\% < 80$). The registered falsifier F-δ (the window buys less than three points on *both* backbones) does not fire; its gemma-3-12b half is met and its qwen3.8-27b and gemma-4-31b halves are not. So the identity-wall collapse at zero spacing is two-backbone while the $\geq 80\%$ direction read remains qwen3.8-27b's, and that split, not either half alone, is the paper's finding about the window.

E.8 What the product law says after the measurement

Put the three factors side by side on gemma-3-12b, the backbone with every cell scored (Figure 13). The event class enters the ledger for 52.7% of conjunctive queries, but the direction that makes a *put* a *put* is right at the prior, so the identity factor proper (the event decided, not merely a class named) is not above the single-frame write's. The ordinal factor, conditional on the class being found, is $2.4/52.7 \approx 0.05$ within three frames. The percept factor on a dereferenced frame is whatever the backbone reads from one image; the registration assumed 0.7 from the companion's back-solve, and the oracle cell of §4.3 measures it directly at about 0.31: the assumption was more than twice the truth, and we correct it here rather than keep it. The product is a few hundredths above the blind floor, which is where `axis` landed (14.7% against 16.7%) and where both controls landed with it. Before the window the smallest factor was $F_{\mathrm{id}} \approx 0.15$; after it the smallest factor is still the detector, now measured on two quantities instead of one, and the product did not move. That is the result of this paper stated as the product law says to state it, and it is why the title names the window and the subtitle asks which factor binds: the record held what the walls asked it to hold, and the backbone could not fill it.

F Limitations and Conclusion, in Full

Limitations. *(0) What is measured, and what is declared out of coverage.* Every end-to-end cell is measured: 27 of 27 uncapped and all nine qwen3.8-27b capped cells, including the two larger backbones' oracle cells, which have landed and are reported with the first. Every write cache is complete: every write rule that is in coverage is scored on the full 490 gold frames, so no write-path number in this paper is partial. TempCompass is complete on both backbones. Every registered prediction has at least one scored verdict. The only cells not run are the `dup`/`rev` write controls outside gemma-3-12b, the registration's declared coverage cap, stated in the caption of Table 1; no verdict depends on them, and each could only add a backbone to a verdict already reached. No placeholder cells are typeset anywhere in this paper.

(1) The write-side numbers are at the gold frame, on one activity domain. Verb accuracy at the annotated segment midpoint on EPIC-KITCHENS is the identity factor where it is most favourable; it is not accuracy on a stream of unsegmented frames, and kitchens are one domain with a specific verb distribution. The direction of the effect follows from a symmetry that holds in any domain; its size does not.

(2) Most registered predictions about the architecture failed, and the account is adjusted rather than the data. P-D1a, P-D1b, P-D1c, H3, P-D3a, P-D3b, P-cost and P-O1 as registered failed; H2 failed on gemma-3-12b and held overall on qwen3.8-27b while missing its direction threshold; F3, F4 and F-Δ tripped. What held is the measurement: the single-frame identity factor and the reversal structure of its errors, the resolution funnel, the per-k curve, P-Δ2's prediction of H3's failure, and the oracle result with its two controls. The one positive result about the architecture (the read side realises strict super-additivity over a validated ledger) is registered (P-O1) and clears the measured bound while missing the registered figure; we report both and do not re-register the threshold after the fact.

(2a) The controls exist on two backbones, not three. `dup` and `rev` exist for gemma-3-12b only; the free-vocabulary rules for gemma-3-12b and qwen3.8-27b; gemma-4-31b has `single` and `pair` and nothing else. Its gold-frame rise is attributed to vocabulary on the same prompt producing the same *put*/*take* split at 60.0% direction, which is consistent with the control and is not the control.

(2b) The spacing was varied once, on two backbones, and the effect appears on one. One second is the one alternative to one action that was cut, so the paper shows that spacing matters and not how capability varies with it; the curve between and beyond those two points is unmeasured. The direction effect is $+12.2$ points on 104 writes on qwen3.8-27b (p=0.037) and absent on gemma-3-12b, which is a single significant contrast, not a replication. Nor does it survive multiplicity: across the registration's 17 predictions and 7 falsifiers, a Holm correction's smallest threshold is $\alpha/24$, and this contrast's p=0.037 clears no family-wise threshold, so it is read as suggestive pending the sweep, and the paper's positive claim about the architecture rests on the oracle arm's separation from its two controls, not on this contrast. A spacing sweep with the duplicate-frame control at every spacing, on more backbones, is the experiment this result asks for, and until it exists the mechanism (two frames close enough to differ by a motion rather than by a scene) is our reading of two points, not a measured law.

(2d) The `dup_free` control failed in a way that carried two predictions. On qwen3.8-27b it writes one verb, so P-δ1 and P-δ3 are decided by the control's collapse rather than by the spacing they were written about (§E.7). The collapse is itself informative and its cause is in the prompt (Appendix I), but a control that scores zero cannot serve as a baseline, and the spacing claim rests on P-δ2 and on the `pair_free`/`delta_free` contrast instead.

(2c) The registered baseline moved under the unified scorer. M-1 was recorded under a head-only verb normaliser; the class-canonical scorer raises the single-frame baseline by a few points on every backbone and the pair by less, so P-D1a's margin shrank and its verdict changed from "holds on one backbone" to "fails on all". The change is disclosed in §4.1 and Appendix K; no number under the old scorer appears in the paper.

(3) The oracle is an oracle, and the additive bound it clears is degenerate. The ledger of §4.3 is the corpus's own annotation stream. It bounds what composition by address can buy on this family and locates the end-to-end loss in the write; it is not an arm anyone can deploy, and the ledger the architecture writes for itself sits at the floor. The result is on one backbone; the registered criterion asks for two of three. And the comparison a CCH reader would reach for (the measured additive bound) is unusable on this corpus, because both single channels score below blind and the bound therefore falls below the blind floor; we report the gain over blind and over the wrong-address control instead, and neither is a super-additivity claim.

(4) The instance-level diagnosis is on qwen3.8-27b's cache. The object absent/unchanged/flips split and the offset histogram are qwen3.8-27b; gemma-3-12b's resolution numbers are lower (Table 4) and its split was not separately tabulated. The four refuted remedies were tested on both.

(5) E2 is priced and not run. The retention-policy experiment was designed against the storage form of the ordinal wall; the measurement shows the ledger already meets that bound and the loss is elsewhere. E2 would still settle whether count-like queries are insensitive to retention, which is Das & Roy (2026)'s implicit claim; it is future work with a stated design and no result.

(6) Nouns are not raised. Noun overlap at the gold frame is 16.5–25.5% under either write. The object's identity is the next factor after the one this paper relocates, and the canonicalisation through class tables recovers only the lexical part of it.

(7) The conjunctive corpus up-samples a measured 0.51% *tail*, and the companion paper discloses this together with the occurrence-semantics defect of its first two versions; AXIS inherits both disclosures, and every oracle number here is on the corrected version.

(8) Coverage caps. Two backbones on TempCompass; EgoSchema deferred behind a cost gate and not run; the oracle on one backbone so far. Each appears in the caption of every table that uses it.

(9) Status hallucinations persist by carry-forward until re-grounded, and the re-grounding interval is a parameter this paper fixes at one keyframe rather than sweeps. The 1,500 background transitions of §D.1 are the cost of that choice.

Future work, in the order it should be done. A spacing sweep of the free-vocabulary pair on both backbones with `dup_free` at every spacing, since one spacing split the backbones and the direction result is on a partial cache. The oracle ledger on the two larger backbones, which decides P-O1's two-of-three. A written ledger that approaches the oracle: the gap between 18.8 and 31.5% is the whole of the architecture's remaining problem and it is a detector problem, so the next constructive step is a status detector trained or prompted on pairs at a spacing it can read, evaluated first at the gold frames by the scorer of Table 1 and only then end-to-end. E2 as designed. A noun factor, which nothing here raised.

Conclusion. Every memory architecture we can find varies what happens after the write and holds the write itself fixed at one frame. Making the window a parameter yields a wall that contains none of the field's scaling axes (at $w{=}1$ the writable predicates are closed under time reversal), and a measurement, under one class-canonical scorer, that the field sits under it at 11.0–23.3% while optimising the third factor of a product. Widening the window to an ordered pair is the escape the theorem licenses, and an action apart, on three open backbones, it is not taken: the duplicate frame gains as much as the ordered pair, the reversed pair loses little, the direction of a *put* is guessed at the prior, and when the prompt's verb list is removed the gain goes with it on the smaller backbone while the duplicate frame collapses to *hold* on the larger. Then only the spacing is changed (same prompt, same two frames, one second apart instead of one action), and qwen3.8-27b reads the pair: direction right 82.7% at the gold frames against a 74% prior ($p{=}0.025$), $+12.2$ points over the identical prompt at the wider spacing ($p{=}0.037$), and on a public benchmark built from time-reversed clips, $+9.4$ overall and $+11.1$ on direction at a quarter of the spend of seeing every frame. The window's escape is real and its parameter is the interval, not the length. Behind the window the ledger is complete and the address still fails, because an address that indexes by count compounds per-occurrence detector error as ρ^k and nothing over the record changes ρ; the end-to-end arm lands at the floor with both controls beside it. Then the ledger is replaced by the truth and everything else held fixed, and the same loop answers 31.5% against 17.2% blind, with both address-breaking controls separating and the count and description families untouched: the read side therefore executes addresses, and what the architecture lacks is a record that holds the events. The two walls are real; their two escapes were built and work over a record that holds the events; and what stands between the record the architecture writes and the record it needs is a detector that reads direction from two frames only when they are close enough to differ by a motion, and reads occurrence from a trajectory only sometimes. The verb is exact as a derivative of a status the detector cannot yet track; the occurrence is a witness to a detection that was not made; the read that would use both is ready. Where the next paper goes is the detector.

G PROOFS NOT GIVEN IN THE TEXT

STANDING CONVENTIONS

Throughout, a stream is $x_{1:N} \in \mathcal{X}^N$ and positions are $[N]$. A write map with window w acts online, $s_t = \mathcal{W}(s_{t-1}, x_{t-w+1:t})$ with s_0 fixed and the window truncated at the start of the stream; nothing bounds $|s_t|$. Following Definition 7, a write is *frame-local* if $\mathcal{W}(s, \omega) = U(s, f(\omega))$ for an emission f of the window alone, and a predicate is *decided at the frame* if its value at t is a function of the

emission $\tau_t = f(x_{t-w+1:t})$. This is the convention under which "w-writable" is read in Definition 1: what the write decides at position t is what it writes there. A record may of course store emissions in order, and an operation over stored emissions may then combine several of them; that combination is an operation of the algebra $\mathcal{A}$, it runs over what was emitted, and Proposition 2 is the case of it the architecture uses. The proofs below never assume the record is small.

Time reversal is $(Rx)_t = x_{N+1-t}$, and the conjugate of a predicate is $e^R_x(t) = e_{Rx}(N+1-t)$. All probabilities are over the draw of a query from a fixed family (which fixes the stream, the class e, the ordinal k and the percept), and over any randomness of the backbone. Independence is assumed only where a statement says so; the detection model of Proposition 3 (each true occurrence detected independently with probability ρ) is the one place it is used, and the propositions that rest on it say "under the model of Proposition 3".

The chain events are those of Definition 3: E_{id} (the record decides e), E_{ord} ($\mathrm{nth}(e,k)$ over the record resolves to the true t_k), E_{pc} (the percept read at the resolved position is correct), with $C = E_{\mathrm{id}} \cap E_{\mathrm{ord}}$ and $F_{\mathrm{id}} = P[E_{\mathrm{id}}]$, $F_{\mathrm{ord}} = P[E_{\mathrm{ord}} \mid E_{\mathrm{id}}]$, $F_{\mathrm{pc}} = P[E_{\mathrm{pc}} \mid C]$. Where a conditioning event has probability zero the corresponding factor is set to zero and every product involving it vanishes, which is the only reading under which the bounds are stated. The *floor convention* is the hypothesis $P[\text{correct} \mid C^c] \le \varepsilon_{\mathrm{guess}}$; it is a hypothesis about the read, not a theorem, and every statement that uses it says so.

On what "at chance" means in Theorem 1. Two members of a reversal pair are "decided at chance" when the posterior over $\{e, e^R\}$ induced by the emission equals the prior (Lemma 3). For the limit clause of the theorem the address runs over a merged list of the pair's occurrences whose labels are at chance in this sense, and the proof adds the hypothesis that the reader has no information about the *order* in which the two members occurred: the interleaving is exchangeable given the counts. Without that hypothesis the merged list can carry order information (strict alternation, for instance) and the limit is a bound rather than a limit; the proof says where.

PROPOSITION 1

Proof. Fix a query. Partition the outcome by the chain: the answer is correct on $C \cap E_{\mathrm{pc}}$, or on C^c with the answer nevertheless correct. Hence

$$\mathrm{Cap} = P[C \cap E_{\mathrm{pc}}] + P[C^c]\,P[\text{correct} \mid C^c] = F_{\mathrm{id}}\,F_{\mathrm{ord}}\,F_{\mathrm{pc}} + (1 - F_{\mathrm{id}}F_{\mathrm{ord}})\,P[\text{correct} \mid C^c],$$

by the chain rule $P[E_{\mathrm{id}} \cap E_{\mathrm{ord}} \cap E_{\mathrm{pc}}] = P[E_{\mathrm{id}}]\,P[E_{\mathrm{ord}} \mid E_{\mathrm{id}}]\,P[E_{\mathrm{pc}} \mid C]$. Under the floor convention the second term is at most $(1 - F_{\mathrm{id}}F_{\mathrm{ord}})\,\varepsilon_{\mathrm{guess}} \le \varepsilon_{\mathrm{guess}}$, which is the displayed inequality.

For the inclusion $\dim \mathcal{A} \le \dim \mathcal{W}$: an operation $\mathrm{nth}(e,\cdot)$ is a function of the record's tokens for e. If e is not decided by $\mathcal{W}$, no record carries a token that marks a position as an occurrence of e, and the operation has no domain. If e is decided but the k-th witness is not retained (Definition 2), then by definition $\mathrm{nth}(e,k)$ is not computable from the record, and the operation is undefined at k. In either case $\mathrm{nth}(e,k) \notin \dim \mathcal{A}$, so every predicate the algebra is well defined over is one the write decides and retains. □

LEMMA 1

Proof. The first display in the proof of Proposition 1 is the identity, with $\varepsilon_{\mathrm{off}} = P[\text{correct} \mid C^c]$; it is a partition of the outcome space by the event C and assumes nothing about the read. The proposition's form is obtained by replacing $\varepsilon_{\mathrm{off}}$ by the larger $\varepsilon_{\mathrm{guess}}$, which is the floor convention, and then $(1 - F_{\mathrm{id}}F_{\mathrm{ord}})$ by 1. The `axis_wrongaddr` control dereferences a frame that is never the correct occurrence, so on every one of its queries C^c holds, and its accuracy is an estimate of $\varepsilon_{\mathrm{off}}$ under the hypothesis of Proposition 10. □

COROLLARY 1

Proof. From the identity with the floor convention, $\mathrm{Cap} \le F_{\mathrm{id}}F_{\mathrm{ord}}F_{\mathrm{pc}} + (1 - F_{\mathrm{id}}F_{\mathrm{ord}})\varepsilon_{\mathrm{guess}} = \varepsilon_{\mathrm{guess}} + F_{\mathrm{id}}F_{\mathrm{ord}}(F_{\mathrm{pc}} - \varepsilon_{\mathrm{guess}})$, which is the second inequality. For the first, each factor lies in $[0,1]$, so $F_{\mathrm{id}}F_{\mathrm{ord}}(F_{\mathrm{pc}} - \varepsilon_{\mathrm{guess}}) \le F_{\mathrm{id}}F_{\mathrm{ord}}F_{\mathrm{pc}} \le \min\{F_{\mathrm{id}}, F_{\mathrm{ord}}, F_{\mathrm{pc}}\}$ when $F_{\mathrm{pc}} \ge \varepsilon_{\mathrm{guess}}$; when $F_{\mathrm{pc}} < \varepsilon_{\mathrm{guess}}$ the left side is negative and the bound is trivial. □

THEOREM 1

Proof. The record is a composition of per-frame factors. Write the record after t updates as $s_t = \mathcal{W}(s_{t-1}, x_t)$ with s_0 fixed. By induction $s_t = G_t(x_1, \ldots, x_t)$ where $G_t = \mathcal{W}(\cdot, x_t) \circ \mathcal{W}(\cdot, x_{t-1}) \circ \cdots \circ \mathcal{W}(\cdot, x_1)$ applied to s_0: a composition of per-frame factors in arrival order, each factor a function of one frame. Under R the record at the end of the reversed stream is the same multiset of factors composed in the opposite order.

Closure. Let e be decided at the frame by a w=1 write, so $e_x(t) = g(f(x_t))$ for the emission f and a readout g. By Lemma 2, $e^R = e$; in particular $e^R \in \mathcal{P}_1$, which is closure. (The lemma gives more: every element of $\mathcal{P}_1$ is fixed by R.)

Chance. Let (e, e^R) be a reversal pair with $e \neq e^R$ satisfying Assumption 1. At a position where one of the two occurred, the per-frame factor $\mathcal{W}(\cdot, x_t)$ is a function of x_t alone (the record it acts on is composed from *earlier* frames and is the same object whichever member occurred at t), and by Lemma 3 the posterior over $\{e, e^R\}$ given that factor's value is the prior. So no readout of the emission at the position separates e from e^R better than the prior, which is chance in the sense of the standing conventions. Nothing in the argument bounds the size of s_t: a larger record composes more information about the same reversal-invariant factors and does not change the per-frame factor at the position.

The limit. Since the write does not separate the pair, the tokens it writes for "e" mark the occurrences of e and of e^R alike; the algebra's list for e is the merged list of both. Let the pair have m occurrences in total, of which the queried e has m_e. Under the exchangeability hypothesis of the standing conventions the k-th entry of the merged list is the true $t_k(e)$ exactly when the first k merged entries are all occurrences of e, an event of probability $\binom{m_e}{k}/\binom{m}{k} \leq (m_e/m)^k$. For a family whose ordinal is drawn among the pair's occurrences, k grows with m and this bound tends to zero; the chain fails, and by Lemma 1 under the floor convention $\mathrm{Cap} \to \varepsilon_{\mathrm{guess}}$. No step involved the index, the retriever or the read budget, which enter only through F_{pc} and multiply a chain probability that is already zero in the limit. Without exchangeability, $(m_e/m)^k$ is replaced by the best success probability of a rule that knows the counts and the interleaving law but not the labels; the theorem's chance clause still holds, and the limit clause holds whenever that probability tends to zero, which it does whenever the interleaving carries no order information.

Where the closure breaks. For w=2 the emission is $f(x_{t-1}, x_t)$. On the reversed stream the window at the mirror position is $((Rx)_{N-t}, (Rx)_{N+1-t}) = (x_{t+1}, x_t)$: the pair reversed and shifted by one. A function of an ordered pair may take different values on (x_{t-1}, x_t) and on (x_{t+1}, x_t), so the induction of the closure step has no base, and Proposition 2 exhibits a predicate in $\mathcal{P}_2$ that is not fixed by R. □

COROLLARY 2

Proof. Inclusion: a write with window w is a write with window w+1 whose emission ignores its oldest frame, so $\mathcal{P}_w \subseteq \mathcal{P}_{w+1}$. Strictness at w=1: Proposition 2 gives $e_{a\to b} \in \mathcal{P}_2$ with $(e_{a\to b})^R \neq e_{a\to b}$, and by Lemma 2 every element of $\mathcal{P}_1$ is fixed by R, so $e_{a\to b} \notin \mathcal{P}_1$; "the object entered the hand" is such a predicate. The order bound: a predicate decided at the frame by a write with window w is a function of $x_{t-w+1:t}$, so two streams that agree on that window at t receive the same value at t; a predicate that distinguishes two orderings of w+1 observations agreeing on every w-subwindow is not such a function. □

COROLLARY 3

Proof. Write $y = Rx$, so $y_u = x_{N+1-u}$. Then

$$s^y_u \;=\; \psi\big(\phi(y_{u-m}), \ldots, \phi(y_u)\big) \;=\; \psi\big(\phi(x_{N+1-u+m}), \ldots, \phi(x_{N+1-u})\big),$$

and with $t = N{+}1{-}u{+}m$ that argument list is exactly the argument list of s^x_t in reversed order. By hypothesis ψ is invariant under that reversal, so $s^y_u = s^x_{N+1+m-u}$: the score sequence of y is a reflection of the score sequence of x. Two sequences related by a reflection have the same multiset of values, so any threshold rule that fixes a constant or a quantile of the score distribution selects reflected index sets, and $T(y)$ is the reflection of $T(x)$. The emission at a selected position is frame-local, so

by Lemma 2 the record written at $u \in T(y)$ is the record written at $N{+}1{+}m{-}u \in T(x)$. The two multisets of records therefore coincide, and any predicate decided from the record takes the same value on x and on Rx, which is the hypothesis Theorem 1 needs. No step constrains $|T|$, so the conclusion holds at every event rate; taking T to be the set of true transition positions gives the oracle case, in which the reflection carries each occurrence of e to an occurrence of e^R and the selected set is mapped onto itself. □

LEMMA 2

Proof. For $w{=}1$ the window at position t is the single frame x_t, and the window at position $N{+}1{-}t$ of Rx is $(Rx)_{N+1-t} = x_t$: the same element of $\mathcal{X}$. Hence for any f, $\tau_t(x) = f(x_t) = f((Rx)_{N+1-t}) = \tau_{N+1-t}(Rx)$. If e is decided at the frame, $e_x(t) = g(\tau_t(x))$ for some g, and then $e^R_x(t) = e_{Rx}(N+1-t) = g(\tau_{N+1-t}(Rx)) = g(\tau_t(x)) = e_x(t)$ for every x and t, which is $e^R = e$. If (e, e^R) is a pair with $e \neq e^R$, then e is not fixed by R and so is not decided at the frame by any $w{=}1$ frame-local write; the same for e^R. That the write "assigns *put* and *take* the same value" is the identity $\tau_t(x) = \tau_{N+1-t}(Rx)$ read at an occurrence of *put*, whose mirror in the reversed stream is an occurrence of *take*. □

LEMMA 3

Proof. Let μ_e and μ_{e^R} be the laws of the frame at an occurrence of e and of e^R; Assumption 1 is $\mu_e = \mu_{e^R} =: \mu$. For measurable f the laws of $f(x_t)$ under the two hypotheses are the pushforwards $f_{\#}\mu_e = f_{\#}\mu_{e^R} = f_{\#}\mu$, so the two likelihoods coincide. By Bayes' rule, for τ in the support,

$$P[e \mid f(x_t) = \tau] = \frac{\pi \, \mathrm{d} f_{\#}\mu(\tau)}{\pi \, \mathrm{d} f_{\#}\mu(\tau) + (1-\pi) \, \mathrm{d} f_{\#}\mu(\tau)} = \pi,$$

almost surely with respect to $f_{\#}\mu$. □

COROLLARY 4

Proof. Let δ be a (possibly randomised) rule that depends on the frame through $\tau = f(x_t)$. Conditioning on τ,

$$\begin{aligned} P[\delta \text{ correct}] &= \mathbb{E}\big[\, P[\delta(\tau) = e \mid \tau]\, P[e \mid \tau] + P[\delta(\tau) = e^R \mid \tau]\, P[e^R \mid \tau] \,\big] \\ &\leq \mathbb{E}\big[\, \max\{P[e \mid \tau],\, P[e^R \mid \tau]\}\big] = \max\{\pi, 1-\pi\}, \end{aligned}$$

the last step by Lemma 3. The constant rule naming the majority member attains $\max\{\pi, 1-\pi\}$; any rule that names the minority member with positive probability sits strictly below it, which is the sense in which a rule may sit below the ceiling. □

PROPOSITION 2

Proof. Take the $w{=}2$ frame-local write with emission $f(x_{t-1}, x_t) = (\sigma(x_{t-1}), \sigma(x_t))$; $e_{a\to b}(t) = [\sigma(x_{t-1}){=}a \wedge \sigma(x_t){=}b]$ is a function of this emission, so $e_{a\to b} \in \mathcal{P}_2$. Exactness: given the status trajectory $\sigma(x_{1:N})$, the first difference $(\sigma(x_{t-1}), \sigma(x_t))_{t\geq 2}$ is a deterministic function of it, and $e_{a\to b}$ is the indicator of one row of the transition table applied to that difference; no further observation is consulted. Conjugacy:

$$(e_{a\to b})^R_x(t) = e_{a\to b,\, Rx}(N{+}1{-}t) = [\sigma((Rx)_{N-t}){=}a \wedge \sigma((Rx)_{N+1-t}){=}b] = [\sigma(x_{t+1}){=}a \wedge \sigma(x_t){=}b] = e_{b\to a}(t{+}1),$$

which is $e_{b\to a}$ shifted by one position, the alignment of a two-frame window under reversal. Since $a \neq b$, $e_{a\to b}$ and the shifted $e_{b\to a}$ differ on any stream in which the status enters b at t and stays there at $t{+}1$, so $(e_{a\to b})^R \neq e_{a\to b}$ and by Lemma 2 $e_{a\to b} \notin \mathcal{P}_1$. The two members of the pair are two different rows of the table, so they are separated by the lookup and not by any perception beyond the statuses. □

PROPOSITION 3

Proof. The m occurrence positions of a class are an arbitrary m-subset of $[N]$; answering $\mathrm{nth}(e, k)$ for all k determines the subset, which carries $\log_2 \binom{N}{m} = \Omega(m \log(N/m))$ bits. A count plus a

last-seen timestamp carries $O(\log m + \log N)$ bits, which for $m \geq 2$ and $N \gg m$ is below the requirement, so some interior $\mathrm{nth}(e, k)$ is not determined by the record; perception quality does not enter, since the argument is about what the record can hold rather than what was seen. For $k \in \{1, m\}$ the first and last timestamps suffice, which is the boundary the statement excludes. □

PROPOSITION 3

Proof. Let the true occurrences be $t_1 < \cdots < t_m$ and let each be detected independently with probability ρ. Without false positives the detected list $d_1 < d_2 < \cdots$ is a subsequence of the true list.

Exact resolution. $d_k = t_k$ if and only if $t_1, \ldots, t_k$ are all detected, an event of probability ρ^k by independence. Equivalently, let M be the number of misses among $t_1, \ldots, t_{k-1}$, so $M \sim \mathrm{Bin}(k{-}1, 1{-}\rho)$; when t_k is detected it sits at detected index $k - M$, and $d_k = t_k$ iff $M = 0$.

Neighbourhood read. Reading the resolved occurrence together with its r neighbours on either side in the detected list, $\{d_{k-r}, \ldots, d_{k+r}\}$, covers t_k iff t_k is detected and $M \leq r$ (a detected t_k can only sit at an index $\leq k$, so the upper neighbours never help). By independence of the detection of t_k from M,

$$P[\, t_k \in \{d_{k-r}, \ldots, d_{k+r}\}\,] = \rho\, P[M \leq r] = \sum_{j=0}^{r} \binom{k-1}{j} \rho^{k-j} (1{-}\rho)^j = \rho^k \sum_{j=0}^{r} \binom{k-1}{j} \Big(\frac{1-\rho}{\rho}\Big)^j,$$

which for $r{=}1$ is $\rho^k\big(1 + (k-1)(1-\rho)/\rho\big)$. Since $P[M \leq r] \leq 1$ the neighbourhood read never exceeds ρ, the recall of the addressed occurrence itself, with equality at $r \geq k - 1$ when every detection is read. The multiplier over the exact rate is $\sum_{j \leq r} \binom{k-1}{j}((1-\rho)/\rho)^j$, which for ρ in the range the measurement reports and $r{=}1$ is a factor of a few and not an escape: the measured neighbourhood read of §E.3 sits inside this envelope.

False positives. Let false detections occur at rate ϕ per true occurrence, independently of the true detections, and let F be the number of them before t_k. Now $d_k = t_k$ iff t_k is detected and exactly $k{-}1$ detections precede it, i.e. $D + F = k - 1$ with $D \sim \mathrm{Bin}(k{-}1, \rho)$ the true detections among $t_1, \ldots, t_{k-1}$. Hence

$$P[d_k = t_k] = \rho\, P[D + F = k - 1] \leq \rho \max_j P[D = j],$$

the inequality because $P[D + F = k - 1] = \sum_j P[D = j]\, P[F = k - 1 - j] \leq \max_j P[D = j] \sum_i P[F = i]$. The probability is not monotone in ρ: a false positive that lands before t_k stands in for a missed true occurrence, so at ϕ near $1 - \rho$ (false positives arriving at the rate true occurrences are missed) the sum is largest. Its envelope over all ϕ is $\rho \max_j P[\mathrm{Bin}(k{-}1, \rho) = j]$: the largest point mass of a binomial with $k{-}1$ trials, which is non-increasing in k and tends to zero as k grows (of order $k^{-1/2}$ by the local limit theorem). For every fixed $\phi \neq 1 - \rho$ the mean of $D + F$ is $(k-1)(\rho+\phi) \neq k-1$, the event $D+F = k-1$ is a large deviation, and the decay in k is geometric. The statement in the text records the exact form, the compensating rate at which it is largest, and the two decay regimes. □

LEMMA 4

Proof. E_{ord} requires two things of the record. First, that $\mathrm{nth}(e, k)$ be computable from what it holds when every occurrence has been detected: that is the indicator, and when it is zero the operation is undefined at k and E_{ord} is empty, whatever ρ is (Proposition 3 gives the indicator zero at interior k for $O(1)$ words per class). Second, that the k-th kept entry be the true t_k, which under the model of Proposition 3 has probability ρ^k. Both are necessary, so $F_{\mathrm{ord}}(k) \leq \mathbf{1}[\cdot] \cdot \rho^k$; when the indicator is one and detection is the only error the two conditions are the whole of E_{ord} and the bound is an equality. The two regimes are the two values of the indicator. □

COROLLARY 5

Proof. In the propagation regime $F_{\mathrm{ord}}(k) = \rho^k$. At $k{=}1$ this is ρ, and $\mathrm{nth}(e, 1) = t_1$ iff t_1 is detected, so $F_{\mathrm{ord}}(1)$ is the recall of the first occurrence. For every k, $F_{\mathrm{ord}}(k{+}1)/F_{\mathrm{ord}}(k) = \rho^{k+1}/\rho^k = \rho$. Neither reading fits anything: both are the model's ρ read off two points of the curve. □

PROPOSITION 8

Proof. Let D be the detected list and $\Phi(D) \subseteq D$ a sublist. The k-th element of $\Phi(D)$ equals t_k only if $t_k \in \Phi(D) \subseteq D$, so t_k was detected, which has probability ρ; hence $P[\mathrm{nth}_\Phi(e,k) = t_k] \leq \rho$ for every Φ, and the ceiling ρ is the recall of the addressed occurrence, a property of the detector that no map from D to its sublists touches. If moreover $\Phi(D)$ contains no false positive before t_k (in particular if the detector has none), then the $k-1$ elements of $\Phi(D)$ preceding t_k are true occurrences among $t_1, \ldots, t_{k-1}$, so all of $t_1, \ldots, t_{k-1}$ were detected and kept, and the event requires $t_1, \ldots, t_k$ all detected: probability at most ρ^k. A filter therefore acts only through the false positives it removes or the true detections it discards; it cannot add a detection that was not made. For the union of two detectors, an occurrence is detected by the union iff it is detected by at least one, so the union's per-occurrence recall is $\rho_\cup = 1 - P[\text{both miss}] \leq 1 - (1-\rho_1)(1-\rho_2)$, with equality iff the two miss events are independent and strict inequality under positive dependence; applying the no-false-positive bound to the union's list gives $\rho_\cup^k$. When the two detectors' misses coincide, $\rho_\cup = \max\{\rho_1, \rho_2\}$ and the union is the better detector alone. □

PROPOSITION 9

Proof. Under $\oplus_{\mathrm{addr}}$ the read executes an expression of $\mathcal{A}$ over $\mathcal{C}_\mathrm{s}$'s record and dereferences $\mathcal{C}_\mathrm{v}$ at the returned position. The event E_{id} is a property of the record (the class is decided), E_{ord} is a property of the record and the executed expression (the position is right), and E_{pc} is a property of the one dereferenced frame and the read over it. The answer is correct on $E_{\mathrm{id}} \cap E_{\mathrm{ord}} \cap E_{\mathrm{pc}}$, so by the chain rule $P[\text{correct}] \geq F_{\mathrm{id}} F_{\mathrm{ord}} F_{\mathrm{pc}}$, where F_{pc} is $\mathcal{C}_\mathrm{v}$'s percept factor conditional on a correctly addressed frame. No event in the intersection involves the backbone reading two texts: the position is passed as a position. Under $\oplus_{\mathrm{cat}}$ the record is rendered to prose and the frame is chosen by the backbone from that prose; let J be the event that the chosen frame is the resolved position. The corresponding chain is $E_{\mathrm{id}} \cap E_{\mathrm{ord}} \cap J \cap E_{\mathrm{pc}}$, with probability $F_{\mathrm{id}} F_{\mathrm{ord}}\, P[J \mid C]\, F_{\mathrm{pc}}$, and $P[J \mid C]$ is a fourth factor, equal to one only if the backbone always recovers the position from its description. □

PROPOSITION 10

Proof. By Lemma 1, $\mathrm{acc}(\texttt{axis}) = F_{\mathrm{id}} F_{\mathrm{ord}}\, F_{\mathrm{pc}} + (1 - F_{\mathrm{id}} F_{\mathrm{ord}})\, \varepsilon_{\mathrm{off}}$. The control executes the same address and dereferences an occurrence that is never the correct one, so on every query its chain is broken; by hypothesis its conditional accuracy on a broken chain equals that of `axis`, so $\mathrm{acc}(\texttt{axis_wrongaddr}) = \varepsilon_{\mathrm{off}}$. Subtracting,

$$\mathrm{acc}(\texttt{axis}) - \mathrm{acc}(\texttt{axis_wrongaddr}) = F_{\mathrm{id}} F_{\mathrm{ord}}\, (F_{\mathrm{pc}} - \varepsilon_{\mathrm{off}}),$$

whose absolute value is at most $F_{\mathrm{id}} F_{\mathrm{ord}}\, |F_{\mathrm{pc}} - \varepsilon_{\mathrm{off}}| \leq F_{\mathrm{id}} F_{\mathrm{ord}}$ since both terms in the difference lie in $[0, 1]$. □

PROPOSITION 11

Proof. Under $\oplus_{\mathrm{addr}}$ with the floor convention $\varepsilon_{\mathrm{off}} = a_\varnothing$, Lemma 1 gives $a_{\mathrm{addr}} = F_{\mathrm{id}} F_{\mathrm{ord}}\, F_{\mathrm{pc}} + (1 - F_{\mathrm{id}} F_{\mathrm{ord}})\, a_\varnothing$, so $a_{\mathrm{addr}} - a_\varnothing = F_{\mathrm{id}} F_{\mathrm{ord}}\, (F_{\mathrm{pc}} - a_\varnothing)$. Strict super-additivity is $F_{\mathrm{id}} F_{\mathrm{ord}}\, (F_{\mathrm{pc}} - a_\varnothing) > (a_\mathrm{s} - a_\varnothing) + (a_\mathrm{v} - a_\varnothing)$. If $F_{\mathrm{pc}} \leq a_\varnothing$ the left side is non-positive and the inequality fails whenever the single-channel gains are non-negative; otherwise dividing by $F_{\mathrm{pc}} - a_\varnothing > 0$ gives the displayed condition, and the steps reverse, so it is an equivalence.

For the two necessary conditions: since $F_{\mathrm{id}} F_{\mathrm{ord}} \leq 1$, $a_{\mathrm{addr}} - a_\varnothing \leq F_{\mathrm{pc}} - a_\varnothing$, so strict super-additivity forces $(a_\mathrm{s} - a_\varnothing) + (a_\mathrm{v} - a_\varnothing) < F_{\mathrm{pc}} - a_\varnothing$, i.e. $a_\mathrm{s} + a_\mathrm{v} - a_\varnothing < F_{\mathrm{pc}}$. Since $F_{\mathrm{pc}} \leq 1$ this implies $a_\mathrm{s} + a_\mathrm{v} - a_\varnothing < 1$, which is the availability condition of Anonymous authors (2026a); the same conclusion follows for any composite from $a_{\mathrm{addr}} \leq 1$, which is how the companion paper states it. The sharpening to F_{pc} is specific to address composition, whose accuracy is bounded by the percept factor on one correctly addressed frame rather than by one. □

H THE STATUS-TRANSITION TABLE

Fixed before any grid ran; never tuned on a result. Statuses are first mapped to a class by a synonym table (*open*, *opened*, *ajar*, *lid-open* → *open*; *closed*, *shut*, *lid-closed* → *closed*; *in-hand*, *held*, *holding*, *in hand* → *held*; *on*, *running*, *lit* → *on*; *off*, *stopped*, *unlit* → *off*); an unmapped status is unclassified.

transition	derived event	transition	derived event
closed → *open*	open	*open* → *closed*	close
→ *held*	take	*held* →	put
off → *on*	turn-on	*on* → *off*	turn-off

A change between unclassified statuses emits no event and is counted, not absorbed. An object entering the status field for the first time with status *held* emits *take*; an object leaving the field while *held* emits *put*.

I PROMPTS AND GRAMMAR, VERBATIM

The pair write (conjunctive corpus). The `single`, `dup` and `rev` rules use this prompt unchanged with their own frames attached; `single` attaches one image and the sentence about two images is then vacuous by construction, a property the design accepts so that the prompt is one string.

```
You maintain a compact JSON state for a person wearing a head-mounted camera in a
kitchen. It is your ONLY memory of the stream: you will never see these frames again.
You are shown TWO images IN ORDER. The FIRST is the PREVIOUS keyframe. The SECOND is
the CURRENT keyframe. The action is what CHANGED between them: one frame alone cannot
tell putting from taking, or opening from closing, because those are the same
arrangement of hand and object passed through in opposite directions. Only the change
tells you which.
Current state (may be truncated): {state}
Output JSON with these fields IN THIS ORDER:
 "change": one clause naming what is DIFFERENT between the first and the second image
   -- what entered or left the hand, what became more or less open, what was switched,
   what moved. Write "nothing" only if the two images are truly identical. Answer this
   field FIRST and derive the next one from it.
 "action": what the person does in the CURRENT frame, as "<verb> <noun>", derived
   from "change":
     object LEAVES the hand or arrives on a surface -> "put-down <noun>"
     object ENTERS the hand                         -> "take <noun>"
     openable thing is MORE open than before        -> "open <noun>"
     openable thing is LESS open than before        -> "close <noun>"
     switchable thing changed state                 -> "turn-on <noun>" / "turn-off <noun>"
   Use "hold <noun>" ONLY when "change" is "nothing".
 "affordances": REQUIRED and must not be empty whenever any container, appliance, tap,
   drawer, door or hand-held object is visible. One entry per such object, naming its
   status in the CURRENT image, e.g. {"fridge": "open", "drawer": "closed", "tap":
   "off", "cup": "in-hand", "bin": "lid-open"}. Write "unknown" for a status you cannot
   see. Carry forward every entry already in the state and change only what these two
   images show changed. This is the world's state, and an action you could not see
   directly is still recoverable from a status that moved.
 "location": the work surface or appliance the person is at now -- one of sink, hob,
   counter, fridge, table, cupboard, bin, floor. Never "unknown" if any of them is
   visible.
 "objects": the object classes visible in the CURRENT image, each with last_seen={t}.
Output ONLY the JSON object, then on the LAST line only:
CAPTION: <the action as "<verb> <noun>", then the other object class names visible in
the CURRENT image, comma-separated. Names only.>
```

The free-vocabulary write (Amendment A5). `pair_free`, `dup_free` and `delta_free` share this prompt, which differs from the one above only in that the enumerated verb table is gone. The rules differ from one another only in which two frames are attached: the previous keyframe and the current one, the current one twice, or the frame cut one second before the current one and the current one. The instruction that decides the `dup_free` control's behaviour is the parenthesis in `action`:

a writer shown two identical images reports `change` as *nothing* and is then required to write *hold*, which is why qwen3.8-27b's `dup_free` cache contains one verb (§4.1). That is the instruction doing what the theorem says it must, not a defect of the cache.

```
You maintain a compact JSON state for a person wearing a
head-mounted camera in a kitchen. It is your ONLY memory of the stream: you will never
   see
these frames again.
You are shown TWO images IN ORDER. The FIRST is the PREVIOUS keyframe. The SECOND is the
CURRENT keyframe. The action is what CHANGED between them: one frame alone cannot tell
putting from taking, or opening from closing, because those are the same arrangement of
   hand
and object passed through in opposite directions. Only the change tells you which.
Current state (may be truncated): {state}
Output JSON with these fields IN THIS ORDER:
 "change": one clause naming what is DIFFERENT between the first and the second image
   -- what
   entered or left the hand, what became more or less open, what was switched, what
   moved.
   Write "nothing" only if the two images are truly identical. Answer this field FIRST
   and
   derive the next one from it.
 "action": what the person does in the CURRENT frame, as "<verb> <noun>", derived from
   "change". Use the most specific verb you can (any verb, not from a list); write
   "hold <noun>" ONLY when "change" is "nothing".
 "affordances": REQUIRED and must not be empty whenever any container, appliance, tap,
   drawer, door or hand-held object is visible. One entry per such object, naming its
   status
   in the CURRENT image, e.g. {{"fridge": "open", "drawer": "closed", "tap": "off",
   "cup": "in-hand", "bin": "lid-open"}}. Write "unknown" for a status you cannot see.
   Carry
   forward every entry already in the state and change only what these two images show
   changed. This is the world's state, and an action you could not see directly is still
   recoverable from a status that moved.
 "location": the work surface or appliance the person is at now -- one of sink, hob,
   counter,
   fridge, table, cupboard, bin, floor. Never "unknown" if any of them is visible.
 "objects": the object classes visible in the CURRENT image, each with last_seen={t}.
Output ONLY the JSON object, then on the LAST line only:
CAPTION: <the action as "<verb> <noun>", then the other object class names visible in
   the
CURRENT image, comma-separated. Names only.>
```

The read loop and the address grammar.

```
You are an embodied QA agent with a token budget. Your memory of the session is an
INTERACTION LEDGER that recorded every interaction and WHEN it happened, plus the
keyframes themselves. You cannot search the keyframes by description; you address them.
Available actions (output exactly one, on the last line):
ADDR: nth(<verb> <noun>, k)    -- the k-th time that interaction happened
ADDR: first(<object>=<status>) -- when the object first ENTERED that status,
                                  e.g. first(fridge=open), last(tap=on)
ADDR: first(<verb> <noun>)     -- the earliest time it happened
ADDR: last(<verb> <noun>)      -- the latest time it happened
ADDR: count(<verb> <noun>)     -- how many times it happened
  any of the first three may be followed by +d or -d to step d interactions later or
  earlier, e.g.  ADDR: nth(open cupboard, 5)+1
STATE: all                     -- read the compressed scene state
ANSWER: <final answer>
Interactions the ledger can address: {vocab}
Question: {q}
{context}
Round {r}/{R}. Think briefly, then output one action on the last line.
```

The TempCompass write (all three rules) and read.

```
You maintain a compact JSON record of a video. It is your ONLY memory of it: you will
never see these frames again.
You are shown ONE or TWO images. If TWO, they are IN ORDER: the FIRST is the PREVIOUS
keyframe and the SECOND is the CURRENT one, and the motion is what CHANGED between them
-- a single frame cannot tell left from right, rising from falling, or opening from
closing, because those are the same picture passed through in opposite directions. If
ONE, describe what it shows and say so.
Current record (may be truncated): {state}
Output JSON with these fields IN THIS ORDER:
 "change": one clause: what is DIFFERENT between the previous and the current image
   (position, pose, size, colour, count, state). Write "single frame" if only one image
   was given.
 "action": what the main subject is doing NOW, as a short verb phrase, derived from
   "change".
 "motion": the direction of the main subject's movement, one of: left, right, up,
   down, toward the camera, away from the camera, clockwise, anticlockwise, none,
   unknown.
 "attributes": the main subject's current visible attributes as {"size": ...,
   "colour": ..., "count": ..., "brightness": ...}; carry forward what the record
   already has and change only what these images show changed.
 "subjects": the main things visible now, each with last_seen={t}.
Output ONLY the JSON object.

You answered questions about a video from a written record of it, made while it
played. You cannot see the video now. Here is the complete record, one line per
keyframe in order:
{record}
Final record: {state}
Question: {q}
Answer with the option letter only.
```

J PROVENANCE OF THE COMPANION-SUPPLIED INPUTS

This appendix makes the paper self-contained on the three inputs it takes from the companion submissions: two thresholds and the baseline arms. Nothing here re-scores a companion claim; each paragraph states what the number is, where it is load-bearing, and how the released material reproduces it.

The conjunctive fraction (0.51%). The pooled fraction of items, across the embodied multimodal benchmarks the companion measurement audits, whose query needs a locator and a percept jointly: the only family on which the product law has a locator to bound; on the rest of the field's items it has none (§B.8). It is read from the benchmarks' own annotations with no model in the loop, and it is load-bearing here only as the warrant for up-sampling that family (item (7) of §4.4). The per-benchmark counts and the audit script ship with this paper's artifact.

The availability threshold (42.3%). The companion measurement certifies that on this corpus's CONJ family the single channels leave room for the conjunction (the premise Proposition 11 needs), and prices it at 42.3%. It is computed from the corpus's annotations alone, and it is carried forward as a threshold, not as evidence: P-Δ2 turns it into a prediction about status-tracking (§D), P-O1 judges the oracle arm against it (§E.7), and the certifying submission's own headline arm fails against it.

The baseline arms. `blind`, `retr_only`, `state_only` and ASP in §4.2 are the budgeted grid's arms at the same budget on the same keyframes. Their raw transcripts and the scorer that produced their numbers are released with this paper's artifact, so every comparison in §4.2 can be re-derived without access to the submission that first ran them.

Artifact, named file by file. The eight-stream list with segment boundaries, the item generator for the three families, the per-item annotation dumps that reproduce both thresholds from scratch, every write cache, and every read arm's raw output ship with the paper. Concretely: `analysis/a0_conjfrac_probe.py` over `results/a0_conjfrac_probe.json` regenerates the conjunctive fraction; `analysis/a0_bounds.py` over `results/a0_bounds.json`

regenerates the availability threshold; `analysis/a0_corpus.py` regenerates the corpus itself from the public EPIC-KITCHENS-100 annotations; and `results/cells_epic_v3/` holds the baseline arms' per-question records. A reader can therefore verify each companion-supplied input independently of the submissions that produced it.

Nothing here rests on an unpublished companion. The companions are under review and carry no archival identifier, so a claim of ours that depended on one would be a claim a reader cannot check. Every quantity this paper takes from one is either recomputable from the files just named or is carried as a *threshold* the paper argues against rather than as evidence for it. The availability inequality itself is not imported at all: it is derived here, in the proof of Proposition 11, as the special case $F_{\mathrm{pc}} \leq 1$ of this paper's sharper bound, and Anonymous authors (2026a) is cited for having named and measured it rather than for its validity. The companions are cited for priority, for the corpus this paper re-uses, and for the questions this paper explicitly declines to answer.

K THE REGISTRATION, AND ITS AMENDMENTS

The registration file is reproduced here in substance; its predictions and falsifiers are in §D verbatim.

Frozen (§1–2 of the registration). Benchmarks, item counts, keyframe rules (40 uniform on EgoSchema; $\min(16, \lceil \mathrm{dur} \rceil)$ at 1 fps on TempCompass; one per action segment at its midpoint on the conjunctive corpus), scorer (option index; the corpus's judge-free scorer), budget $B{=}4096$ for every arm, the four write rules with their coverage, the three read modes, and the backbone lists.

Measured before registration (§3). M-1 the single-frame identity factor; M-2 the reversal confusions; M-3 single-frame address resolution at 0.4–1.6% exact, 2.0–3.3% count-exact under three lookup rules; M-4 `blind` on the conjunctive corpus; M-5 `blind` on EgoSchema-500 (five backbones); M-6 a 20-keyframe smoke run, explicitly not a result.

Reporting rules (§7). Every cell records its unparsed count and a family above 5% unparsed is reported unscorable; realised per-decision spend is reported for every arm; `single` and `pair` accuracy are computed by one function; coverage caps appear in every caption; frozen quantities change only by timestamped amendment; and if H2 lands and H1 fails the paper is a write-path architecture with a public-benchmark gain on temporal sensitivity and a null on long-horizon QA, not to be reframed.

Cost (§8). Staged: the conjunctive corpus ($\sim$21M tokens), TempCompass ($\sim$24M), EgoSchema-500 ($\sim$80M); the projection is re-checked and reported before the third stage, which has not started.

Amendment A1 (before any grid cell). W1′ added and made authoritative; the offline cache extended to record per-frame statuses; four partial episodes discarded and relaunched; the address grammar extended with state addressing; P-Δ1, P-Δ2 and F-Δ registered.

Amendment A2 (offline $16/40$ complete, before any E3 cell was read). The early write-path verdicts of Table 1; the four refuted remedies; the error-propagation reading; the prediction that H3 fails on both backbones through P-Δ2; the vocabulary collapse reported as a design cost.

Amendment A3 (on the code-and-process review). The normalisation defect and its repair by canonical class tables; the cost claim corrected; one pre-repair E3 cell quarantined and E3 restarted under one code state; the instance-level diagnosis recorded as not a code artefact.

Amendment A4 (the morning after the overnight grids). The verdicts of §4 recorded against each registered criterion; P-D1a's attribution to the window withdrawn on the `dup`/`rev` controls and the direction test; H2's verdicts on both backbones with the clip/time-reverse motion test, and the pair-spacing δ registered as the next experiment; F3 and F4 tripped on gemma-3-12b; qwen3.8-27b's cells declared unscorable under reporting rule 1 with the budget-exhaustion cause; a per-round output cap registered as a new arm, not applied mid-run; and the paper's standing under reporting rule 6

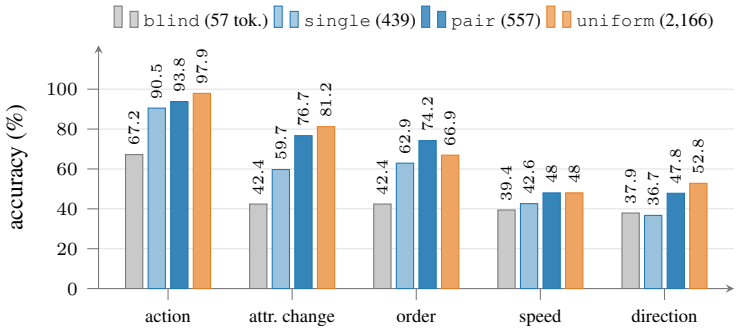


Figure 14: TempCompass by dimension, qwen3.8-27b, the backbone on which H2's overall criterion holds. `pair` over `single`: $++9.4$ overall and $++11.1$ on direction (registered criterion ≥ 15: held on overall, missed on direction), against `blind` at 46.1%. The pair write reaches 68.0% at 557 tokens against `uniform`'s 69.4% at 2,166 (Fig. 12b); direction at 47.8% stays ten points above blind and ten below the `uniform` arm's 52.8%. gemma-3-12b, where the same move does nothing on direction, is in Fig. 11.

Table 11: TempCompass multi-choice, every arm, per dimension. 1,580 items, 410 clips, B=4096; keyframes at 1 fps, at most 16. `uniform` attaches the frames themselves and is the reference the write rules are not expected to reach; `single`/`dup`/`pair` are record-only reads over a ledger written by the rule named; `dup` runs on gemma-3-12b only (the registration's declared coverage cap). *direction* is the dimension the pair was designed for.

backbone	arm	all	action	attribute	direction	order	speed	tokens	unp.
gemma-3-12b	`blind`	43.0	58.3	38.9	36.4	38.4	41.6	54	4
gemma-3-12b	`uniform`	61.6	94.4	61.8	43.9	57.3	49.2	2,382	0
gemma-3-12b	`single`	55.2	85.2	60.8	34.6	55.0	40.4	512	25
gemma-3-12b	`dup`	53.6	83.4	53.5	32.8	54.3	43.2	504	35
gemma-3-12b	`pair`	55.5	86.1	64.6	34.9	55.0	36.9	513	26
qwen3.8-27b	`blind`	46.1	67.2	42.4	37.9	42.4	39.4	57	0
qwen3.8-27b	`uniform`	69.4	97.9	81.2	52.8	66.9	48.0	2,166	2
qwen3.8-27b	`single`	58.6	90.5	59.7	36.7	62.9	42.6	439	0
qwen3.8-27b	`pair`	68.0	93.8	76.7	47.8	74.2	48.0	557	0

stated: a write-side wall with its measurement, an architecture refuted by its own controls, and the relocation of the binding constraint to the detector, not to be reframed as a gain.

Amendment A5 (before any A5 cell was paid for): the positive-result programme. Three free-vocabulary write rules (`pair_free`, `dup_free`, `delta_free` at one second) on two backbones, with P-δ1 (`delta_free` $-$ `dup_free` ≥ 8), P-δ2 (direction $\geq 80\%$ under `delta_free`, $\leq 76\%$ under `dup_free`), P-δ3 (`pair_free` within ± 3 of `dup_free`) and F-δ (less than 3 points on both backbones kills the window as a lever at any spacing we can cut). Three oracle-ledger read modes on the class-semantics corpus, with P-O1 (oracle CONJ $> 42.3\%$ on two of three backbones), P-O2 (oracle $-$ wrongaddr ≥ 15, oracle $-$ noderef ≥ 8), P-O3 (within ± 5 of the best single-channel arm off-family) and F-O (oracle $\leq 25\%$: read side also broken). The per-round completion cap as a new arm. Two addenda: the companion's prediction that class semantics would at least double ordinal resolution over the existing pair caches failed (resolution within three frames moved from 2.4 to 1.6% and 8.2 to 7.5%), recorded as the finding that the model ledger and not the corpus semantics limits resolution; and the scorer-unification finding of §4.1, with the retirement of the head-only scorer and the re-emission of every write-path number, including M-1, under the unified one. A post-registration replication of the three free-vocabulary rules on gemma-4-31b is judged under the same thresholds and marked *(post-reg.)* in the ledger. *Verdicts:* Table 10.

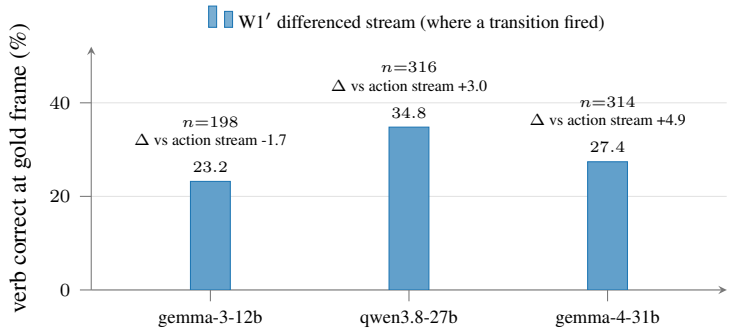


Figure 15: The scaffold's derivative is backbone-dependent as a verb writer, and the registration said what would happen if it were worse. Verb accuracy of the W1′ differenced stream on the sub-population where a difference is defined (n=198/316/314 of 490 gold frames), against the model's own action stream under the identical pair prompt (Δ in -1.7/+3.0/+4.9). P-Δ1 registered: worse on any backbone ⇒ drop W1′; the honest reading of a small deficit on one backbone against gains on two is *backbone-dependent*, and the paper neither drops nor claims it. The statuses the derivative differences are written by the same detector whose direction sense the controls measured at the prior; its ceiling is the action field's.

L SECONDARY EXPERIMENTS AND SENSITIVITY

TEMPCOMPASS, EVERY ARM BY DIMENSION

Table 11 is the generated form of Table 7: same cells, transposed, with unparsed counts and spend per arm. Figure 14 plots the qwen3.8-27b columns of the same grid. It is written directly from the cell files, whereas the body table is typeset from macros entered when those cells landed; where the two differ in the last digit (gemma-3-12b `single`, 55.3 against 55.2) the appendix is the rounding of the cell file and the body table rounds a value that was recorded one digit earlier. The generated table is authoritative.

THE DIRECTION STATISTIC, AND WHY ITS n VARIES

The direction column of Table 1 conditions twice: on gold frames whose verb is *put* or *take*, and on writes that chose one of those two verbs. The second condition is what makes the statistic a test of the window rather than of the vocabulary: a rule that never writes *put* or *take* has no direction to be right about, and its n says so. The listed prompt puts one of the two verbs on most such frames; the single-frame and free-vocabulary rules put one there on a minority, and on qwen3.8-27b's `dup_free` on none. The majority prior of 74% is the share of *put* among the *put*/*take* gold frames at the locators, so a writer that always said *put* would score it; a rule at the prior has not read the direction, whatever its verb accuracy.

WHAT THE FREE-VOCABULARY CACHES WROTE

Under the free prompt the written vocabulary is 35 and 42 distinct canonical verbs at the gold frames against 7 and 11 under the listed one; the most frequent free-prompt verb on both backbones is *hold*, which the class tables treat as a verb of its own and which is never a gold verb at a locator, since locators are transitions. The collapse of qwen3.8-27b's `dup_free` to *hold* on every frame of its partial cache is the limiting case: the duplicate-frame control, which under the listed prompt scored above the pair, scores nothing once the list is gone, because the list was supplying the transition. That is the A4 controls' attribution stated a second way, by removal rather than by addition.

THE CLASS-SEMANTICS RE-SCORING OF THE PAIR CACHES

Re-running the canonical-key resolution analysis of Table 4 on the companion's class-semantics items from the same pair caches: the event class is found in the union ledger about as often as before, the count is resolvable about as often, and the address is exact within three frames *less* often (1.6 and 7.5% against 2.4 and 8.2%). The companion's registered prediction that class semantics would at

least double resolution fails, and the failure is informative: per-occurrence detection at $\rho \approx 0.2$–0.35 is a property of the written ledger and is unchanged by how the corpus counts, which is what made the oracle ledger, rather than a further corpus repair, the decisive read-side test.

M REPRODUCTION

Everything below is in the repository and runs from a fresh clone with the corpus videos fetched by the scripts named; no number in the paper is typed by hand into a table, and no placeholder cells are typeset: a rule outside a backbone's declared coverage cap simply has no row, and the cap is stated in the table's caption.

- `scripts/methods/axis.py`: the write prompt, the status classes and transition table, the ledger with its collapse and canonicalisation, the address grammar and resolver, the read loop, and the two controls.
- `scripts/run_axis.py`: `--phase offline` (writes, cached per rule, backbone and episode, including the three free-vocabulary rules), `--phase probe` (Table 1, one function over every cache), `--phase main` (E3 cells; `AXIS_DATA=data/epic_v3` for the oracle modes, `AXIS_ROUND_CAP=256` for the capped arm).
- `scripts/fetch_epic_delta_frames.py`: the frame one second before each keyframe, for `delta_free`.
- `analysis/axis_tables.py`: every table and every write-path, oracle, E3 and A5-ledger macro in this paper, from `results/axis_write.json` and the cell files; it also re-emits the preamble's write-path macros so that no number predates the unified scorer.
- `analysis/axis_probe.py`: single-frame address resolution over the companion caches (M-3).
- `scripts/fetch_tempcompass.sh` and `scripts/run_tempcompass.py`: the keyframe extraction at the frozen rule, and the five arms of Table 7.
- `notes/registered_axis.md`: the registration and its five amendments, append-only.
- `data/epic/EPIC_100_noun_classes.csv`, `EPIC_100_verb_classes.csv`: the published class tables used for canonicalisation.

Caches are keyed by (rule, backbone, episode) and a cell is reproduced from its cache without a write call; the write calls themselves are at temperature 0 through one provider, and the companion papers price the residual re-sampling movement of that provider at about one point of mean absolute cell movement.

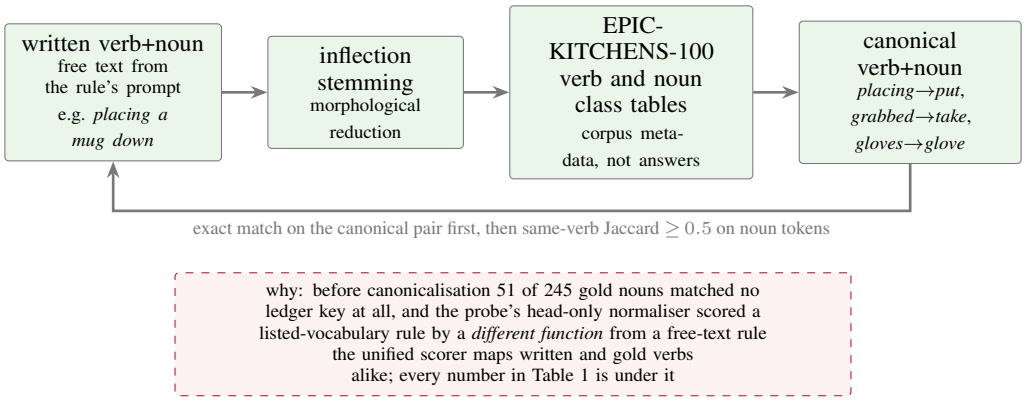


Figure 16: One scorer reads every cache. Written and gold verbs are mapped alike through the corpus's published class tables with inflection stemming, so a rule cannot gain by spelling and a listed-vocabulary rule is scored by the same function as a free-text rule. The registration's original M-1 was recorded under the head-only scorer; the unified scorer is the registered baseline that moved, and the change is disclosed in Appendix K.

N Additional Figures

The remaining figure of the full manuscript that the main text does not carry: the class-canonical scorer (Fig. 16).