MEMORY WITHOUT TRAINING: A COMPLEMENTARY LEARNING SYSTEM ON FROZEN MULTIMODAL ENCODERS

Anonymous authors
Paper under double-blind review

ABSTRACT

We report a complementary learning system that is never trained, and whose every behaviour is derived before it is measured. A fast store binds role–filler pairs by circular convolution and completes them through an attractor read; a slow store accumulates the same content in a Hebbian outer-product matrix under passive decay; replay of confidence-gated retrievals carries episodes from one to the other. Binding capacity, the forgetting horizon, the consolidation schedule, the onset of semanticisation and the magnitude of a priming effect are each predicted from a single interference calculation, with no fitted parameter. Content comes from frozen encoders: MiniLM and CLIP codes enter through a codec whose crosstalk we predict from the codes' power spectra alone, and an untrained random projection into a shared workspace binds language and vision slots into one episode, within 0.0107 of the derived accuracy and with no new constant. Three findings stand out. Binding capacity belongs to the workspace dimension rather than to the encoder that produced the content, so projection alone buys 0.1530 of accuracy at depth. The spacing effect reverses direction with the clock it is measured from, and both directions follow from the same law. And the full circuit adds nothing over the storeless codec law, agreeing within 0.0093 at every load, so end-to-end behaviour is the product of independently derived parts. We registered all 23 predictions before measuring them; four failed, and we report the failures, the diagnoses they carried, and three corrected laws re-validated on grids they had never seen.

1 INTRODUCTION

A frozen sentence or image encoder is an excellent perceptual front end and a poor memory. It maps content to a vector, and that is all it does: it cannot say that *this* filler occupied *that* role in *this* episode, cannot hold a set of such bindings apart from a neighbouring set, and has no notion of an episode fading, being rehearsed, or turning into a generality. Systems that add these abilities almost always add training — fast-weight layers, memory controllers, learned retrievers — and with training comes a familiar cost: the system's capacity, its forgetting curve and its interference behaviour become empirical properties, measured after the fact and specific to the checkpoint.

We take the opposite position. This paper reports a complementary learning system in which nothing is trained and, more importantly, in which every quantity the system exhibits is *derived in advance* from one calculation about interference between superposed bindings. Capacity, the age at which a trace becomes unreadable, how much a rehearsal schedule extends that age, when repeated exposures stop reconstructing an exemplar and start reconstructing a prototype, how large an effect a single weak exposure leaves behind, and what happens when slots in one episode come from different encoders — each is a prediction with no free parameter, written down before the corresponding measurement exists.

That discipline is the paper's methodological core, and we hold ourselves to it in a checkable way. Every prediction was registered with its tolerance, its comparator and its date, and hashed into a lock file, in a commit that precedes the commit carrying the data it is compared against. Of the 23 predictions reported here (Appendix A lists each with its bound, date and outcome), four failed.

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.

We report them as failures, with the diagnosis each carried, and we report what happened when the corrected laws were re-registered on grids they had never seen: two held exactly, one within a single event. A law that is fitted cannot fail this way, and a law that has failed and then held on fresh ground is a different kind of evidence from a curve that matches.

The system runs on real content. Filler codes come from frozen MiniLM and CLIP encoders, and enter the memory through a codec whose crosstalk we predict from the codes' power spectra alone. Because binding spreads a code's structure across the whole episode dimension, encoder anisotropy turns out to matter far less than one would guess: the codec is effectively modality-blind on real codes, and language and vision slots can be bound into a single episode after nothing more elaborate than a fixed random projection into a shared workspace.

Our contributions are:

- A training-free complementary learning system — fast binding store, slow Hebbian store, confidence-gated replay between them — specified completely by closed-form operators, with no learned parameter anywhere (§2).
- Parameter-free laws for its behaviour: capacity and forgetting (§3), consolidation schedules including a sign reversal of the spacing effect (§4), and priming with an amnesia dissociation (§5) — each registered before measurement.
- A codec and workspace that admit frozen multimodal encoders without a new constant, with capacity shown to belong to the workspace dimension rather than the encoder's (§6).
- Evidence that the composition is transparent: the full attractor circuit matches a storeless calculation at every load, so the system's end-to-end behaviour is the product of its independently derived parts (§7).

2 THE SYSTEM, AND ONE INTERFERENCE CALCULATION

Two stores. The fast store holds an episode as one vector. Given k role–filler pairs (r_j, z_j) with unit-norm codes, it binds each pair by circular convolution, marks slot j by a fixed random permutation raised to the jth power, weights it, and superposes:

$$x \;=\; \sum_{j=0}^{k-1} w_j \, \pi^j (r_j \circledast z_j)\,, \qquad w_j = \gamma^{\,k-1-j}. \tag{1}$$

A permutation is used rather than a cyclic shift because a shift commutes with circular convolution, which would make position indistinguishable from a relabelling of the role. Reading slot j inverts the two operations in order and decodes against the codebook:

$$\hat{z}_j \;=\; r_j^{\dagger} \circledast \pi^{-j}(x), \qquad \hat{v}_j \;=\; \arg\max_v \, \langle \hat{z}_j / \|\hat{z}_j\|,\; z_v \rangle. \tag{2}$$

Across episodes, retrieval is an attractor read: a cue is completed against the stored set by a sharp softmax before Eq. 2 is applied.

The slow store holds the same content as a matrix. It writes by Hebb's rule, decays multiplicatively once per discrete event, and reads by running the same synapses forward:

$$W \leftarrow W + \eta \, z \, r^{\top}, \qquad W \leftarrow \gamma W, \qquad \hat{z} = W r. \tag{3}$$

It has no cleanup and no episode separation, deliberately. Two bindings of the same role interfere *coherently* — their roles overlap at cosine one, not at the root-of-d overlap of random roles — so the read returns their decay-weighted blend. In the fast store that would be a defect; here it is the mechanism by which repetition of a regularity survives while accidents decay.

One calculation. Unbinding a superposition returns the sought filler plus a sum of crosstalk terms from every other binding. For traces carrying weights w_i and near-orthogonal random roles, the retrieved cosine is

$$\cos(\hat{z}_j, z_j) \;\simeq\; \frac{w_j}{\sqrt{w_j^2 + S_j/d}}, \qquad S_j = \sum_{i \neq j} w_i^2, \tag{4}$$

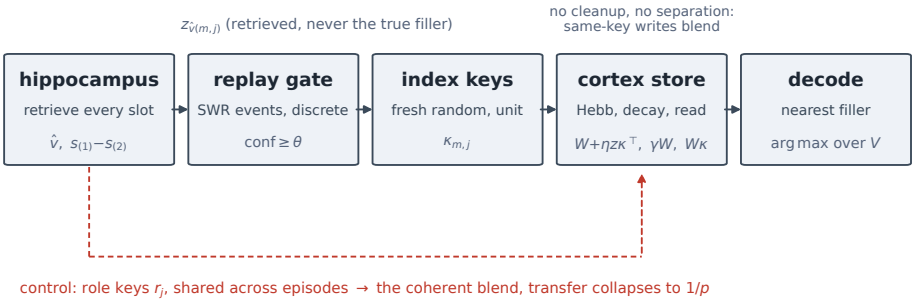

Figure 1: The slow store and the bridge that writes it. Retrievals from the fast store pass a confidence gate, are keyed by fresh index codes, and are accumulated into one Hebbian matrix. The dashed path is the control: the same retrievals keyed by the shared role vectors instead collapse to the blend.

and a decode succeeds while that signal clears the strongest of $V-1$ distractor similarities, whose median is $z_V/\sqrt{d}$ with $z_V = \Phi^{-1}\big(2^{-1/(V-1)}\big)$ — for the codebook size used throughout, 2.78. Everything the rest of this paper derives is Eq. 4 evaluated with different weights: equal weights give capacity, geometrically decaying weights give a forgetting horizon, a schedule of repeated writes gives a summed weight, a single small η gives priming, and shared roles collapse the denominator into the coherent-blend case. Nothing is fitted at any point; the only constant, z_V, is a property of the codebook size.

The bridge. Consolidation replays what the fast store retrieves into the slow one. Each retrieval carries a label-free confidence — the gap between the top two decode similarities — and only retrievals above a threshold are replayed. Keys are fresh per-event index codes rather than the roles themselves, following hippocampal indexing accounts (Teyler & DiScenna, 1986); §4 shows what happens when that choice is reversed. Figure 1 draws the path.

3 The slow store: capacity, forgetting, semanticisation

Capacity, and a marginal failure. Reading one of N undecayed pairs returns signal one against crosstalk of power $(N-1)/d$, so the store holds pairs while $1 \geq z_V\sqrt{(1+(N-1)/d)/d}$, giving $N^* = d\,(d/z_V^2 - 1) + 1$ — the per-episode capacity multiplied by the dimension, and quadratic in d where the episodic law is linear. Registered at a tenth of ratio tolerance, this *failed*: worst deviation 0.118 over 4 grid points. The miss is at the smallest dimension only, and is conservative — the store holds more than derived — so the quadratic shape stands while the small-d correction remains underived. We report it as a failure rather than adjusting the tolerance.

Forgetting is a horizon, and its edge was wrong. Under decay the age-a trace carries weight γ^a, so Eq. 4 predicts an *edge* rather than a gradient: ages up to some a^* are readable and beyond it none are. The edge's position, registered at one event of tolerance, failed decisively — worst deviation 16 events over 12 points — and the failure carried its diagnosis. The derivation had the aged trace competing only against an incoherent field, but the decode also competes against recently written fillers whose similarities fluctuate at $\sim 1/\sqrt{d}$ while the signal has decayed to γ^a. The law is therefore exact wherever $\gamma^a\sqrt{d}$ stays large and misses in proportion as it falls, which is precisely the observed pattern.

Adding the stored-competitor population turns the threshold into a product of Gaussian tails: not closed form, but still parameter-free in (γ, d, V, T). Registered afresh on a grid sharing no parameter value with the first, it *held* at 1 event over 8 points, and — the part that matters — the old law's failure signature is absent, so the diagnosis is confirmed and not merely the tolerance. Figure 2 shows both laws against the measured curves.

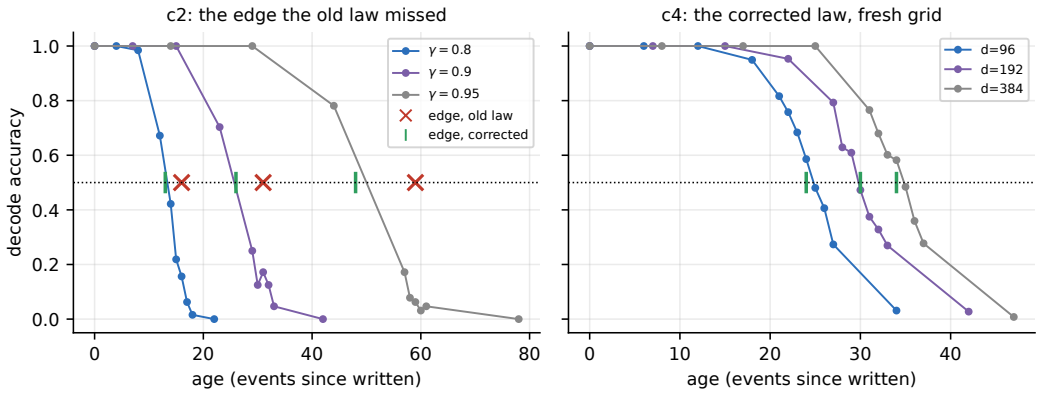

Figure 2: Forgetting is a horizon, and where its edge lies. Left: measured curves with the falsified law's predicted edges struck through, beside the corrected law's. Right: the corrected law on a grid it had never seen.

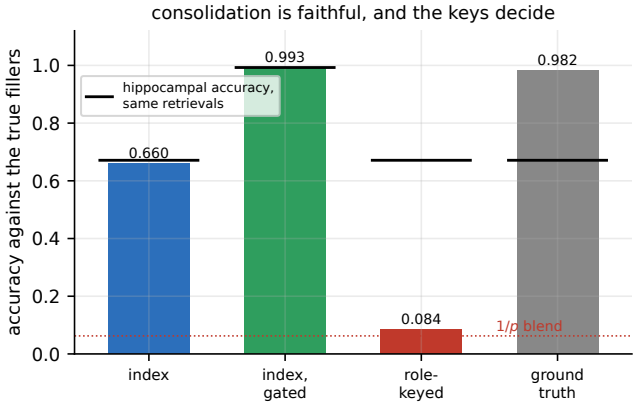

Figure 3: Consolidation is faithful, and the keys decide. Each bar is a replay arm; the tick above it is the hippocampal accuracy of the same retrievals.

Semanticisation has a derivable onset. Binding one role repeatedly to exemplars drawn at cosine ρ around a prototype, the read is closer to the prototype than to any stored exemplar once $m > (1+\rho)/\rho$. Dimension cancels from the inequality entirely, which is the bet: the crossing depends on ρ alone. It held, worst deviation 1.0 exemplar over 10 cells, with the crossing invariant across dimensions. What would be a binding failure in the fast store is, here, generalisation with a predicted onset.

4 Consolidation

The channel is faithful, and the keys decide. Index-keyed replay transfers what the fast store retrieved without loss and without repair: cortical accuracy 0.6596 against the hippocampal 0.6712 of the same retrievals, a gap of 0.0116. Gating replay by confidence carries the gain undiminished — 0.9927 against 0.9927, at coverage 0.130 — so a selection mechanism in the fast store becomes an end-to-end property of the pair. Keying the same retrievals by the *shared* role vectors instead collapses transfer to 0.0840, an eightfold loss from one change of keys with the retrievals held identical. The isolating control, replaying ground-truth fillers through the same store, retains 0.9820; hippocampal accuracy times that retention lands on the index arm's measured value to three decimals, so the channel composes multiplicatively rather than interacting (Figure 3).

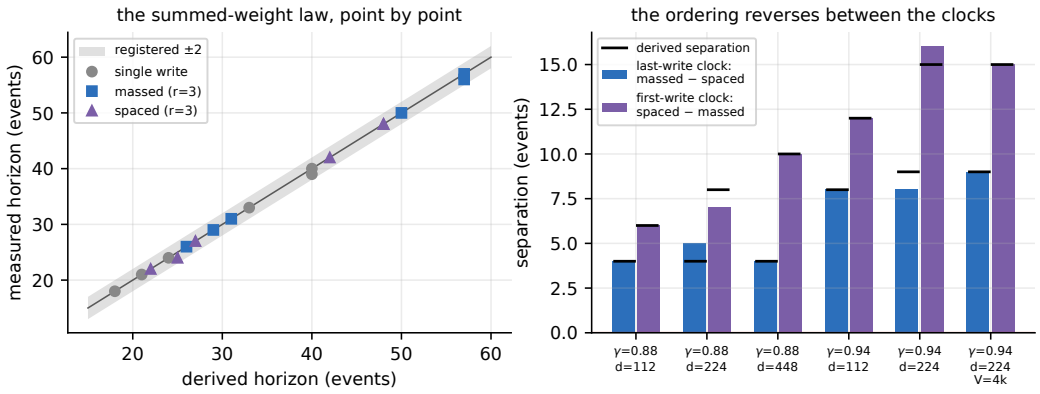

Figure 4: Replay schedules on the summed-weight law, and the two-clock reversal. Left: measured against derived horizons. Right: the separation on each clock with its derived tick.

Schedules, and a spacing effect that reverses. A trace replayed at several ages is written repeatedly to the same key with the same content, so its weights sum coherently and Eq. 4 takes $w = \sum_i \gamma^{a_i}$ unchanged. Horizons predicted this way land within 1 event over 18 scheduled horizons.

The interesting consequence concerns massed versus spaced rehearsal. At equal age since the *last* write, a massed schedule's summed weight dominates a spaced one term by term, so massed must win — the opposite of the textbook spacing effect. On the shared timeline, however, the spaced schedule's last write falls later, and that offset exceeds what the weight deficit costs, so total survival from *first* exposure orders the other way. Both orderings are consequences of the same law, and both were registered: measured 0 inversions on either clock, with minimum separations 4 and 6 events landing on the derived minima (Figure 4). The spacing effect, in this system, is a statement about which clock is being read rather than about which schedule is better.

Consolidation in time, and a second falsification repaired. With decay running and fresh content still arriving, the consolidated store forgets on a horizon of its own. Composition survives: index accuracy equals hippocampal accuracy times the truth arm's retention at every sampled age, worst gap 0.0213, across a fourfold accuracy range, with the gated arm never below the ungated one (0 inversions). The *position* of the horizon, however, failed: derived 30 events against a measured 27, a deviation of 3 events and early. The diagnosis was again in the miss. The law had spread the background's crosstalk evenly over the codebook, but each background write lands on *one* entry, so the load is skewed and the max-of-V decode is driven by the few entries holding a fresh trace. Evaluating the same expectation over the actual assignments reproduces the measured curve; re-registered on decay rates the corrected law had never seen, it gave crossings 19 and 34 at $\gamma = 0.87$ and $\gamma = 0.93$, and both were exact: deviation 0 (Figure 5). Geometry must enter these laws whole rather than compressed — a lesson that returns in §7.

5 Priming, and what survives amnesia

A weak exposure is, on this account, not a special kind of trace but an ordinary one born old: a write at weight $\eta < 1$ enters Eq. 4 at $w = \eta\gamma^a$, and nothing else about it differs. Three consequences follow and all three held. Weak writes decode at the graded probability the law derives, worst deviation 0.0292 over 24 cells sampled in the informative band — half of what measurement noise alone was budgeted, and far from the tenth-scale miss a genuine subthreshold regime would produce. The store therefore has no threshold, and priming needs none. The weak trace's horizon sits earlier than a full-weight trace's by exactly $\log \eta / \log \gamma$ events, worst deviation 1 event over 3 points: decay is a trace-independent scalar, as the homeostatic reading requires.

The third is the dissociation. With the fast store removed, explicit episodic report is at floor by construction — that half is architecture, not a finding — while the magnitude of what remains is

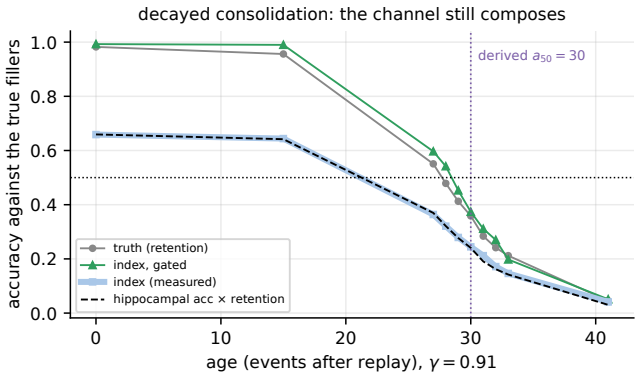

Figure 5: Decayed consolidation still composes. The dashed curve is hippocampal accuracy times measured retention, drawn over the measured index arm.

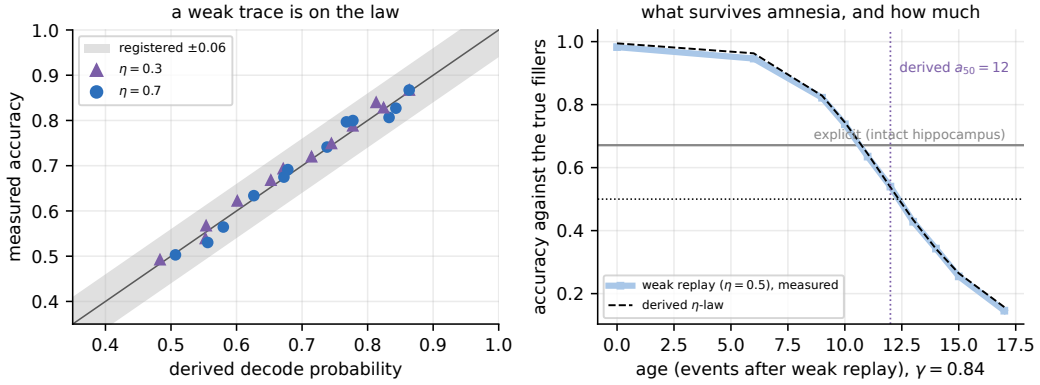

Figure 6: Priming as the law's lower reaches, and the amnesia dissociation. Right: the flat line is the explicit accuracy a lesion removes; the curve is what it leaves.

the bet. A weak replay's retained accuracy tracks the corrected retention law at $\eta\gamma^a$ to 0.0166 across ages spanning near-ceiling to near-floor, with a derived crossing at 12 events (Figure 6). The performance an amnesic system retains is, here, a number the law fixes rather than a residue.

6 Real codes, mixed codes, and the workspace

Crosstalk from spectra. Binding with a random unitary role randomises a code's Fourier phases while leaving its power spectrum intact, so a bound term's covariance is the circulant of the modality's average spectrum, and protection conjugates that covariance by the permutation. Crosstalk between two protected bound slots is therefore

$$\mathrm{Var}\left\langle \pi^i(r_i \circledast z_i),\ \pi^j(r_j \circledast z_j)\right\rangle \ =\ \mathrm{tr}\big(P^{\Delta} C_a P^{-\Delta} C_b\big), \qquad \Delta = i - j, \tag{5}$$

with C the circulant of each spectrum — computable from the codes alone. Eq. 5 held to 0.066 of ratio over 8 pairs spanning white codes to a synthetic modality concentrated 4.586 times white. The naive spectral overlap is *not* the law: it over-predicts same-spectrum crosstalk, because the protection permutation partially whitens, and Eq. 5 carries exactly that.

Why the codec is modality-blind. Real encoder codes are strongly anisotropic in their own basis, yet after centring their power spectra are nearly flat: self-overlaps 1.011 for MiniLM, 1.012 for CLIP text and 1.070 for CLIP image, against one for white. The Fourier basis is arbitrary for a learned embedding space, so anisotropy spreads across bins and binding whitens what the encoder concentrated. The decode consequence follows: per-slot accuracy in pure and mixed episodes lands within

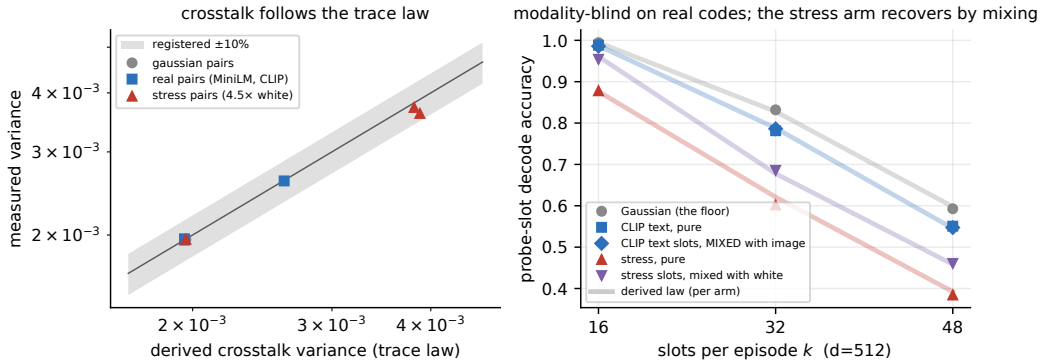

Figure 7: The codec under real and mixed codes. Left: crosstalk on the trace law across a spectral range. Right: decode accuracy against load, with mixed real markers on their pure line.

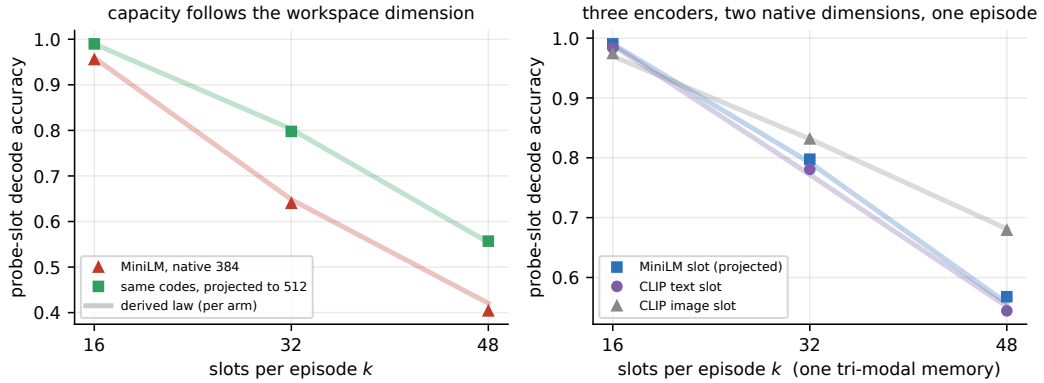

Figure 8: One dimension for every modality. Left: the same codes native and projected, each on its own law. Right: the tri-modal episode, every probe on the law.

0.0140 of the derived law over 15 cells, including episodes carrying CLIP text and CLIP image slots together. The synthetic stress modality is what makes this evidence rather than insensitivity: at 4.586 times white it loses tens of points in pure episodes and *recovers* by the derived margin when mixed with white codes, tracked to 0.0193 over 9 cells (Figure 7). Mixing is relief, with a predicted sign and size.

A workspace, and where capacity lives. MiniLM and CLIP produce codes of different dimension, so no episode can carry both until they share a space. A fixed random projection per modality — never trained — suffices. Geometry survives it on the Johnson–Lindenstrauss law (Johnson & Lindenstrauss, 1984): both registered cosine statistics land within 0.034 relative in both directions, up-projection and down. With that, MiniLM, CLIP text and CLIP image slots bind into one episode and decode on the existing law with no new constant, worst deviation 0.0107 over 9 tri-modal cells.

The projection also settles where capacity lives. Because a role spreads crosstalk over the dimension of the *episode*, not of the source, projecting content up buys real binding capacity: the same MiniLM codes, native and up-projected, each track their own derived law to 0.0167 over 6 cells, and differ by 0.1530 at the deepest load (Figure 8). Capacity is a property of the workspace, not of the encoder that filled it.

7 The Full Stack Adds Nothing

The results above measure the codec without the attractor store. Running the tri-modal codebook through the complete circuit tests two things at once, and they resolve in opposite directions.

Capacity stated in terms of a single compressed geometry statistic — the empirical threshold z_{emp}, which is 4.00 for the tri-modal pool against 2.92 for a Gaussian one, implying a twofold spread of capacities — *failed*: measured capacity is flat across every arm, worst ratio deviation 1.242 over 3 arms. The pool's near-duplicate tail inflates a median-of-maxima statistic without moving the bulk of the distractor field, so the one-number law overcorrects exactly where structure concentrates. This is the same lesson as the decayed horizon: geometry must enter whole.

The law that keeps the whole distractor geometry, meanwhile, *held*: the full circuit's end-to-end accuracy equals the storeless calculation at every load, worst deviation 0.0093 over 12 cells. The attractor read, which is the one stage every earlier derivation treated as exact, adds nothing on real tri-modal content — so the corrected capacity law is that same held curve read at one half, and the system's end-to-end behaviour is the product of parts derived independently. The practical consequence is worth stating plainly: real multimodal codes cost about three percent of capacity through the full stack, not the near-half the compressed statistic predicted.

8 Related work

Binding without training. Representing structure by binding and superposition in a fixed vector space is the vector-symbolic tradition: tensor-product representations (Smolensky, 1990), holographic reduced representations (Plate, 1995; 2003), hyperdimensional computing (Kanerva, 2009), and their modern surveys and capacity analyses (Kleyko et al., 2022; Frady et al., 2018). We use the standard operators; what is new here is not the algebra but that a whole memory *system* — two stores, a gate and a consolidation policy — is specified in it and then predicted from it in advance, on frozen encoder codes rather than random ones.

Attractor retrieval. The completion step is a modern Hopfield read (Ramsauer et al., 2021), whose dense-associative predecessors (Krotov & Hopfield, 2016) extend the classical network (Hopfield, 1982). Our finding in §7 is a negative one about this component in this setting: at every load we measured, the attractor read contributes nothing beyond what a storeless calculation predicts.

Complementary learning systems. The division of labour between a fast, sparse, pattern-separating store and a slow, overlapping, statistical one is the CLS account (McClelland et al., 1995; Kumaran et al., 2016), itself built on earlier hippocampal theory (Marr, 1971; Treves & Rolls, 1994). Consolidation by replay of retrieved episodes is the evidenced mechanism (Wilson & McNaughton, 1994; Girardeau et al., 2009), and keying by episode rather than by content is hippocampal indexing (Teyler & DiScenna, 1986). We instantiate all three without training and derive their consequences; the motivating problem, catastrophic interference in a single store (McCloskey & Cohen, 1989), appears here as an interference term with a closed form.

Memory phenomena as predictions. The behavioural regularities we predict have long empirical literatures: distributed practice (Ebbinghaus, 1885; Cepeda et al., 2006), serial position (Murdock, 1962; 1982), and the preserved priming of amnesic patients (Warrington & Weiskrantz, 1970; Tulving & Schacter, 1990). Our claim is narrow and specific: in this system these appear as consequences of one interference calculation, with magnitudes fixed in advance rather than fitted. The plasticity and decay rules are the classical local ones (Hebb, 1949; Anderson, 1972; Kohonen, 1972; Turrigiano, 2008).

Learned memory, and frozen encoders. Differentiable memories learn their read and write policies (Graves et al., 2016), as do fast-weight and linear-attention variants (Schlag et al., 2021). Our system trains none of them, which costs adaptivity and buys predictability. The content comes from standard frozen encoders (Reimers & Gurevych, 2019; Radford et al., 2021), used strictly as front ends.

9 Discussion and limitations

Three limitations bound these results, and we state them plainly.

The visual stimuli are procedurally rendered rather than natural. The encoder is real and frozen and the codes are genuinely visual, but the images are synthetic, and the honest label is *real frozen encoder, synthetic stimuli*. Whether a natural-image codebook, with its heavier near-duplicate tail, still satisfies Eq. 5 is open, and §7 shows exactly which statistic would be at risk.

The system is a memory, not an agent. It has no policy, no objective and no mechanism for deciding what is worth storing beyond the confidence gate — and that gate selects by retrieval confidence, whereas the biology it is drawn from selects by novelty and reward. We test selective replay's consequence, a cleaner store, not its criterion.

Finally, the evaluation is probe accuracy under controlled interference, not a downstream task. That is a deliberate trade: it is what makes each measurement attributable to a term in Eq. 4, and it is why a failure such as the forgetting edge could be diagnosed rather than merely observed. It is also the clearest direction for the next step.

What the four failures taught is more useful than any single confirmation. Two of them — the forgetting edge and the decayed horizon — failed because a population of competitors had been summarised rather than counted, and both were repaired by putting the omitted population back. The third, the compressed capacity statistic, failed the same way at the level of the whole codebook. The recurring lesson is that these laws tolerate no compression of the interference geometry: the calculation is exact when it counts every competitor and wrong in a diagnosable direction when it does not. That is a property one can only discover by writing the prediction down first.

Reproducibility Statement

Every number in this paper is a generated macro read from a committed results file; no measured value is typed into the prose, and a check in the build fails the paper if one is. Every prediction was registered with its value, comparator and date, hashed into a lock file, in a commit that precedes the commit carrying the results it is compared against, so the ordering is auditable rather than asserted. Each claim names the test that protects it, and settled claims — including falsified ones — retain a test pinning the outcome, so reinstating a claim requires a deliberate change. Sweep cells are cached under a key combining the configuration hash and a digest of the source files that determine the value, so a change to a model file invalidates exactly the cells it can affect. Filler codes come from frozen encoders read from an on-disk cache whose key includes the model identifier and checkpoint commit; the code, the registered predictions and the results files accompany the submission.

Use of Large Language Models

The author used a large language model as a research tool throughout this work: implementing experiment code, carrying out derivations that were then checked against simulation, proposing candidate predictions, and drafting the manuscript. The registration protocol of §1 was applied identically whatever a prediction's source, every reported number is generated from committed data by the accompanying code, and the author is responsible for all content.

References

James A. Anderson. A simple neural network generating an interactive memory. *Mathematical Biosciences*, 14(3–4):197–220, 1972.

Nicholas J. Cepeda, Harold Pashler, Edward Vul, John T. Wixted, and Doug Rohrer. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132 (3):354–380, 2006.

Hermann Ebbinghaus. *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Duncker & Humblot, 1885.

E. Paxon Frady, Denis Kleyko, and Friedrich T. Sommer. A theory of sequence indexing and working memory in recurrent neural networks. *Neural Computation*, 30(6):1449–1513, 2018.

Gabrielle Girardeau, Karim Benchenane, Sidney I. Wiener, György Buzsáki, and Michael B. Zugaro. Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, 12 (10):1222–1223, 2009.

Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, et al. Hybrid computing using a neural network with dynamic external memory. *Nature*, 538 (7626):471–476, 2016.

Donald O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Wiley, 1949.

John J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

Pentti Kanerva. Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Computation*, 1(2):139–159, 2009.

Denis Kleyko, Dmitri A. Rachkovskij, Evgeny Osipov, and Abbas Rahimi. A survey on hyperdimensional computing aka vector symbolic architectures, part I: Models and data transformations. *ACM Computing Surveys*, 55(6), 2022.

Teuvo Kohonen. Correlation matrix memories. *IEEE Transactions on Computers*, C-21(4):353–359, 1972.

Dmitry Krotov and John J. Hopfield. Dense associative memory for pattern recognition. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

Dharshan Kumaran, Demis Hassabis, and James L. McClelland. What learning systems do intelligent agents need? complementary learning systems theory updated. *Trends in Cognitive Sciences*, 20(7):512–534, 2016.

David Marr. Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 262(841):23–81, 1971.

James L. McClelland, Bruce L. McNaughton, and Randall C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995.

Michael McCloskey and Neal J. Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In Gordon H. Bower (ed.), *Psychology of Learning and Motivation*, volume 24, pp. 109–165. Academic Press, New York, 1989.

Bennet B. Murdock. The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5):482–488, 1962.

Bennet B. Murdock. A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89(6):609–626, 1982.

Tony A. Plate. Holographic reduced representations. *IEEE Transactions on Neural Networks*, 6(3): 623–641, 1995.

Tony A. Plate. *Holographic Reduced Representation: Distributed Representation for Cognitive Structures*. CSLI Publications, 2003.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning (ICML)*, volume 139, pp. 8748–8763, 2021.

Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Thomas Adler, Lukas Gruber, Markus Holzleitner, Milena Pavlović, Geir Kjetil Sandve, Victor Greiff, David Kreil, Michael Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. Hopfield networks is all you need. In *International Conference on Learning Representations (ICLR)*, 2021.

Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP)*, pp. 3982–3992, 2019.

Imanol Schlag, Kazuki Irie, and Jürgen Schmidhuber. Linear transformers are secretly fast weight programmers. In *International Conference on Machine Learning (ICML)*, volume 139, pp. 9355–9366, 2021.

Paul Smolensky. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, 46(1–2):159–216, 1990.

Timothy J. Teyler and Pascal DiScenna. The hippocampal memory indexing theory. *Behavioral Neuroscience*, 100(2):147–154, 1986.

Alessandro Treves and Edmund T. Rolls. Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4(3):374–391, 1994.

Endel Tulving and Daniel L. Schacter. Priming and human memory systems. *Science*, 247(4940): 301–306, 1990.

Gina G. Turrigiano. The self-tuning neuron: Synaptic scaling of excitatory synapses. *Cell*, 135(3): 422–435, 2008.

Elizabeth K. Warrington and Lawrence Weiskrantz. Amnesic syndrome: Consolidation or retrieval? *Nature*, 228(5272):628–630, 1970.

Matthew A. Wilson and Bruce L. McNaughton. Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172):676–679, 1994.

A THE REGISTER OF PREDICTIONS

Table 1 lists every prediction this paper reports, as registered: its identifier in the accompanying code, the one-sentence claim, the tolerance and comparator it was locked with, the date of registration, and the outcome. The four that failed are marked; each is discussed where it arises in the text, together with the diagnosis it carried and, for three of them, the corrected law re-registered on a fresh grid. Every row is generated from the same claim registry the lock file hashes, so the table cannot drift from what was registered.

Table 1: The 23 pre-registered predictions and their outcomes, as registered.

claim	registered prediction	bound	registered	outcome
COR-CAP-01	The cortical store's capacity is the hippocampal per-episode law times the dimension: a stored pair decodes until $N^* = d\,(d/z_V^2 - 1) + 1$ pairs are superposed, with the same threshold constant z_V and nothing fitted.	$\leq$ 0.1 ratio dev	2026-08-30	**failed**
COR-FOR-01	Forgetting under passive decay is a horizon, not a gradient: with per-event decay γ and one pair per event, the number of retrievable recent events is $\#\{a : \gamma^{2a}(d/z_V^2 - 1) \geq S/d\}$ with $S = \sum_i \gamma^{2i}$, fixed by (γ, d, V) with nothing fitted.	$\leq$ 1 events	2026-08-30	**failed**

claim	registered prediction	bound	registered	outcome
COR-FOR-02	The corrected horizon law holds: retrievability of age a requires beating both the incoherent field over the unstored codes and the $T-1$ stored competitors, whose fillers carry the coherent term COR-FOR-01 missed, giving $a^* = \max\{a : (V-T)\log\Phi(\gamma^a/\sigma_u) + \sum_j \log\Phi(\gamma^a/\sigma_j) \geq \log\frac{1}{2}\}$ with everything fixed by (γ, d, V, T).	$\leq$ 2 events	2026-08-30	held
COR-CON-01	Index-keyed replay consolidates faithfully: the cortical store's accuracy against the TRUE fillers equals the hippocampal retrieval accuracy of what was replayed – the cortex neither cleans up nor loses, because the matrix retention at this load is derivably at ceiling.	$\leq$ 0.02 accuracy	2026-08-30	held
COR-CON-02	Selective replay works end to end: gating replay by M8's label-free confidence carries the hippocampal gate's accuracy gain into the cortical store undiminished.	$\leq$ 0.02 accuracy	2026-08-30	held
COR-CON-03	The keying is the division of labour: replaying the same retrievals keyed by the SHARED role vectors instead of per-event index codes collapses episodic recall to the blend – at or below $2/p$ – because every episode's binding lands on one key and the read returns their prototype, not any member.	$\leq$ 0.125 accuracy	2026-08-30	held
COR-SEM-01	Semanticisation has a derivable onset: rebinding one role to m exemplars drawn at cosine ρ around a prototype, the read is closer to the prototype than to any stored exemplar once $m > (1+\rho)/\rho$, and the crossing does not move with dimension.	$\leq$ 1 exemplars	2026-08-30	held
COR-CON-04	A replay schedule moves the forgetting horizon exactly as the summed-weight law says: a trace replayed at ages a_i carries weight $\sum_i \gamma^{a_i}$ – same key, same filler, so the writes sum coherently – and that weight drops into the corrected horizon law unchanged.	$\leq$ 2 events	2026-09-02	held
COR-CON-05	Massed and spaced replay order both ways, by the clock: at equal replay budget, massed outlasts spaced measured from the last write, and spaced outlasts massed measured from the first – the spacing effect is a statement about which clock, not about which schedule is better.	$=$ 0 inversions	2026-09-02	held
COR-CON-06	Decayed consolidation composes: at every sampled age, the index arm's end-to-end accuracy equals the hippocampal retrieval accuracy times the truth arm's measured retention – decay adds interference to the channel, not interaction between the stages.	$\leq$ 0.05 accuracy	2026-09-02	held
COR-CON-07	The consolidated store forgets on a horizon fixed by $(\gamma, \mathrm{load}, d, V)$: under ongoing on-vocabulary writing, the truth arm's retention crosses one half at the age the per-entry retention law derives, with nothing fitted.	$\leq$ 2 events	2026-09-02	**failed**

claim	registered prediction	bound	registered	outcome
COR-CON-08	The corrected retention law holds where the falsified one missed: with the background ladder's crosstalk assigned to entries as it actually lands – skewed, a few entries carrying a fresh trace – the consolidated store's half-retention crossing follows $(\gamma, \mathrm{load}, d, V)$ with nothing fitted, on decay rates the law has never seen.	≤ 2 events	2026-09-02	held
COR-PRI-01	A weak exposure is a real trace on the decode law: a single write at weight $\eta < 1$ decodes with exactly the graded probability the law derives for the weight $\eta\gamma^a$ – there is no separate subthreshold regime, and priming is the law's lower reaches.	$\leq$ 0.06 accuracy	2026-09-03	held
COR-PRI-02	Priming has a horizon, and it sits where the decayed weight says: a weight-η trace's last retrievable age lies earlier than the full-weight horizon by $\log \eta / \log \gamma$ events – the weak trace is not fragile, just born old.	≤ 2 events	2026-09-03	held
COR-PRI-03	Priming survives amnesia at the derived magnitude: a weak replay through the bridge leaves a cortical trace that decodes at exactly the level the corrected retention law derives for weight $\eta\gamma^a$ – the performance a system retains when explicit retrieval is gone is a number the law fixes, not a residue.	$\leq$ 0.05 accuracy	2026-09-03	held
EC-MIX-01	Crosstalk between protected bound slots follows the trace law: binding with a random unitary role randomizes a code's Fourier phases and keeps its power spectrum, protection conjugates by the permutation, and the pairwise variance is $\mathrm{tr}(P^{\Delta} C_a P^{-\Delta} C_b)$ with C the circulant of the modality's average spectrum – nothing fitted, across pairs spanning white to a 4.5-fold spectral concentration.	$\leq$ 0.1 ratio	2026-09-03	held
EC-MIX-02	The codec is modality-blind on real codes, and the s-matrix says why: per-slot decode accuracy in pure and MIXED frozen-encoder episodes – MiniLM beside Gaussian, CLIP text beside CLIP image – lands on the derived spectral decode law, whose inputs (spectra within seven percent of white) predict that mixing costs nothing measurable.	$\leq$ 0.05 accuracy	2026-09-03	held
EC-MIX-03	The same law where the effect is large: a stress modality concentrated to 4.5x white loses exactly the derived accuracy in pure episodes, and mixing it with white RECOVERS its slots by the derived margin – mixture as relief, sign and size from the law, which is what makes the real-code agreement evidence rather than insensitivity.	$\leq$ 0.05 accuracy	2026-09-03	held
ASSOC-01	The workspace projection preserves codebook geometry on the Johnson-Lindenstrauss law: a fixed random projection into the workspace dimension concentrates every pairwise cosine on its raw value with $1/\sqrt{d}$ jitter, so the projected pool's mean absolute cosine and empirical threshold z_{emp} are derivable from the raw gram – in both directions, MiniLM up to 512 and CLIP text down to 384.	$\leq$ 0.05 relative	2026-09-03	held

claim	registered prediction	bound	registered	outcome
ASSOC-02	The workspace makes the impossible episode possible, on the existing law: MiniLM slots (projected from 384), CLIP text slots and CLIP image slots bound into ONE 512-dimensional episode decode at the accuracy the spectral law derives for the projected pools – tri-modal memory with no new mechanism and no new constant.	$\leq$ 0.05 accuracy	2026-09-03	held
ASSOC-03	Binding capacity lives in the workspace dimension, not the source's: up-projected MiniLM codes at $d = 512$ decode on the 512 law – better than their native 384 selves at every load past the ceiling, by the derived margin – because the role spreads crosstalk over the episode's dimension, wherever the content came from.	$\leq$ 0.05 accuracy	2026-09-03	held
MIX-STORE-01	Through the live DG-CA3-CA1 stack, the tri-modal union pool holds exactly the capacity its own empirical threshold derives: measured k^* follows $\alpha\, d/z^2_{\mathrm{emp}}$ with one constant calibrated on the Gaussian identity arm alone – the 440 Part II law, now with three encoders in one codebook and the engine's own circuit.	$\leq$ 0.15 ratio	2026-09-03	**failed**
MIX-STORE-02	The stack adds nothing the codec did not: end-to-end accuracy of the full circuit on identity-mode arms equals the storeless codec law on the same substrate at every load – M6's finding that CA3 never fails, carried to real tri-modal codes, end to end.	$\leq$ 0.05 accuracy	2026-09-03	held