Drosophila Connectome Topology Provides No Measurable Language-Model Advantage: A Preregistered Negative Experimental Program

Vladislav Matyukhin 1,*

1 Independent researcher.

Correspondence: dagterevvladislav@gmail.com

ORCID: 0009-0008-7319-9336

* Corresponding author.

Abstract

Whether the synaptic wiring of an animal brain can serve as an inductive bias for language models is an empirical question, not a metaphor. We tested that question in a preregistered experimental program that transferred *Drosophila melanogaster* anatomy—and, later, a *Drosophila*-shaped memory interface—into small language-model and memory computers. Win rules, controls, and seed counts were written before each queue. Across the closed tests, no protocol showed that a fly-derived operator reduced pair-token negative log-likelihood (NLL) or cross-entropy relative to both a no-graph baseline and a non-brain control at the preregistered margin.

A washed diffusion prior built from top-degree FlyWire nodes never beat scrambled or random graphs on lexical, factual, or morphological probes and is architecturally near-identity after a forced unit diagonal. Raw syntax kernels of size 64 × 12 and 256 × 12 beat a non-brain graph on holdout area under the receiver operating characteristic (AUROC) on three of three seeds, and the 64 × 12 kernel beat the same control on TinyLM pair-NLL on three of three seeds, but never beat the no-kernel model by the required margin of 0.02. Replacing the TinyLM feed-forward block with a frozen antennal-lobe projection neuron → Kenyon-cell expansion and hard *k*-winner-take-all was worse than a dense feed-forward network and worse than Erdős–Rényi and degree-matched expansions on three of three seeds. A trainable projectome residual first failed to enter computation under a degenerate zero initialization. After a non-degenerate reparameterization the communication channel was used (relative residual norm $\rho \approx 0.02$–0.06), yet fly topology won Quality, Anatomy, and Specificity on zero of three seeds. Using the same expansion as a retrieval hash lost to random expansion plus the same winner-take-all (primary recall@1 at noise 0.2 ≈ 0.31 versus 0.37) and to dense cosine lookup (≈ 0.85). Isolated central-complex heading persistence did not meet preregistered attractor criteria. A silent-engram interface is realizable by construction; a learned recovery operator failed already on training data; a family-bank expression computer that beat weak controls was matched by a generic constrained selector (mean success 0.80 on all learned arms, gap 0, five seeds).

We conclude that trainable family-level sparse communication is viable, but the *Drosophila* projectome topology provides no measurable pair-NLL advantage over vanilla, random or degree-matched topology, or learned sparsity under the C1-R interface. Mushroom-body expansion is not a language feed-forward substitute and is not a superior associative hash in these tests. The *Drosophila*-specific silent-engram claim does not survive a matched constrained control. An 800-million-parameter model's Russian multiple-choice score is not a connectome result: no same-corpus ablation without the blueprint was run. Efficiency, latency, and training cost were not measured and are not claimed.

Keywords: connectome; *Drosophila melanogaster*; language models; mixture of experts; inductive bias; negative results; mushroom body; silent engram; projectome; preregistration

1. Introduction

Whole-brain wiring diagrams now exist for an adult fruit fly [1,2]. That fact has revived an old hope in machine learning: that evolution has already searched a space of sparse, modular, and typed communication architectures, and that copying or constraining an artificial network with those constraints will improve

learning. The hope is not empty in neuroscience. Connectome-constrained models of the fly visual system can predict neural activity when unknown biophysical parameters are optimized for a task [3]. Theory also warns that a connectome often fails to pin down dynamics: one wiring diagram is compatible with many computations [4]. In computer science, the fly olfactory expansion followed by winner-take-all (WTA) is a locality-sensitive hash for retrieval, not a ready-made language transform [5].

These three results already suggest a discipline. A measured adjacency matrix is a constraint, not a drop-in weight tensor. Task optimization still has to do work. Retrieval is not generation. None of that discipline was present in the first engineering transfer we study here: a production mixture-of-experts (MoE) language model whose router prior was a washed diffusion kernel over 256 high-degree FlyWire nodes, later aggregated onto 32 or 128 experts, with the diagonal of the kernel forced to one. That construction is close to the identity and cannot be read as a *Drosophila* verdict.

The scientific question is nevertheless well posed. Does any well-specified use of fly wiring—or of a fly-shaped memory interface that is *not* a mask—improve a language-adjacent objective relative to strong non-brain controls, under a rule written before the run?

We report a closed experimental program that asked that question in stages. Early queues tested washed and raw graph priors as attention or router biases in TinyLM. Later queues replaced the feed-forward network (FFN) with a frozen biological expansion, then tested a trainable projectome residual whose biological adjacency was a mask rather than a weight. After those language-model (LM) transfers failed or failed to activate, the program moved the same anatomy into an associative-hash computer, then left anatomy and tested a silent-engram interface motivated by a 2026 mushroom-body physiology result: forgotten traces can remain silent in a neutral mushroom-body output neuron and become expressible only through a separate dopaminergic reminder pathway, which can also create a false memory under a misleading reminder [6]. That biology is a claim about storage versus accessibility. It is not a connectome mask.

This paper is a measurement paper. “Better” means a preregistered inequality on a named metric, a named control, a named seed count, and a named margin. Suggestive single-seed gaps are reported as such. Incomplete queues are reported as incomplete, not as anatomy failures. The 800-million-parameter production model is described for context and is not an experimental arm of any connectome ablation.

The locked reading of the program is as follows. Trainable family-level sparse communication can enter the forward pass. The *Drosophila* projectome topology provides no measurable pair-NLL advantage over vanilla, random or degree-matched topology, or learned sparsity under that interface. Frozen antennal-lobe projection neuron (ALPN) → Kenyon-cell (KC) expansion is a worse language FFN than random sparsity and a worse retrieval hash than random expansion plus the same WTA. The *Drosophila*-specific silent-engram claim from a weaker family-bank computer is closed by a matched constrained selector. The research direction “is there any computational structure in this brain that could become an inductive bias?” is not answered by these implementations. The implementations that were tested are answered.

2. Related Work

2.1. Fly wiring as a scientific resource

The FlyWire reconstruction of an adult female *Drosophila melanogaster* brain reports on the order of 139,000 neurons and 5 × 10^7 chemical synapses, with cell-class and neurotransmitter annotations and a derived projectome [1,2]. We use the public v783 connectivity table. That resource was built to explain fly computation, not to initialize transformers.

2.2. Connectome-constrained models

Lappalainen et al. constructed a differentiable network with the measured connectivity of 64 cell types in the fly motion pathway, left single-neuron and single-synapse parameters free, and optimized them for motion detection [3]. Predictions agreed with a large body of physiology. The lesson they draw is the one we adopt: connectivity is a constraint; parameters are fit to a task. Beiran and Litwin-Kumar analyzed student–teacher

networks that share weights but differ in biophysical parameters and showed that a connectome often does not substantially constrain recurrent dynamics unless additional activity measurements remove degeneracy [4]. Inserting measured synapse counts into an unrelated architecture and an unrelated objective is the opposite of that program.

2.3. Expansion and winner-take-all as hashing

Dasgupta, Stevens, and Navlakha treated the antennal-lobe → mushroom-body expansion plus WTA as a locality-sensitive hash [5]. The algorithmic object is a tag for nearest-neighbor search. It is not a theorem that the same matrix, used as a frozen FFN or as a language-model expert map, will reduce next-token loss.

2.4. Silent traces and reconstructive recovery

Yang et al. showed that forgotten aversive memories in *Drosophila* persist as silent traces in behaviorally neutral mushroom-body output neurons [6]. A reminder recovers the original, behavior-guiding trace. A misleading reminder can produce a false memory. True recovery and false formation are gated by separable dopaminergic pathways. The computational content we take from that paper is the interface: storage is not accessibility; recovery is a separate pathway; a wrong reminder is a first-class failure mode. We do not transfer dopaminergic weights, and we do not treat a connectome mask as a silent engram.

2.5. Brain-inspired language-model memory

Memory-augmented networks already implement external writes, forget gates, and fast weights [7–9]. Those mechanisms do not by themselves implement a silent bank that is unreadable by key similarity and is restored only by a cue pathway. That distinction matters for claims of biological novelty. It does not license a fly uniqueness claim until a generic computer with the same output rights fails the same bars.

3. Program Design

3.1. Object and non-object

The production model Mara-MoE-800M is a 22-layer MoE transformer with model width 2112, 16 heads, 128 experts of hidden width 24, eight experts active, vocabulary 49,152 (SmolLM2 byte-pair encoding [10]), and a connectome blueprint that maps 256 nodes to 128 experts. A later continued-pretraining checkpoint used a Cyrillic-expanded tokenizer (51,260 types) and about 9.3 × 10^9 tokens. No experiment in this paper compares that 800-million-parameter checkpoint to an otherwise identical run without the blueprint. No ironclad queue trained or modified that checkpoint. Statements about 800-million-parameter quality (Section 5.8) are therefore not connectome ablations.

Closed stages of the program are summarized in Figure 1. The experimental computers are small. TinyLM language tests use width 256, four layers, eight heads, sequence length 256, 1500 AdamW steps, and learning rate 3 × 10^−4 unless a queue states otherwise. Circuit tests use typed FlyWire subgraphs (optic lobe, central complex, mushroom body, descending neurons, antennal lobe) without a language-model loss. Memory tests use frozen factor embeddings and small encoders; they have no next-token objective.

3.2. Graph construction

Edges come from FlyWire v783 (`Presynaptic_Index`, `Postsynaptic_Index`, `Connectivity`). Edge weight is log(1 + Connectivity). Early production priors selected top-degree nodes, built a column-normalized adjacency A, formed the diffusion kernel $K = A + 0.5\,A^2$, forced $\mathrm{diag}(K) = 1$, and row-normalized. Aggregation from 256 nodes onto 32 or 128 experts was a block average followed by another row-normalization. The forced unit diagonal is an architectural error, not a biological claim: after this wash the prior is close to the identity. Late syntax, plant, and routing queues disable that diagonal fill. Frozen expansion tests do not use the top-degree soup: they take class-labeled ALPN → Kenyon-cell edges with Connectivity ≥ 1, then the top-degree 256 projection neurons and 1024 Kenyon cells (6956 nonzero weights; median row occupancy 7; no empty Kenyon-cell rows).

3.3. Controls

Three different objects were called “shuffle” on different days. They are not interchangeable.

1. Degree-preserving rewiring of the fly graph, then the same kernel. This is still a fly graph. Comparing real versus this rewire is A versus A.

2. An Erdős–Rényi-like random graph of the same size.

3. A non-brain source with the same dimension and sparsity, sometimes degree-matched.

Early syntax and TinyLM queues used (1). Closed claims in this paper that concern anatomy use (2), (3), a no-map or no-kernel arm, a dense FFN, or a learned-sparse topology, as named per queue. Degree-preserving rewire is retained only as a diagnostic.

3.4. Decision rules

A language-model win requires, on a preregistered seed set, that the fly arm is better than the no-graph arm *and* better than every named non-brain control, usually by a margin of 0.02 in pair-token NLL or by a named accuracy gap. Pair-token NLL is computed on Universal Dependencies pairs (typically `nsubj`, `amod`, and later `nmod`) from SynTagRus [11,17]. Full cross-entropy is logged and does not judge. A queue with range of pair-NLL across arms below 0.01 is fatal for that seed: the head is too weak to read anatomy, and the seed is not a fly failure.

Circuit and memory wins are queue-specific and are stated with the result. Binary judges are accompanied by signed deltas against each control. A gap of 0.0001 is not an effect.

3.5. What was not measured

No queue measured 800-million-parameter inference speed, training cost, tokens per second, latency, peak memory, out-of-distribution robustness, long-context recall, or continual-learning forgetting. A possible advantage of sparse biological routing on those axes is untested. The sentence “the connectome did not make the language model better” refers only to the quality metrics that were judged.

3.6. Invalid tests, not negative anatomy

Several early tests are broken instruments. They are listed so that they are not recycled as fly failures.

- Washed Mara prior: unit diagonal plus row-normalization. A null on that prior is not a null on *Drosophila*.
- Mushroom-body association with every dopaminergic and output neuron driven to 0.8: both output channels saturate near 0.99 and 0.98. Association cannot be read.
- Central-complex “memory” after a full gold write into the state that a linear probe later reads: the probe reads the write, not autonomous integration (`fatal_sanity = true`).
- A four-way heading probe whose southern class is empty on some seeds: a class hole, not anatomy.
- Central-complex residual in TinyLM with six examples per seed: underpowered, not a language-model verdict.

4. Materials and Methods

4.1. Data sources

Connectomic edges and class labels are from FlyWire v783 and the associated atlas [1,2]. Language-graph probes use Cultura-derived same-sentence names [12], RuWordNet 2.0 lexical relations [13], NEREL 1.1 person, organization, and place mentions [14], Wikidata capital and selected relational pairs [15], RuWordNet derivational morphology, and gold Universal Dependencies relations from SynTagRus [11]. Mel’čuk-style probes used SynTagRus trees plus closed OPER/MAGN lists from the Russian National Corpus; they are not a licensed lexical-function layer of SynTagRus. Associative-memory keys and silent-engram payloads are synthetic and are generated by the released protocol scripts.

4.2. Graph-to-probe protocol (QKQL)

For each candidate kernel, a linear probe Q is trained to score entity pairs. Twenty percent of pairs are held out. Negatives are matched. Q is retrained on the scrambled or non-brain map. The probe win rule is holdout AUROC(real) − AUROC(control) > 0.05. This is not next-token loss and is not the 800-million-parameter model.

4.3. TinyLM prior injection (plant)

Configs that first passed the non-brain AUROC gate (64 × 12 and 256 × 12 raw kernels) were injected as attention bias in TinyLM. The tokenizer is the production merge-32 tokenizer. The diagonal fill is off. The judge is pair-token NLL on 96,972 pairs for plant-1 and 128,865 pairs for plant-2 (the latter adds `nmod`). A seed wins if fly < none − 0.02 and fly < non-brain − 0.02. Ironclad plant-1 requires three of three seeds. Plant-2 uses a learnable strength (initialized at 0.2, clamped to [0, 2]) and an auxiliary alignment loss; ironclad plant-2 requires at least four of five seeds. An infer-only diagnostic evaluates a finished no-kernel checkpoint under a foreign kernel and does not judge.

4.4. Frozen expansion as FFN (BSEG-T1)

The TinyLM FFN is replaced by a frozen ALPN → Kenyon-cell lift, hard k-WTA with $k = \max(8, \lfloor 0.05 \times 1024 \rfloor) = 51$, a scale and bias only on live Kenyon cells, and a tied contract back to model width. Attention is standard. There is no central complex, no dopaminergic neuron, no projectome MoE, and no router prior. Arms share one AdamW run: fly, same-nnz Erdős–Rényi, same-nnz degree-matched, and a dense 4× FFN. Anatomy requires fly < Erdős–Rényi − 0.02 and fly < degree-match − 0.02 on three of three seeds. Quality requires fly < dense − 0.02 on three of three. Ironclad is both. Expand-arm trainable parameter count is 19,262,208; dense is 21,351,168.

4.5. Projectome residual (C1 and C1-R)

Five expert families correspond to optic, central-complex, mushroom-body, descending, and antennal-lobe blocks. Family residuals communicate as

$$r' = r + \alpha F\_\{M \odot W\}(r),$$

where M is a 5 × 5 projectome mask, W holds allowed-edge weights, and α is a scalar gate. Vanilla is the exact submodel at α = 0.

C1 initialized α = 0 and $W = 0$. That joint origin is first-order degenerate: $\partial L/\partial \alpha \propto F\_W(r) = 0$ at $W = 0$ and $\partial L/\partial W \propto \alpha = 0$ at α = 0. Logged α remained exactly 0.0 on every topology arm and seed. C1 therefore does not test anatomy.

C1-R is a separate experiment, not a patch. Allowed weights are initialized to masked random values with fan-in and root-mean-square matching. An interface assay (D0), without a language-model loss, verified: α = 0 reproduces vanilla logits bit for bit; $\partial L/\partial \alpha \neq 0$ at initialization; the first AdamW step moves α; forbidden mask entries have zero gradient and zero update; and a synthetic family-transport task is solved by a cross-family edge (accuracy 1.0) but not by self-loops (accuracy 0.465, chance). Seeds are 20260921, 20260922, and 20260923. The judge is the same pair-NLL, 1500 steps, fatal range 0.01, and Quality / Anatomy / Specificity margins of 0.02. Quality is fly < none − 0.02. Anatomy is fly < Erdős–Rényi − 0.02 and fly < degree-match − 0.02. Specificity is fly < learned-sparse − 0.02. Ironclad is all three on three of three seeds. Telemetry includes α and the relative residual norm $\rho = \|\alpha F\_W(r)\| / \|r\|$. Telemetry is not a win criterion.

4.6. Expansion as retrieval hash (M0)

M0 is not T1. Synapse counts are not FFN weights. The ALPN → Kenyon-cell mask is a hash. Arms are mushroom-body expansion plus WTA, random same-nnz expansion plus the same WTA, dense cosine, and a learned projection. Memory size and dimension are matched. The primary metric is recall@1 at input noise

0.2, margin 0.02, seeds 20260931–20260933. Efficiency (bytes per query, lookup operations, latency) is logged. M1, a language retrieval test, opens only if M0 is mushroom-body-specific against random expansion.

4.7. Central complex

The central-complex graph has 2878 nodes. Tests write a heading into the ellipsoid-body / protocerebral-bridge ring, hold with leaky dynamics, and read a linear four-way probe. Controls are Erdős–Rényi, degree-matched, and later non-brain graphs. Persistence gold requires either a no-dynamics readout ≥ 0.85 after 16 null steps or cosine to the gold write ≥ 0.90. Cue tests compare the held fly state with the sparse encoder itself (keep fraction 0.08). These tests are not language-model losses.

4.8. Silent-engram computers (Phase S, R, B)

S0 implements a bank whose silent traces are unreadable by key similarity and are restored only by a cue pathway $R(S, c)$. Bars are suppression (old value not returned by the key alone), recovery (true cue restores the old value), fidelity, and false lure (a key-similar lure must not restore). The critical control is `archive_key`: an active/silent flag plus cosine(reminder, KEY). If that control matches the silent bank, there is no *Drosophila*-shaped interface.

S1-core learns a recovery operator R_θ on frozen factor embeddings. Held-out slices are new entities and new entity × jurisdiction / era compositions. False memory is scored only on hard lures (wrong jurisdiction, wrong era, neighbor entity, close old rule, cue of another silent item). Capability and fly-novelty are separate claims. S2, a language-model insertion, requires fly novelty on three of three seeds and is not opened.

S1-X is an information audit, not a trainer search. It asks whether a deterministic map from (S, raw cue) to the named trace exists, whether a structural oracle ranks the true slot first, and whether the trained R_θ extracts that ranking.

R0 asks a global addressing question that the biology in [6] does not have to solve: $(S, c) \rightarrow i^*$ on new entity vectors. B0 gives the family bank as already found and tests expression. B1 matches output rights: every arm may point only at an existing family trace; invention is forbidden; physical slot identifiers are independently permuted per family and permuted again after training without refit; payloads travel with the version. Arms are silent expression, a generic constrained selector, a two-stage context gate, and an oracle. Seeds are 20261101–20261105. The fly-uniqueness claim from B0 is closed if the silent arm does not beat the strongest generic constrained selector.

4.9. Software, seeds, and statistics

Protocols live in `scripts/ironclad_*.py`, `scripts/proof_*.py`, and the phase judges under `artifacts/phase_*_judge.json`. Seeds are integer dates or reserved blocks (20260907–09, 20260921–23, 20260931–33, 20261001–13, 20261021–23, 20261031–33, 20261101–05). We do not treat a single seed as ironclad. We do not average bars that were defined as a joint contract. We do not retune a closed judge after seeing numbers.

4.10. Ethics, animals, and humans

This work did not conduct new experiments on humans, non-human animals, or plants. Connectomic measurements are taken from a previously published, community-curated electron-microscopic reconstruction [1,2]. Linguistic corpora are public research datasets used under their existing licenses. No personally identifying unpublished data are released.

4.11. Manuscript preparation

Writing-assistance software was used only to organize wording from recorded experimental artifacts. It is not an author. Every numerical claim was checked against those artifacts. The corresponding author is accountable for the text.

5. Results

5.1. Graph probes: syntax plants; lexicon and facts do not

On Cultura definitions and synonyms, the washed 256-node prior loses to scrambled maps (AUROC gaps −0.17 and −0.15). Raw same-sentence names can beat scrambled (clean entities: fly-256 prior +0.053; fly-32 raw +0.127). RuWordNet synonyms, hypernyms, and definitions all lose. NEREL facts lose (fly-256 prior gap −0.114). Wikidata capitals and a mixed relational set lose. Derivational morphology loses. A Mel'čuk-style layer loses. The only probe that clears the 0.05 bar against scrambled on a washed-or-raw sweep of structure is raw syntax at 256 × 12 (gap +0.080). The washed Mara prior wins none of these tasks (`mara_prior_wins = []`).

A node sweep against degree-preserving rewire, one seed, found a raw 64 × 4 peak (holdout AUROC 0.696 versus 0.523, gap 0.172). That comparison is A versus A. A later three-seed non-brain queue (`syntax_nb`) is the anatomy test:

- 64 × 4 versus non-brain: 1/3 seeds, fail.
- 64 × 12 versus non-brain: gaps +0.121, +0.083, +0.073; win.
- 256 × 12 versus non-brain: gaps +0.149, +0.064, +0.075; win.

Against rewire the old 64 × 4 peak is positive on 3/3; against a non-brain graph it is not. A relation sweep shows that 64 × 4 is not "syntax in general": `obj` never plants; `amod` is the relation that still likes 256 nodes; a union of relations kills 64 × 4 and prefers 64 × 12. These AUROC facts are properties of a graph plus a retrained probe. They are not language-model quality.

5.2. Attention and router priors do not beat no-graph

A raw 256 × 12 `amod` attention bias in TinyLM beat rewire and lost to no bias (test loss 3.976 versus 4.045 versus 3.956). A sparse bias at strength 0.25 did not beat rewire. Three-seed TinyLM queues with a raw 64-expert router lost under both rewire and non-brain controls: on no seed was real < shuffle and real < no-map together. A washed 256 → 32 production prior versus a non-brain graph lost on 3/3 seeds. An ABCD battery that required the fly cost to beat Erdős–Rényi, degree-match, and language-rewire with relative gap ≥ 0.05 failed A, B, C, and D. A single local Tiny-MoE seed in which raw-64 fly < rewire and fly < no-map (4.680 < 4.723 < 4.741) was not replicated on any three-seed queue.

Plant-1 transferred the 64 × 12 AUROC win into TinyLM pair-NLL (Table 1; Figure 5). Fly pair-NLL is below non-brain on 3/3 seeds. Fly is below none − 0.02 on 0/3 seeds. Mean pair-NLL is fly 3.294, non-brain 3.451, none 3.289 (fly − non-brain ≈ −0.157; fly − none ≈ +0.005). Four-way pair choice stays at 0.192–0.199 on every arm, below chance 0.25. Infer-only substitution of a foreign kernel into a finished no-kernel checkpoint raises pair-NLL, especially for the non-brain kernel. The 256 × 12 plant config completed one seed only (fly − none = +0.030) and is not a 3/3 close.

Plant-2 (learnable strength plus alignment) was aborted after two seeds so that the frozen-expansion test could run. Seed 20260907 had pair-NLL range 0.0078 < 0.01 and is fatal (not a fly fail). Seed 20260908 had fly worse than none by 0.015 and worse than non-brain by 0.013; learned strength did not collapse to zero (0.50–0.59); the alignment auxiliary remained near 1.01. Plant-2 is incomplete (0 wins of 2 completed; 4/5 required) and does not rescue plant-1.

Table 1. Plant-1 TinyLM pair-token NLL for the raw 64 × 12 kernel. Judge: fly < none − 0.02 and fly < non-brain − 0.02. *n*_pair = 96,972.

Seed	Fly	Non-brain	None	Fly − none	Fly − non-brain	Win
20260907	3.28286	3.45908	3.28874	−0.00588	−0.17622	No
20260908	3.28422	3.35702	3.28146	+0.00276	−0.07280	No
20260909	3.31493	3.53730	3.29742	+0.01751	−0.22237	No

5.3. Frozen ALPN → Kenyon-cell expansion is a worse language FFN

BSEG-T1 completed three seeds with no fatal flag and n_pair = 176,087 (Table 2; Figure 3). Order is stable: dense < degree-match < Erdős–Rényi < fly. Holdout sparsity is 0.0497–0.0498 on every arm, equal to k / n_KC, so WTA fired. Anatomy 0/3. Quality 0/3. The biological matrix is worse than random sparsity with the same number of nonzeros and worse than a larger dense FFN. A follow-on recurrent-filter test was not opened. This closes frozen mushroom-body expansion as a language FFN. It does not close expansion-plus-WTA as a retrieval algorithm; that is M0.

Table 2. BSEG-T1 pair-token NLL. Margin 0.02. Sparsity is holdout fraction of Kenyon cells > 0.

Seed	Fly	Degree-match	Erdős–Rényi	Dense	Fly − deg	Fly − ER	Fly − dense
20260907	4.10312	3.88956	3.94805	3.34126	+0.214	+0.155	+0.762
20260908	4.11849	3.90234	3.92959	3.35326	+0.216	+0.189	+0.765
20260909	4.07264	3.90981	3.98334	3.35659	+0.163	+0.089	+0.716

5.4. Projectome routing: C1 never entered; C1-R entered and did not help

C1 is a negative result about the interface, not about anatomy. On all three seeds α was exactly 0.0 at steps 1, 250, 500, and 1500. Forward curves of fly, Erdős–Rényi, and degree-match coincide bit for bit. Seeds 20260907 and 20260908 have inter-arm ranges 0.0070 and 0.0088, below the fatal threshold 0.01, and cannot be used for or against the projectome. Seed 20260909 has fly below none by 0.0136, short of the Quality margin 0.02, and identical to Erdős–Rényi and degree-match, so Anatomy is absent. Specificity on that seed is fly 3.446387 versus learned 3.447868 ($\Delta = -0.0015$), far from fly < learned − 0.02. Quality, Anatomy, and Specificity are each 0/3. Because the channel was off, C1 does not say that the projectome mask is worse than a random mask. It says the model selected the vanilla submodel. A local probe of the same residual, without a new language-model run, reproduces the dead state: from (α, W) = (0, 0), both gradients are zero and one AdamW step leaves both at zero.

D0 passed. C1-R therefore has the right to answer a content question (Figure 2). At step 1500, ρ on topology arms is 0.02–0.06 on every seed; W_max remains about 0.6–0.8. Masks entered computation (Table 3). Quality 0/3: only seed 23 has fly < none, and then by 0.002, an order of magnitude below the margin 0.02. Anatomy 0/3: fly is never better than both random controls; on seed 21 degree-match beats fly by ≈ 0.024. Specificity 0/3. Learned sparsity does not win systematically at comparable ρ (best learned advantage 0.009 on seed 23, below margin). All three seeds are valid; none is fatal.

This is a genuine negative, not technical invalidity. Anatomy was presented to the model. The causal chain is: projectome structure → used computation → no topology-specific language-model benefit. That chain is stronger than C1, where topology never entered. C2, a 190-mask search, 70-million-parameter scale-up, and 800-million-parameter scale-up are not opened. Projectome-mask → pair-NLL is closed.

Table 3. C1-R pair-token NLL and relative residual ρ at step 1500. Δ is fly minus the named control. Quality / Anatomy / Specificity margins are 0.02.

Seed	None	Fly	ER	Deg	Learned	Δnone	ΔER	Δdeg	Δlearned	ρ_fly
21	3.4700	3.4771	3.4722	3.4530	3.4720	+0.0071	+0.0050	+0.0241	+0.0052	0.061
22	3.4845	3.4851	3.4730	3.4782	3.4786	+0.0006	+0.0121	+0.0069	+0.0065	0.038
23	3.4499	3.4479	3.4422	3.4442	3.4387	−0.0020	+0.0057	+0.0037	+0.0092	0.050

5.5. Mushroom-body expansion is not a better associative hash

M0 primary recall@1 at noise 0.2 is mushroom-body 0.313 / 0.326 / 0.305, random+WTA 0.376 / 0.370 / 0.367, dense 0.863 / 0.841 / 0.846, learned 0.403 / 0.398 / 0.395 (Figure 4). Wins 0/3. Beats-random 0/3. At noise 0 every arm is 1.0; a 0.35 distractor mix is also near ceiling and is not a fly credit. Sparse storage is about 836 kB versus 1.05 MB dense (not ≤ 0.5×). Lookup operations fall from about 262 k to about 11 k, but recall collapses and latency is not better. M1 is not opened as a mushroom-body-specific language claim. Threshold fishing after the lock is not a result.

A repaired mushroom-body association circuit (drive a subset of Kenyon cells and dopaminergic neurons; Hebbian update on half the output neurons; score = mean(MBON_A) − mean(MBON_B) on pattern A) meets its own file rule on 3/3 seeds against an Erdős–Rényi graph (scores 0.180, 0.186, 0.213 versus 0.131, 0.134, 0.126). Specificity was not in that rule. Scores on pattern B have the same sign and nearly the same magnitude (0.160, 0.163, 0.197). The test shows that saturation can be removed. It does not show that A is bound and B is not.

5.6. Central complex: a ring code is not an attractor

A five-department projectome (optic, central complex, mushroom body, descending, antennal lobe) is structurally unlike random maps of the same size (fly NLL 0.350 versus Erdős–Rényi ≈ 3.23–3.45 and degree-match ≈ 1.73–2.23). That is a map fact, not a language-model fact.

On department-native probes, only the central-complex ring resultant passed its L1 rule (resultants 0.434 / 0.427 / 0.421, minimum gap 0.30). Optic, mushroom-body (saturated), descending (readout = whole graph), and antennal-lobe (cosine collapse ≈ 0.997) did not. A six-example “own textbook” follow-up did not replicate 3/3. Live loom detectors have fly activations on the order of 10^−3; `prefer_loom = 1` is the sign of a tiny difference, not a strong detector. An eight-way language-model holdout never cleared a +0.20 gap over controls with probe accuracy ≥ 0.70.

Heading hold after a full write is linearly readable on some seeds and some metrics (`hold_via_hold` 0.49–0.66 versus lower shuffle and degree-match). Persistence gold fails: after 16 null steps, no-dynamics readout is 0.30–0.41 and cosine to gold is ≈ 0.46 (mean inhibition) or ≈ 0.31 (ReLU). The fly state leaves the write earlier than controls. A four-new-seed confirmation of a 16-step linear probe went 2/4, with two seeds invalid because a label shuffle still read heading. Quality queries after hold are 0.50–0.73 against an interface parser at 1.00: the graph loses information relative to its input. Sparse-cue hold (0.20–0.29) is worse than the cue encoder itself (0.51–0.71). Isolated signed attractor and controlled persistence are not established. Central-complex recurrence is not inserted into a language model on the basis of these tests.

5.7. Silent engram: an interface exists; fly uniqueness does not

S0, by construction, realizes storage ≠ accessibility (Table 4). All five arms suppress the old value after a new write. `archive_key` does not recover on a true cue that is not the key (recovery ≈ 0.001) and almost always recovers on a key-lure (false lure ≈ 0.999). A learned gate opens on any reminder, including the lure. Dense overwrite and recency cannot return a silent trace. This shows that the four-bar profile is a different object from a flag plus key reactivation. It is not a transfer of fly physiology and not a reason to put the mechanism in a language model.

Table 4. S0 mean profile, three seeds. Suppression leak is P(old | key, no reminder).

Arm	Leak	Recover	Fidelity	False lure
Silent engram	0.00	1.00	1.00	0.00
Archive key	0.00	0.001	0.001	0.999
Learned gate	0.00	1.00	1.00	1.00
Dense overwrite	0.00	0.00	0.00	0.00
Recency	0.00	0.00	0.00	0.00

S1-core fails the joint contract on 3/3 seeds for both the silent arm and a capacity-matched dual-key (22,817 versus 22,784 parameters). No temperature meets training recovery ≥ 0.90 and hard-lure false memory ≤ 0.05; the fallback is argmax. Hold-entity recovery is 0.000–0.017 (silent) and 0.000–0.067 (dual-key). Hold-compose recovery is 0.000–0.083. Retention is 1.0 and suppression is 0.0; the broken axis is recovery. Dual-key false memory is ≈ 1.0 without an entity filter. Failure is already visible on the training set (silent training recovery 0.00–0.08). S2 is not opened.

S1-X shows that the failure is not missing information. A deterministic map from features to the named trace exists on 3/3 seeds (no identical states; maximum off-diagonal cosine ≈ 0.61; no opposite-label collisions). A structural oracle has rank-1 = 1.0 on train, hold-entity, and hold-compose; situation → slot linear fit cosine is 1.0; recovery is 0.88–1.00; margin versus the competitor is about +0.6. The oracle's residual "false memory" (0.19–0.25) is almost entirely retrieval of another real silent fact of the same entity for which the lure is a true reminder (wrong jurisdiction 0.28–0.42, wrong era 0.37–0.44, close old rule 0.25–0.45). Neighbor entities and another item's cue are refused. The trained R_θ has rank-1 0.00–0.08, median rank 20–63, and a margin versus the competitor of −3 to −7 already on train. A diagnostic multilayer perceptron reaches training recovery ≈ 0.95 and hold-entity recovery ≈ 0.02: it memorizes training pairs and does not acquire the entity-plus-situation rule. Phase S is not closed as "no signal in the features." S1b, a trainer hunt, is not opened.

R0 asks global $(S, c) \rightarrow i^*$ on new entity vectors. The oracle factorization has rank-1 1.0 on every slice, including entity permutation, so a rule exists. Learned factorized and multilayer-perceptron arms partially recombine seen entities and fail on new entity vectors. Dual-key and bilinear arms do not recombine. Capability 0/3. This is not the fly mechanism in [6], which is not obliged to solve global entity binding.

B0 gives the family bank as found. On that weaker computer, silent expression meets six bars at 1.0 / 0.0 on hold-entity, 3/3 seeds (8321 parameters). A two-stage context gate fails incomplete recovery (0.75–0.83) and lure resistance (0.08–0.17). A metadata selector fails incomplete recovery (0.50–0.75) and invents (0.25–0.28). Slots are shared $P@$(era, jurisdiction). The win is architectural inside a known four-trace family, not new-entity binding.

B1 removes that weakness (Table 5). Output rights, features, and parameter count are matched (silent 11,490; constrained selector 11,393; gate 11,488). Invention is forbidden. Slots are independently permuted per family and permuted again after training. Official scores on all three learned arms and all five seeds are identical: suppression 1.00, complete 1.00, incomplete 1.00, wrong-version 0, invention 0, sibling/hard-lure 1.00, re-permutation 1.00, compose 0, mean success 0.80, joint error 1.00. Gap silent minus best generic is 0. The oracle compose score is 1.00: the map exists; the split does not create it. STOP 5/5. The *Drosophila*-specific claim from B0 is closed. B0 remains a useful factorization on a weaker computer. It is not a unique fly mechanism. S2 is not opened.

Table 5. B1 official bars, all learned arms, all five seeds. Oracle compose = 1.00 is not a contender.

Bar	Silent	Constrained selector	Context gate
Suppression	1.00	1.00	1.00
Complete recovery	1.00	1.00	1.00
Incomplete recovery	1.00	1.00	1.00
Wrong-version	0	0	0
Invention	0	0	0
Sibling / hard lure	1.00	1.00	1.00
Re-permutation	1.00	1.00	1.00
Compose recovery	0	0	0
Mean success	0.80	0.80	0.80

5.8. An 800-million-parameter score is not a connectome effect

On a Russian six-task multiple-choice suite (22 June 2026, acc_norm, continued pretraining ≈ 9.3 × 10^9 tokens, $k = 8$, no supervised fine-tune), Mara 800M scores 36.0%, between Qwen2.5-0.5B (36.6%) and TinyLlama-1.1B (35.8%), and below Qwen2.5-7B (47.2%), Mistral-7B-v0.3 (43.6%), and Qwen2.5-1.5B (39.8%) [18–20]. Task scores for Mara are RCB 34.5%, ruOpenBookQA 33.4%, MathLogicQA 19.3%, PARus 51.0%, ruWorldTree 28.7%, and BPS 49.2%. A different checkpoint (8.8 × 10^9 tokens plus SQL continued pretraining and supervised fine-tune, vocabulary 51,260) reached WikiSQL exact-match 59.8% English and 54.4% Russian on 2000 test examples; the merge-32 9.3 × 10^9-token base on a different SQL format is 46.5%. Those numbers must not be mixed, and none of them is an ablation of the blueprint. Cyrillic tokenizer expansion (0 Cyrillic types in the base 49,152; 2000 added; Russian tokens per word 5.778 → 4.133) is also not a connectome win.

6. Discussion

6.1. What was tested, and what the numbers license

The program tested four transfers of fly structure into computation and one transfer of a fly-shaped interface that is not a mask.

1. Fixed washed and raw graph priors in attention or routers. The washed prior is nearly the identity and never wins a probe. Raw syntax kernels contain reproducible structure relative to a wrong graph. That structure does not become a pair-NLL advantage over the absence of a graph.

2. Frozen biological expansion as an FFN. The operator is active and worse than random sparsity and a dense FFN on 3/3 seeds.

3. Trainable projectome communication. After the degeneracy of C1 is removed, the channel is used and the fly mask is not better than vanilla, random, degree-matched, or learned-sparse topology.

4. The same expansion as a retrieval hash. Random expansion plus the same WTA is better; dense cosine is much better.

5. Storage versus accessibility. The interface is realizable. A learned recovery map fails on ranking already in training. Expression inside a given family bank is not unique to a silent split once every arm is forbidden to invent.

The locked sentence for routing is: trainable family-level sparse communication is viable, but the *Drosophila* projectome topology provides no measurable pair-NLL advantage over vanilla, random or degree-matched topology, or learned sparsity under C1-R. The locked sentence for expansion-as-FFN is: frozen ALPN → KC plus hard k-WTA is not a language FFN. The locked sentence for expansion-as-hash is: the fly mask is not better than random+WTA and both lose to dense. The locked sentence for silent engrams is: B1 closes *Drosophila* uniqueness under matched output rights; B0 remains a weaker-computer factorization; S2 is not opened.

6.2. Why a naïve weight transfer is a weak hypothesis

FlyWire v783 is a typed, signed, neuropil-resolved graph with tens of millions of synapses [1,2]. The production Mara pipeline kept 256 hubs, applied $\log(1 + \cdot)$, a two-step diffusion, a unit diagonal, and a block average. Type, excitation/inhibition, neuropil, motifs, rich-club structure, and time constants are discarded. Two hundred and fifty-six hubs are not a brain. Lappalainen et al. succeed in a setting where the architecture *is* the circuit and the task *is* the circuit's task [3]. Beiran and Litwin-Kumar show why wiring alone is often not a function [4]. Dasgupta et al. already located expansion-plus-WTA in retrieval [5]. The present negatives are consistent with that literature. They are not a proof that no other interface exists.

6.3. Discovery versus confirmation

The 64 × 12 and 256 × 12 AUROC wins were selected after a sweep of size, degree, and relation. A later 3/3 rule on those configs is not an independent confirmation of a special linguistic structure in the fly connectome. Late queues reduce that problem by writing the judge before the job. They do not erase it. We therefore treat "graph better than a wrong graph" as a discovery signal that failed to become "graph better than no graph" under a prewritten language-model rule.

6.4. What remains open, and what is not a sequel

Efficiency, robustness, and out-of-distribution behavior were not measured. That is a gap in the quality claim, not evidence for scaling the closed operators. A residual heading-ring signal in a linear probe is not an attractor and does not justify inserting the central complex into a language model. Compose recovery of 0 on B1, with an oracle at 1.0, does not imply that a larger multilayer perceptron would create composition: the split does not produce it. Global named-trace addressing (R0) is not the mechanism in [6] and is not reopened here. Generic three-factor plasticity of the form $\Delta W = \eta \cdot \text{pre} \cdot \text{post} \cdot \text{dopamine}$ was not tested.

A directed `amod`-binding comparison against degree-matched, learned-sparse, and bilinear computers is outside this report.

6.5. Scope of the negatives

The results are negative on the named computers. They are not a claim that *Drosophila* has no useful algorithms, and they are not a claim that brains are useless for machine learning. They show that these transfers, under these judges, did not improve the measured objectives.

7. Conclusions

A preregistered program asked whether *Drosophila* wiring, or a *Drosophila*-shaped silent-engram interface, improves language-adjacent objectives relative to non-brain controls. The closed answers are:

1. A washed hub prior is not a *Drosophila* test and does not win.

2. Raw syntax kernels can beat a wrong graph and do not beat the absence of a graph on pair-NLL at margin 0.02.

3. Frozen ALPN → KC expansion is a worse language FFN than random sparsity and a dense FFN (0/3 anatomy, 0/3 quality).

4. Under a live trainable residual, projectome topology provides no measurable pair-NLL advantage over vanilla, random or degree-matched topology, or learned sparsity (0/3 Quality, 0/3 Anatomy, 0/3 Specificity).

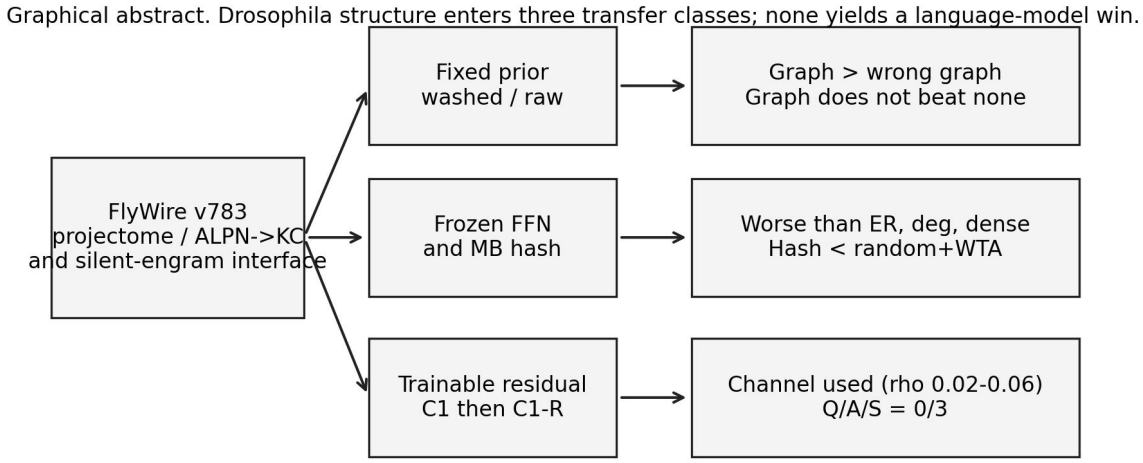
5. The same expansion is not a better associative hash than random expansion plus the same WTA (0/3).

6. Isolated central-complex persistence does not meet attractor criteria.

7. A silent-engram interface is realizable; a learned recovery operator fails on ranking; fly uniqueness of expression dies under a constrained selector (gap 0, five seeds).

The closed tests do not support scaling the washed prior, the raw fixed prior, or frozen BSEG-T1 into an 800-million-parameter model, and they do not attribute that model's 36.0% Completed-6 score to the connectome. S2, a retuned B1, and reopenings of C1-R, M0, or R0 lie outside the present evidence.

Figures



Graphical abstract. Drosophila structure is transferred in three classes. None yields a language-model win under the preregistered judges.

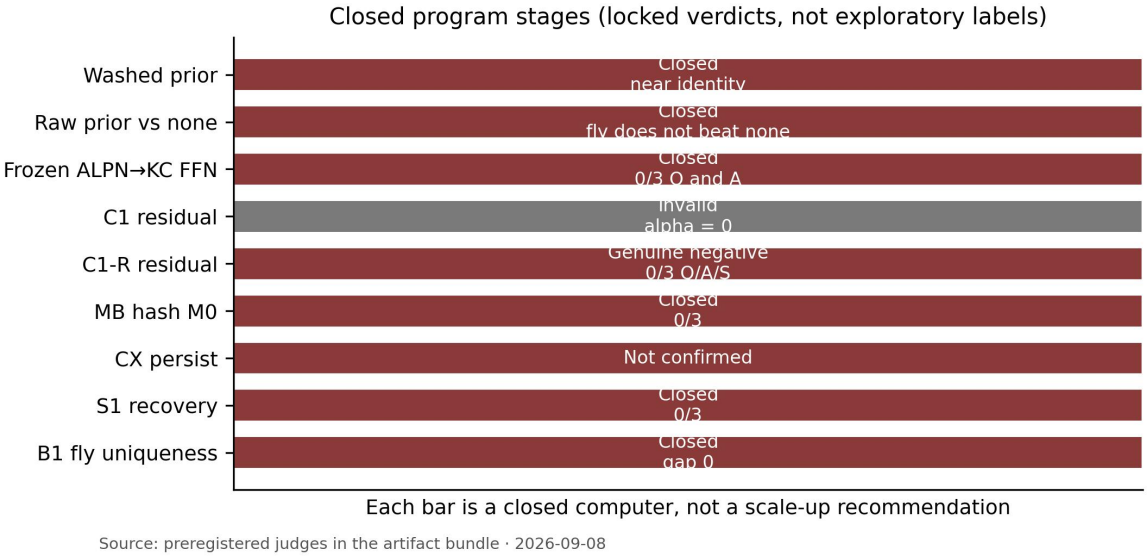


Figure 1. Closed experimental stages and locked verdicts. C1 is an invalid anatomy test (gate α = 0). C1-R is a genuine negative with a live channel.

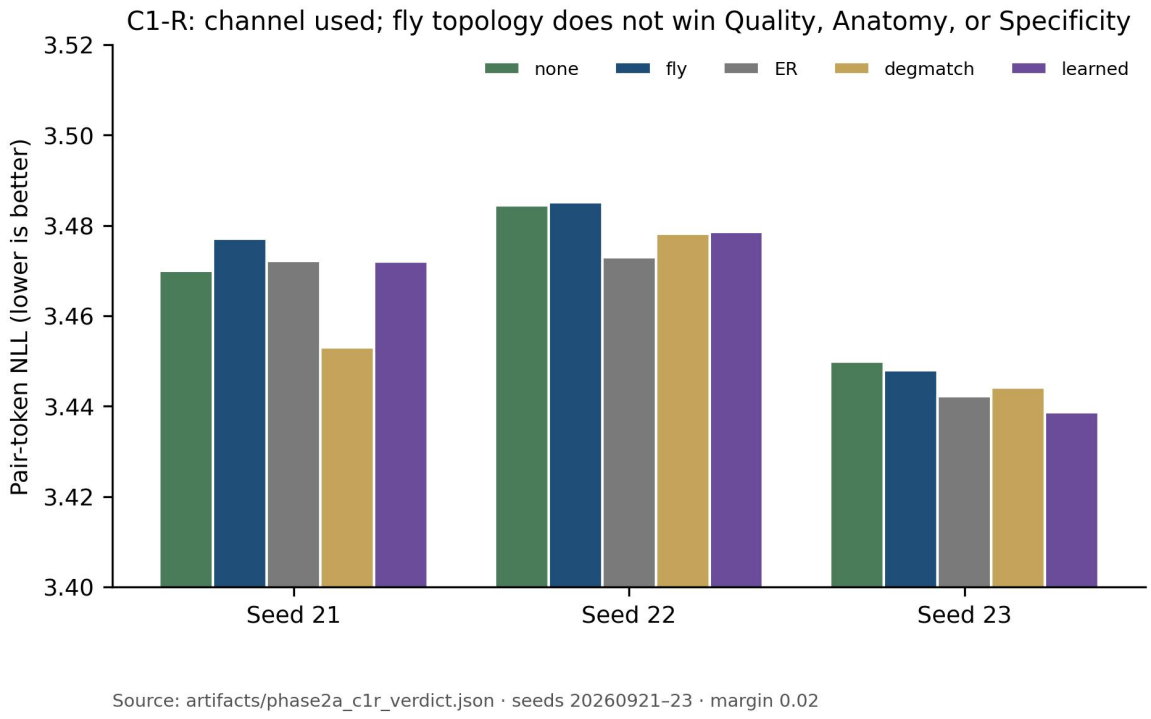


Figure 2. C1-R pair-token NLL by seed and topology arm. Lower is better. Quality / Anatomy / Specificity margins are 0.02.

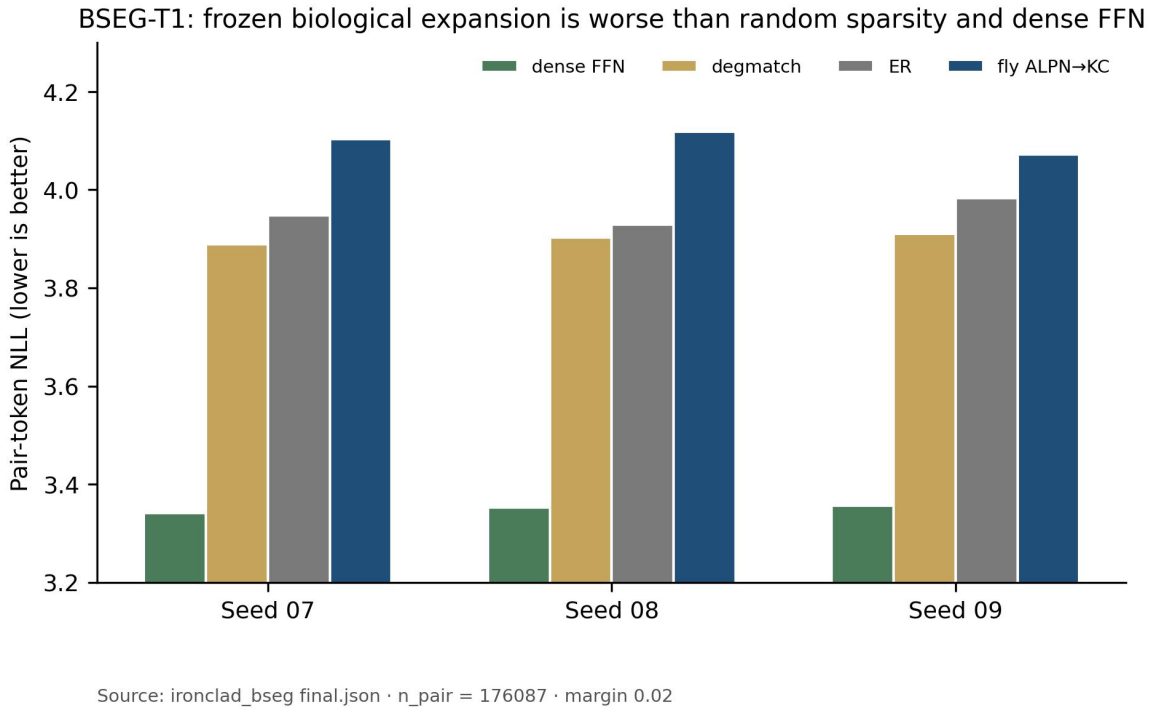


Figure 3. BSEG-T1 pair-token NLL. Order is dense < degree-match < Erdős–Rényi < fly on three of three seeds.

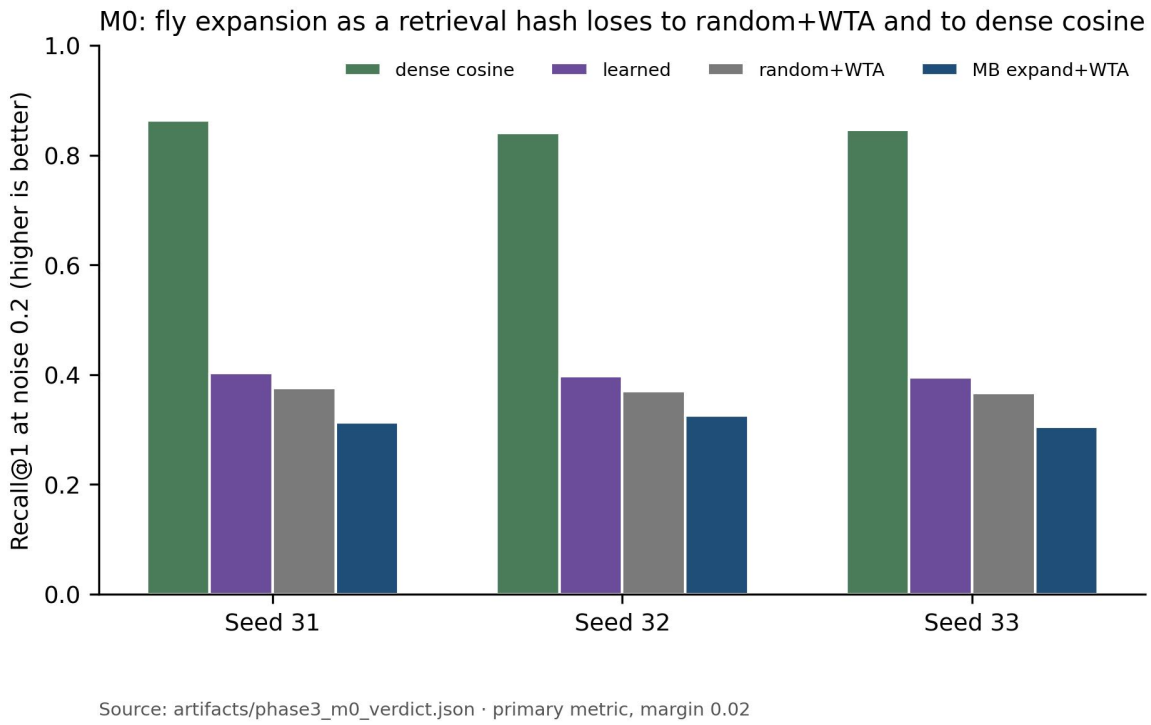


Figure 4. M0 primary recall@1 at noise 0.2. Higher is better. Fly expansion plus WTA loses to random expansion plus the same WTA and to dense cosine.

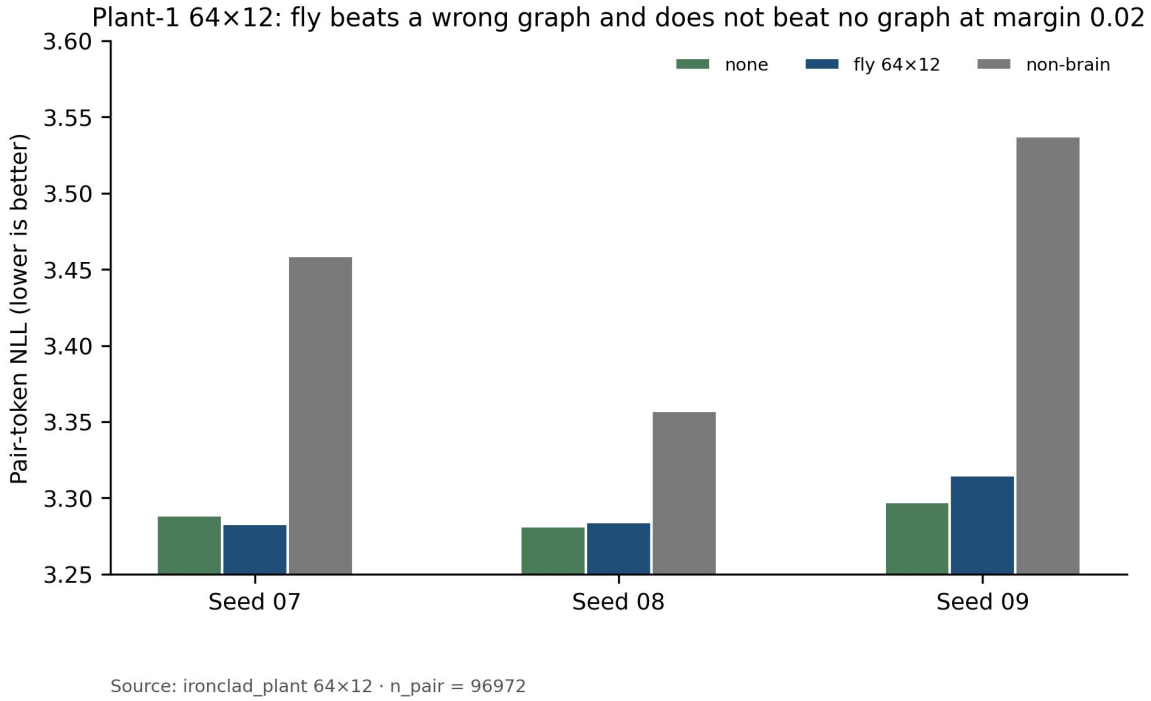


Figure 5. Plant-1 64 × 12 pair-token NLL versus none and a non-brain kernel. Fly beats the wrong graph and does not beat no graph at margin 0.02.

Supplementary Materials

Protocol files, preregistered judges, per-seed result tables, and locked verdicts for C1, C1-R, M0, S0, S1, S1-X, R0, B0, and B1 are provided as Supplementary Files with this preprint.

Author Contributions

The author meets ICMJE authorship criteria [16]. Conceptualization, V.M.; methodology, V.M.; software, V.M.; investigation, V.M.; formal analysis, V.M.; data curation, V.M.; writing—original draft, V.M.; writing—review and editing, V.M. The author has read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Ethics approval

Not applicable. No new human or animal experiments were conducted.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

All data generated or analyzed during this study are included in this article and its supplementary information files. FlyWire v783 connectivity and annotations are public connectomic resources [1,2]. SynTagRus, NEREL, RuWordNet, Cultura, and Wikidata are public datasets [11–15,17]. There are no legal or confidentiality restrictions on the synthetic memory payloads. Third-party corpora remain under their original licenses and are not redistributed as full text. The raw FlyWire table should be obtained from Codex (https://codex.flywire.ai) and cited as [1,2].

Acknowledgments

The author thanks the FlyWire community for the public v783 reconstruction and annotations [1,2] and the creators of the linguistic resources cited above [11–15].

Competing interests

The author declares no competing interests.

References

1. Dorkenwald, S.; Matsliah, A.; Sterling, A.R.; Schlegel, P.; Yu, S.-C.; McKellar, C.E.; Lin, A.; Costa, M.; Eichler, K.; Yin, Y.; et al. Neuronal wiring diagram of an adult brain. *Nature* **2024**, *634*, 124–138. https://doi.org/10.1038/s41586-024-07558-y

2. Schlegel, P.; Yin, Y.; Bates, A.S.; Dorkenwald, S.; Eichler, K.; Brooks, P.; Han, D.S.; Gkantia, M.; dos Santos, M.; Munnelly, E.J.; et al. Whole-brain annotation and multi-connectome cell typing of *Drosophila*. *Nature* **2024**, *634*, 139–152. https://doi.org/10.1038/s41586-024-07686-5

3. Lappalainen, J.K.; Tschopp, F.D.; Prakhya, S.; McGill, M.; Nern, A.; Shinomiya, K.; Takemura, S.; Gruntman, E.; Macke, J.H.; Turaga, S.C. Connectome-constrained networks predict neural activity across the fly visual system. *Nature* **2024**, *634*, 1132–1140. https://doi.org/10.1038/s41586-024-07939-3

4. Beiran, M.; Litwin-Kumar, A. Prediction of neural activity in connectome-constrained recurrent networks. *Nat. Neurosci.* **2025**. https://doi.org/10.1038/s41593-025-02080-4

5. Dasgupta, S.; Stevens, C.F.; Navlakha, S. A neural algorithm for a fundamental computing problem. *Science* **2017**, *358*, 793–796. https://doi.org/10.1126/science.aam9868

6. Yang, W.; Zattera, B.; Pavão-Delgado, M.; Warnecke, C.; Schweizer, J.A.; Ricciuti, A.; Çoban, B.; Felsenberg, J. Creating true and false memories from forgotten information in *Drosophila*. *Nat. Neurosci.* **2026**. https://doi.org/10.1038/s41593-026-02381-2

7. Graves, A.; Wayne, G.; Reynolds, M.; Harley, T.; Danihelka, I.; Grabska-Barwińska, A.; Colmenarejo, S.G.; Grefenstette, E.; Ramalho, T.; Agapiou, J.; et al. Hybrid computing using a neural network with dynamic external memory. *Nature* **2016**, *538*, 471–476. https://doi.org/10.1038/nature20101

8. Sukhbaatar, S.; Szlam, A.; Weston, J.; Fergus, R. End-to-end memory networks. In *Advances in Neural Information Processing Systems 28*; 2015.

9. Ba, J.; Hinton, G.E.; Mnih, V.; Leibo, J.Z.; Ionescu, C. Using fast weights to attend to the recent past. In *Advances in Neural Information Processing Systems 29*; 2016.

10. Allal, L.B.; Lozhkov, A.; Bakouch, E.; Blázquez, G.M.; Penedo, G.; Tazi, L.; et al. SmolLM2: When Smol goes big—from 135M to 1.7B parameters. Hugging Face Blog / technical report, 2024–2025.

11. Droganova, K.; Lyashevskaya, O.; Zeman, D. Data conversion and consistency of monolingual corpora: Russian UD treebanks. In *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018)*; Linköping University Electronic Press: Linköping, Sweden, 2018; pp. 52–65. SynTagRus developed at IITP RAS; UD release: https://universaldependencies.org/treebanks/ru_syntagrus/

12. Nguyen, T.; Nguyen, C.V.; Lai, V.D.; Man, H.; Ngo, N.T.; Dernoncourt, F.; Rossi, R.A.; Nguyen, T.H. CulturaX: A cleaned, enormous, and multilingual dataset for large language models in 167 languages. *arXiv* **2023**, arXiv:2309.09400.

13. Loukachevitch, N.; Lashevich, G.; Gerasimova, A.A.; Ivanov, V.V.; Dobrov, B.V. Creating Russian WordNet by conversion. In *Computational Linguistics and Intellectual Technologies: Dialogue 2016*; 2016; pp. 405–415. RuWordNet 2.0: https://www.ruwordnet.ru/

14. Loukachevitch, N.; Artemova, E.; Batura, T.; Braslavski, P.; Ivanov, V.; Manandhar, S.; Pugachev, A.; Rozhkov, I.; Shelmanov, A.; Tutubalina, E.; et al. NEREL: A Russian information extraction dataset with rich annotation for nested entities, relations, and Wikidata entity links. *Lang. Resour. Eval.* **2023**. https://doi.org/10.1007/s10579-023-09674-z

15. Vrandečić, D.; Krötzsch, M. Wikidata: A free collaborative knowledgebase. *Commun. ACM* **2014**, *57*, 78–85. https://doi.org/10.1145/2629489

16. International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Available online: https://www.icmje.org/recommendations/ (accessed on 8 September 2026).

17. Nivre, J.; de Marneffe, M.-C.; Ginter, F.; Goldberg, Y.; Hajič, J.; Manning, C.D.; McDonald, R.; Petrov, S.; Pyysalo, S.; Silveira, N.; et al. Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of LREC 2016*; ELRA: Portorož, Slovenia, 2016; pp. 1659–1666.

18. Jiang, A.Q.; Sablayrolles, A.; Mensch, A.; Bamford, C.; Chaplot, D.S.; de las Casas, D.; Bressand, F.; Lengyel, G.; Lample, G.; Saulnier, L.; et al. Mistral 7B. *arXiv* **2023**, arXiv:2310.06825.

19. Yang, A.; Yang, B.; Hui, B.; Zheng, B.; Yu, B.; Zhou, C.; Li, C.; Li, C.; Liu, D.; Huang, F.; et al. Qwen2.5 technical report. *arXiv* **2024**, arXiv:2412.15115.

20. Zhang, P.; Zeng, G.; Wang, T.; Lu, W. TinyLlama: An open-source small language model. *arXiv* **2024**, arXiv:2401.02385.

21. Zhong, V.; Xiong, C.; Socher, R. Seq2SQL: Generating structured queries from natural language using reinforcement learning. *arXiv* **2017**, arXiv:1709.00103.

22. Dorkenwald, S.; McKellar, C.E.; Macrina, T.; Kemnitz, N.; Lee, K.; Lu, R.; Wu, J.; Popovych, S.; Mitchell, E.; Zung, J.; et al. FlyWire: Online community for whole-brain connectomics. *Nat. Methods* **2022**, *19*, 119–128. https://doi.org/10.1038/s41592-021-01330-0

23. Turner-Evans, D.B.; Jayaraman, V. The insect central complex. *Curr. Biol.* **2016**, *26*, R453–R457.

24. Green, J.; Adachi, A.; Shah, K.K.; Hirokawa, J.D.; Magani, P.S.; Maimon, G. A neural circuit architecture for angular integration in *Drosophila*. *Nature* **2017**, *546*, 101–106.

25. Aso, Y.; Hattori, D.; Yu, Y.; Johnston, R.M.; Iyer, N.A.; Ngo, T.-T.; Dionne, H.; Abbott, L.F.; Axel, R.; Tanimoto, H.; et al. The neuronal architecture of the mushroom body provides a logic for associative learning. *eLife* **2014**, *3*, e04577.

26. Hige, T.; Aso, Y.; Modi, M.N.; Rubin, G.M.; Turner, G.C. Heterosynaptic plasticity underlies aversive olfactory learning in *Drosophila*. *Neuron* **2015**, *88*, 985–998.

27. Erdős, P.; Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **1960**, *5*, 17–61.

28. Shazeer, N.; Mirhoseini, A.; Maziarz, K.; Davis, A.; Le, Q.; Hinton, G.; Dean, J. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv* **2017**, arXiv:1701.06538.

29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems 30*; 2017.

30. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

31. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.

32. Mel'čuk, I.A. *Dependency Syntax: Theory and Practice*; SUNY Press: Albany, NY, USA, 1988.

33. Lyashevskaya, O.; Droganova, K.; Zeman, D. A report on the Russian Universal Dependencies treebanks and the SynTagRus conversion. Universal Dependencies project documentation, https://universaldependencies.org/treebanks/ru_syntagrus/ (accessed on 8 September 2026).

Appendix A. Closed queues at a glance

Queue	**Judge**	**Result**
Washed prior probes	AUROC vs scrambled	No wins
syntax_nb 64 × 4	AUROC vs non-brain, 3/3	Fail (1/3)
syntax_nb 64 × 12 / 256 × 12	AUROC vs non-brain, 3/3	Pass (graph+*Q* only)
TinyLM CE, rewire and non-brain	real < both, 3/3	Fail
Plant-1 64 × 12	pair-NLL vs none and non-brain	Fail vs none (3/3 vs non-brain)
Plant-2	pair-NLL, ≥4/5	Aborted; 0/2 completed wins
BSEG-T1	Anatomy and Quality, 3/3	0/3 and 0/3
C1	Q / A / S, 3/3	Channel off; no anatomy verdict
C1-R	Q / A / S, 3/3	Channel on; 0/3 / 0/3 / 0/3
M0	recall@1, noise 0.2	0/3; M1 not opened
CX persist / t16 / cue	attractor and cue rules	Not confirmed
S0	four-bar interface	3/3 by construction; not an LM
S1-core	capability and fly	0/3 and 0/3
S1-X	information vs learning	Information present; *R*_θ fails ranking

R0	global addressing	0/3
B0	family-bank expression	3/3 vs weak controls
B1	matched constrained selector	STOP 5/5; fly uniqueness closed

Appendix B. Incomplete or skipped items

Plant 256 × 12 seeds 08 and 09 were not completed. Plant-2 seeds 09, 1501, and 1502 were not started. `cue_ring` was skipped because persist was not fly-only. BSEG-T2 was not opened. An 800-million-parameter blueprint ablation was not run. Speed, memory, and dollars per token were not run. SQL probes on native neurons were not run locally.