

# When does a biological circuit transfer?

Pre-registered, adversarially reviewed tests of *Drosophila* mushroom-body and ring-attractor models on text, chess, image, associative-recall and code-clone tasks

Thiago Patzdorf

Independent researcher; Fábrica de Sites, Brazil

thiagosleman@gmail.com

## Abstract

We report a methods-and-negative-results study of two circuits of the *Drosophila* connectome, the mushroom body (MB) and the central-complex ring attractor, measured against trivial and strong baselines on text, chess, photographs, associative recall and code clones. Every experiment from EXP-002 onward was pre-registered; the first text cycle and the photograph application were not and are marked as such. Each ran with a trivial baseline, a blocking ablation and, where wiring was involved, a degree-preserving null, then adversarially reviewed and verified. The protocol caught instance leakage that let a dictionary lookup (MCC +0.374) beat every model; a Kenyon-cell layer whose removal did not change the result; a strong baseline (GRU) trained for 50 optimizer steps, whose retraining erased most pre-registered effects; and a naive null whose replacement by a degree-preserving null flipped a +13.53 pp connectome advantage to −7.11 pp. The MB is not a competitive classifier on text, chess or a synthetic conjunctive task; with a dense encoder its sparse code recognises near-duplicates, but exact search beats it wherever exact search fits (on photographs the two are indistinguishable and exact search wins the paired test), and on code clones a supervised linear model wins outright. The ring’s parameter-efficiency headline fell to a 3-parameter vector integrator (0.11–0.20% loss against the Bayesian optimum versus the ring’s 0.9%) and to a retrained GRU; its topology-versus-shuffled ablation (+44.49 pp) and untrained inductive bias survive. In chess, a small geometric grid beats logistic regression at full data but loses to a zero-parameter capture rule and to a dense network. The transferable rule: a strong baseline must receive more training budget than the proposed model.

## 1 Introduction

Biologically derived architectures are usually reported when they win. This paper reports a research program whose positive results kept falling to stronger controls, and argues that the record of how they fell is the useful product. The program’s question was narrow: which computational principles of a measured insect circuit transfer to artificial learners, and under what conditions do they increase capability per unit of data, time, memory or compute? A well-measured negative answer is as informative as a positive one.

Two circuits and one derivative were tested. The **mushroom body** (MB) is the fly’s associative-learning centre: a sparse, high-dimensional expansion of olfactory input into Kenyon cells (KCs), global inhibition by the APL neuron, and dopamine-modulated readout by output neurons (MBONs). It has an established machine-learning lineage as a locality-sensitive hash (Dasgupta et al., 2017), a novelty detector (Dasgupta et al., 2018), a hashing objective (Ryali et al., 2020), a word-embedding model (Liang et al., 2021) and a continual learner (Shen et al., 2023). The **ring attractor** of the central complex maintains heading by integrating angular velocity and correcting against landmarks (Seelig and Jayaraman, 2015; Kim et al., 2017; Turner-Evans et al., 2017); a Bayesian ring attractor with a closed-form optimum exists as a benchmark (Kutschireiter et al., 2023). The **spatial grid** is a static two-dimensional adaptation of the ring’s excitation/inhibition principle, built for chess. The MB models were built on, or checked against, the FlyWire FAFB v783 connectome (Dorkenwald et al., 2024; Schlegel et al., 2024): 5,177 KCs, 96 MBONs, 285 antennal-lobe projection neurons,

2 APL neurons and 331 dopaminergic neurons in the MB extract. The ring uses the canonical ellipsoid-body/protocerebral-bridge topology, not measured wiring.

The experiments span five task families: text (SWE-smith agent trajectories labelled by test execution), chess (Lichess puzzle move classification and Elo tournaments), images (catalogue photographs), associative recall (can a KC layer stand in for transformer attention?) and code-clone detection (BigCloneBench). The MB fails as a classifier on text, on chess and on a synthetic conjunctive task; on dense images it was not tested against a strong adversary. The paper is organised around the protocol (Section 3), the MB as classifier (Section 4), the MB as similarity index (Section 5), the ring attractor and the grid (Section 6), and the record of what the protocol caught (Section 7).

**Contributions.** (1) A protocol for testing bio-derived architectures (pre-registration with dated amendments, mandatory trivial and strong baselines, blocking ablation with an isomorphism control, degree-preserving nulls, adversarial review, numeric verification, an append-only ledger) together with the record of what each rule caught. (2) The MB fails as a classifier on text, chess, code clones and a synthetic conjunctive task, and reproduces, in three new modalities, the published pattern: with a live encoder it beats weak hashes and, in most settings, its own ablation, and loses to exact search and to a cheap supervised model; its value envelope as a near-duplicate recogniser is quantified. (3) The ring attractor’s parameter-efficiency headline falls to a 3-parameter comparator and to a retrained GRU; its topology ablation and untrained inductive bias survive. (4) In chess, a 33-parameter geometric grid beats logistic regression at full data but loses to a zero-parameter capture rule and to a dense MLP; classification skill is not playing strength. (5) A rule: a strong baseline must receive more training budget than the proposed model.

Every number in this paper is traceable to a file shipped as ancillary material (`fontes.md` maps each number to a source path and line). Where the record disagrees with itself, both values are footnoted.

## 2 Related work

Dasgupta et al. (2017) abstracted the MB as a sparse  $40\times$  expansion followed by winner-take-all (WTA) retention of the top 5% and showed that this “fly hash” preserves neighbourhoods better than classical LSH at matched compute; later work turned it into a novelty filter (Dasgupta et al., 2018), a hashing objective (Ryali et al., 2020), sparse word embeddings (Liang et al., 2021) and a single-pass continual learner (Shen et al., 2023). The critical literature matters as much: Sharma and Navlakha (2018) found the fly hash impractical without multi-probe; Kleyko and Rachkovskij (2024) studied the design choices of similarity-preserving sparse randomized embeddings of this family and found that projection density and imitation of the biological wiring heterogeneity changed little, while centring and normalisation before the WTA step and the choice of ranking metric changed much; Ram and Sinha (2022) found that the fly classifier’s gain over  $k$ -NN on 70 datasets is statistically insignificant while its gain over its own ablation is large. In neuroscience, the MB’s behaviour is set by its encoder (Lin et al., 2014; Hige et al., 2015; Cayco-Gajic and Silver, 2019), APL inhibition is local (Amin et al., 2020), the PN→KC projection is not random at current sample sizes (Zheng et al., 2022) (contra Caron et al., 2013), and the optimal claw count for a random expansion has a theory (Litwin-Kumar et al., 2017); the MB connectome is described by Li et al. (2020); Modi et al. (2020).

Ring-attractor models of head direction go back to Skaggs et al. (1995); Zhang (1996); Xie et al. (2002); the fly bump, its dynamics and its shift were measured by Seelig and Jayaraman (2015); Kim et al. (2017); Turner-Evans et al. (2017), the connectome by Hulse et al. (2021), and Kutschireiter et al. (2023) derived the circular Kalman filter as the normative optimum. Continuous attractors are fragile to mis-tuning (Renart et al., 2003; Vafidis et al., 2022; Noorman et al., 2024); generic RNNs develop compass/shifter populations (Cueva et al., 2020) but structure rarely emerges unless planted (Schaeffer et al., 2022); Negre Saura et al. (2025) used a ring to structure the action space of a deep RL agent. Bricken and Pehlevan (2021) showed that softmax attention approximates the

Table 1: The rules as they stood at the end of the program, and what each one answers to.

| rule                                                                                                                                                                      | origin                                                                                                             |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| Pre-registration committed before the first training run, with minimum effect and CI requirement; amendments appended with date and reason (the ring carries five, E1–E5) | a winner was declared on noise twice in one night                                                                  |
| Trivial baseline mandatory (dictionary, majority, last observation, content hash, and the zero-parameter rules of the task); leave-one-out oracle printed as the ceiling  | a dictionary beat two architectures; a capture rule beat every chess grid                                          |
| Strong baseline with more training budget than the proposed model; convergence check before any crossing claim                                                            | the GRU episode (Section 6.3); the chess MLP cut at epoch 10 (Section 6.6)                                         |
| Blocking ablation: removing the component under test must degrade by the pre-registered margin or all other $\Delta$ s are void                                           | a KC layer that contributed 0.013                                                                                  |
| Isomorphism control: permuting topology and input/readout indices together must reproduce the intact model                                                                | separates misalignment from graph damage                                                                           |
| Degree-preserving null (Maslov and Sneppen, 2002) whenever measured wiring is compared with random wiring; the naive shuffle is only the weak control                     | the null flipped the sign (Section 4.4); it is uninformative when the tested wiring is itself random (Section 5.4) |
| Group split, never row split; a lookup on the grouping key as leakage sentinel, $ \text{MCC}  \leq 0.05$                                                                  | 59.6% of test instances were in training                                                                           |
| Encoder diagnostic before the learner: Jaccard within/between class, active claws per KC, largest-claw share, distinct codes, content density                             | 93.5% of active KCs had one active claw                                                                            |
| Joint criteria with a floor on each quantity (retention and plasticity)                                                                                                   | three single-quantity metrics gamed in one day                                                                     |
| Stable hashing (CRC32) and explicit integer seeds                                                                                                                         | <code>hash()</code> is salted per process                                                                          |
| Append-only ledger: one entry per run with commit, dirty-tree flag, command, machine, dataset SHA, seeds, times, peak memory; single versioned file                       | five divergent per-worktree ledgers                                                                                |
| Adversarial review with a mandate to run experiments; numeric verification of every number                                                                                | self-review found none of the three worst errors                                                                   |
| No conclusion from a subset preview; paired comparisons, not overlapping marginal intervals; tie credit by expected value                                                 | Sections 4.1, 5.3                                                                                                  |

read of Kanerva’s sparse distributed memory (Kanerva, 1988), a line related to dense associative memories (Krotov and Hopfield, 2016; Ramsauer et al., 2021) and extended to continual learning (Bricken et al., 2023); a KC layer with a Hebbian readout is a fixed random-feature linear attention (Katharopoulos et al., 2020; Choromanski et al., 2021) whose recall-versus-state trade-off is the one measured by Arora et al. (2023, 2024). Degree-preserving rewiring as the null model for network structure follows Maslov and Sneppen (2002); Dhiman (2026) showed that a connectome-constrained visual model (Lappalainen et al., 2024) loses its reported advantage under such a null.

### 3 Protocol

The protocol was not designed in advance; each rule was added after its absence produced a result that was believed for some time (Table 1; Section 7).

**Metric and intervals.** Classification results use the Matthews correlation coefficient (MCC; 0 for any constant predictor) after F1 selected candidates that predicted the positive class everywhere; the clone experiment reports F1 for comparability with its benchmark, with MCC alongside. Intervals are bootstrap percentiles, two-level where seeds exist (test items paired across models, seeds resampled per model).

**Review and verification.** The correction report, the referee report and the numeric verification were produced by LLM-based agents of the same model family as the experimenters, in separate checkouts with separate mandates: independent in context, not in provenance. The verification checked 78 numeric claims of the first article: 68 matched, 2 diverged (a column labelled “IC95” that was the min–max across seeds; an inverted paraphrase of a literature mechanism), 0 lacked a source, 8 were ambiguous; one table of six configurations had no ledger record and was rescued by deterministic re-execution (18 of 18 cells). A second verifier checked 233 mapped numbers of this manuscript’s first draft against their source files: 231 matched exactly, 2 were ambiguous in the source and are corrected inline, and one unmapped count was wrong and is corrected; a further pass covered the sections added in revision (Sections 5.4, 6.6 and 9; 118 items; 113 matched; the rest corrected). The four reports ship as ancillary files: `PARECER.md` and `VERIFICACAO.md` (first article), `PARECER-ARXIV-2026-09-27.md` and `VERIFICACAO-ARXIV-2026-09-27.md` (this manuscript). Experiments, code and this manuscript were produced by LLM agents under the author’s direction; Section 9 gives the full disclosure.

**Compute.** All experiments ran on CPU (a 12-core Windows workstation and an 8-core Linux server). The longest single runs were the 82.5-minute clone experiment, the 53-minute connectome experiment, the 44-minute text retest (EXP-006, SVD arm) and a 4 h 48 min chess tournament.

## 4 The mushroom body as a classifier

### 4.1 Text: SWE-smith

Given the description of a software-engineering task from SWE-smith, predict whether the agent’s patch resolves it (label by test execution); 49,773 trajectories from 13,469 instances, the `instance_id` present in 100% of texts, leave-one-out oracle MCC +0.841. The first result (fly +0.296 vs balanced logistic regression +0.338 on a row split) was invalid on five grounds found by adversarial review: instance leakage (a dictionary on `instance_id` scored +0.374; 59.6% of test instances were in training), a dead KC layer (93.5% of active cells with one active claw; ablation  $\Delta = 0.013$ ), a process-salted hash (four “networks” read as one), a stream-shift test with a zero ceiling by construction, and a gameable metric. Each is a row of Tables 10 and 14; the original result file was marked invalid and kept.

**EXP-002.** The biological claim is one-shot associative memory. EXP-002 ran the full file pre-quentially (predict, then learn), grouped by instance, 3 seeds, against a hash table on `instance_id`, exact 1-NN on TF-IDF and a logistic regression retrained every 2,000 items (Table 12, appendix). Both pre-registered abandonment criteria were met with wide margins: fly minus hash table on instances seen 1–3 times  $-0.618$  [ $-0.627, -0.610$ ]; fly minus logistic regression on unseen instances  $-0.057$  [ $-0.068, -0.047$ ]; and the KC layer was harmful,  $-0.023$  [ $-0.027, -0.020$ ] against the no-KC arm. The wiring-faithful v3 activated 19 of 5,177 KCs (0.4%) and scored 0.000.

### 4.2 The fair retest: a live encoder, and the fly as front-end

The abandonment was by criterion, but the autopsy showed the test had not tested the architecture. Two untried conditions were pre-registered (commit 4c1388b24): a *dense* input (TF-IDF→SVD-256 over the whole text; `bge-small-en-v1.5` 384-d embeddings truncated at 512 tokens) and the MB as *front-end* (KC code → trained linear readout, the “insect cyborg” configuration of Delahunt and Kutz, 2019). Hyper-parameters were searched on validation only (18 configurations  $\times$  5 regularisers for the fly; 5 for logistic regression); the test set was read once. Pre-registered effects: H-A classifier minus logistic regression  $\geq +0.05$ ; H-B front-end minus logistic regression  $\geq +0.03$ ; H-C front-end minus hash table on items seen 1–3 times  $\geq +0.05$ ; blocking gate P2, KC contribution  $\geq +0.02$ .

**The encoder is alive.** On the previous champion configuration (20,000 KCs,  $\rho = 5\%$ , 6 claws), changing only the input, active claws per selected KC went from 1.07 to 6.00 of 6 and the largest-claw

Table 2: EXP-006, SWE-smith, split by instance (test  $n = 10,297$ ), MCC with 95% bootstrap CI. LOO oracle +0.841; majority class and instance lookup both +0.000.

| model                                                      | TF-IDF→SVD-256                  | bge-small-384           |
|------------------------------------------------------------|---------------------------------|-------------------------|
| no KC (rule on dense input; ablation of classifier)        | +0.133 [+0.114, +0.151]         | +0.106 [+0.088, +0.125] |
| random expansion with WTA (4,096)                          | +0.241 [+0.221, +0.259]         | +0.156 [+0.136, +0.174] |
| 1-NN                                                       | +0.210 [+0.192, +0.229]         | +0.176 [+0.157, +0.193] |
| logistic regression on dense input (ablation of front-end) | + <b>0.273</b> [+0.255, +0.291] | +0.188 [+0.172, +0.209] |
| random expansion without WTA (4,096)                       | +0.273 [+0.256, +0.290]         | +0.199 [+0.182, +0.219] |
| MB classifier                                              | +0.243 [+0.225, +0.263]         | +0.194 [+0.176, +0.213] |
| MB front-end (KC code → linear readout)                    | +0.222 [+0.204, +0.237]         | +0.205 [+0.186, +0.224] |

share of the drive from  $\approx 1.00$  to 0.420 (SVD) and 0.345 (bge). The Jaccard overlap between *sibling* trajectories (same instance) minus non-siblings is +0.685 [+0.651, +0.718] (SVD) and +0.918 [+0.897, +0.936] (bge): siblings share 71.4% or 94.5% of active KCs against 2.8% between instances. The within- minus between-class Jaccard is +0.00090 [+0.00058, +0.00126] and +0.00026 [+0.00008, +0.00035]: the code recognises instances and barely separates classes. Everything below follows from that line.

**EXP-006, group split (generalise).** Table 2 gives the test scores ( $n = 10,297$ ; leakage sentinel exactly +0.000). The KC contribution reversed sign, +0.110 [+0.090, +0.131] (SVD) and +0.088 [+0.067, +0.112] (bge), so P2 passed. H-A failed in both representations ( $-0.030$  [ $-0.046, -0.013$ ]; +0.006 [ $-0.015, +0.027$ ]) and so did H-B ( $-0.052$  [ $-0.072, -0.031$ ]; +0.016 [ $-0.006, +0.042$ ]). The attribution is the informative line: with the same random projection and  $n_{\text{KC}} = 4,096$ , a linear readout on the dense drive scores +0.273 in SVD, equal at three decimals to logistic regression on the raw input, and the same readout on the WTA-sparsified binary code scores +0.241; the WTA step costs  $-0.032$  [ $-0.051, -0.013$ ] (SVD) and  $-0.044$  [ $-0.066, -0.023$ ] (bge). The operation that defines a Kenyon cell discards the drive magnitude, and in this task the linear signal lives in the magnitude. An exploratory extension outside the grid found nothing in SVD that reaches logistic regression (best of 31 configurations +0.265 vs +0.273); in bge one out-of-grid configuration beat it by +0.035 [+0.013, +0.056] on a single test read, recorded as post-hoc and not counted. Both sides converged (front-end readouts used 14–71 of 200 iterations; validation scores flat across three orders of magnitude of budget).

**EXP-007, repeated exposure (recognise).** The prequential protocol was rerun with the dense encoders; the hash table reproduced its earlier numbers to within 0.002, so the runs are comparable line by line (Table 3). The bge front-end scores +0.839 on items seen 1–3 times, five times its no-KC ablation (+0.165) and +0.582 above the earlier fly. The “insect cyborg” effect is real and large: with identical flow, retraining and seeds, a linear readout on the KC code versus on the dense vector gives  $\Delta = +0.578$  on items seen 1–3 times and  $-0.004$  on unseen items, with the sign the two Jaccard measurements predicted. It still wins nothing that counts: H-C failed,  $-0.034$  [ $-0.038, -0.031$ ] against the hash table (it had been  $-0.618$ ), and exact 1-NN on the same embedding scores +0.880. In the band of items seen three or more times the front-end edges both (0.692 vs 0.689 and 0.656); no CI was computed for this band, it was not pre-registered, and it is not counted. The circuit was doing approximate similarity search against two exact and cheaper methods on a corpus of 13,469 instances that fits in memory. One error made along the way is recorded: the first front-end search fixed the regulariser at  $C = 1$  while varying  $n_{\text{KC}}$  and therefore measured regularisation, not capacity (+0.124 vs +0.203 at 8,000 KCs); the run was discarded whole and redone with a joint search.

**Corrected verdict.** The first article said the transposition preserved no measurable property; that is now measured as false. It preserves one, near-duplicate recognition, worth +0.578 of MCC as a feature generator in the recognition regime. What does not transfer is utility: where the property serves, a dictionary and exact 1-NN do better and cheaper; where it does not, the APL’s global inhibition is a lossy quantiser costing 0.032–0.044. Continual learning on dense images (EXP-003/004)

Table 3: EXP-007, prequential, bge-small-384, 3 seeds, MCC. Front-end: 20,000 KCs,  $\rho = 0.05$ , 12 claws,  $C = 3.0$ . In SVD-256 the front-end reaches only +0.462 on items seen 1–3 $\times$  (hash table +0.873).

| exposure band       | MB classifier | MB front-end  | no KC  | hash table    | 1-NN          | logistic (dense) |
|---------------------|---------------|---------------|--------|---------------|---------------|------------------|
| never seen          | +0.151        | +0.194        | +0.124 | −0.006        | +0.173        | <b>+0.198</b>    |
| seen once           | +0.193        | +0.880        | +0.154 | <b>+0.936</b> | +0.930        | +0.255           |
| seen $\geq 3\times$ | +0.248        | <b>+0.692</b> | +0.238 | +0.656        | +0.689        | +0.293           |
| seen 1–3 $\times$   | +0.207        | +0.839        | +0.165 | +0.873        | <b>+0.880</b> | +0.261           |
| all                 | +0.219        | +0.623        | +0.188 | +0.590        | <b>+0.634</b> | +0.257           |

is reported in Appendix B: the pre-registered thresholds passed and the interpretation fell, because the criterion was satisfiable by a model that stops learning.

### 4.3 Chess: two encodings, two failure mechanisms

Given a Lichess puzzle position and a candidate move, is it the solution or a sampled legal alternative? 9,000 puzzles (18,000 balanced examples), split 80/20 by puzzle, MCC with group-bootstrap CIs; success required MCC within 0.02 of logistic regression *and* a KC contribution  $\geq 0.03$ . With 12 piece planes  $\times$  64 squares plus move one-hots (897 dimensions) into 2,000 KCs  $\times$  6 claws with top-5% WTA, the fly scored +0.101 [+0.079, +0.124] against +0.245 for logistic regression, +0.387 for a 123,265-parameter MLP and +0.185 for its own no-KC ablation. The encoder diagnostic explains it: the vector is dense in format but 2.3% non-zero in content (20.8 of 897 positions), so 86.5% of (KC, example) pairs had zero active claws and the class-separation Jaccard was −0.0006; the mechanism of Section 4.1 in an input that looked dense. A second round with a signed piece value per square (64 dimensions, 28.6% non-zero) left the Jaccard separation at +0.0002, dropped the fly to +0.016 and the MLP from +0.387 to +0.190, because the scalar encoding is ambiguous; content density and information preservation are different axes. Per the pre-registered rule, no playable engine was built from the MB.

### 4.4 Does the measured wiring help? Connectome versus degree-preserving null

Every earlier MB version drew the PN $\rightarrow$ KC projection at random, “because in the fly it is random”; that assumes the null hypothesis about the connectome rather than approximating it. The extractor was rebuilt to include the input layer of FlyWire v783: 285 antennal-lobe projection neurons (ALPN)  $\rightarrow$  5,177 KCs (22,298 edges, 366,946 synapses), KC $\rightarrow$ KC (24,169 edges, never modelled before), KC $\rightarrow$ 96 MBONs (26,937 edges) and the two APLs; the extractor gives 4.69 ALPN claws per KC against  $5.2 \pm 1.6$  in Zheng et al. (2022). The task is a synthetic odour mixture with a conjunctive label (eight components as graded sparse patterns over the 285 ALPNs, three parity bits as an 8-class label), not linearly separable in component space; in the pre-training gate logistic regression on the ALPNs scored 12.8–15.0% against a 13.2% majority class and the exact Bayes optimum 100% (the gate and the experiment are separate runs, hence the small differences from Table 4). Two regimes (neutral; “convergent”, where signal concentrates on high-degree ALPNs), 5 seeds, 5 reshuffles per null, 4,000 training and 2,000 test items, shared initialisation. The pre-registered prediction, following Zheng et al. (2022) and Dhiman (2026), was a tie.

All four gates passed, so  $\Delta_1$  is interpretable, and the connectome *loses*: −4.30 pp and −7.11 pp (Table 4). In the convergent regime the same circuit, training and seeds beat the naive null by +13.53 pp, a publishable-looking advantage; only the control changes and the sign changes with it, because the naive shuffle destroys the degree distribution (Gini of ALPN out-degree 0.501  $\rightarrow$  0.066,  $z = +203$ ). The measured wiring is far from random: with degrees fixed, KC $\rightarrow$ KC reciprocity is 0.4912 against 0.0018 (270 $\times$ ,  $z = +1,744$ ), clustering 57 $\times$ , and ALPN $\rightarrow$ KC co-convergence variance 13.142 against 5.854 ( $z = +64.7$ ; Zheng et al., 2022 report  $z = 13.8$ ). The mechanism of the loss is one number: the real wiring covers 11,464 distinct ALPN pairs where the shuffled wiring covers 17,390, 34% fewer conjunctions; for a random expansion, coverage of distinct conjunctions is the

Table 4: Measured FlyWire MB wiring against two nulls, 8-class parity task, accuracy (%), 5 seeds;  $\Delta$  rows are paired differences with two-level bootstrap 95% CI. The MLP (28,232 parameters vs 27,713 for the connectome model) was not pre-registered.

| arm                                                                  | neutral                 | convergent                     |
|----------------------------------------------------------------------|-------------------------|--------------------------------|
| Bayes optimum                                                        | 100.00                  | 100.00                         |
| MLP 285→96→8                                                         | 99.96                   | 99.98                          |
| naive shuffle null (mean of 5)                                       | 77.33                   | 65.84                          |
| degree-preserving null (mean of 5)                                   | 71.61                   | 86.48                          |
| connectome                                                           | 67.31                   | 79.37                          |
| isomorphism control                                                  | 67.36                   | 79.33                          |
| no KC (logistic on ALPNs)                                            | 13.26                   | 13.45                          |
| $\Delta_1$ connectome – degree null (pre-registered: $\geq +2.0$ pp) | −4.30 [−6.12, −2.32]    | −7.11 [−8.27, −5.94]           |
| $\Delta_2$ connectome – naive null (descriptive)                     | −10.02 [−11.88, −7.96]  | <b>+13.53</b> [+12.08, +14.95] |
| $\Delta_3$ isomorphism ( $ \Delta  \leq 0.5$ pp)                     | −0.05 [−0.20, +0.10]    | +0.04 [−0.03, +0.16]           |
| $\Delta_4$ KC layer vs no KC (blocking, $\geq +5$ pp)                | +54.05 [+52.14, +55.98] | +65.92 [+64.21, +67.59]        |
| $\Delta_5$ active ALPN inputs per KC ( $\geq 2.0$ )                  | 2.339                   | 2.976                          |

scarce resource (Litwin-Kumar et al., 2017), and the measured wiring trades coverage for redundancy. A first run had given a larger deficit (−7.36/−10.47 pp) for a reason that was not topology: the FlyWire CSV has one row per (pre, post, neuropil), and treating rows as edges gave the null 83% more partners and 49,236 plastic KC→MBON parameters against 26,937, violating the pre-registered parameter equality through a file format; the erroneous artefacts are kept beside the corrected ones. The encoder diagnostic did not predict the result (the connectome’s mean Jaccard separation, 0.0072, exceeds the null’s 0.0052 because both its within- and between-class overlaps are higher), and an MLP with matched parameter count solves the task at 99.96%.

## 5 The mushroom body as a similarity index

### 5.1 The value envelope and the decision rule

Five fronts converged on one description (a sixth, Section 5.4, later confirmed it). The MB code is an excellent near-duplicate recogniser (Jaccard between near-duplicates +0.685 to +0.918) and an almost null category separator (+0.0009 between classes); it offers cost and memory fixed in  $N$ , single-exposure writes, novelty detection for free, and about  $0.1r$  identifiable items for  $r$  cells. Its measured limit is that when exact search fits the budget, exact search wins (dictionary +0.873 and exact 1-NN +0.880 against +0.839). The decision rule recorded in the ledger: “have I seen *exactly* this?” → a dictionary; “have I seen something *similar*, and can I store everything?” → exact 1-NN or an embedding index; “similar,  $N$  too large, fixed memory?” → only then the circuit. None of the program’s five application candidates reached the third line; the largest corpus in play was about 4,000 items.

### 5.2 Can a KC layer replace attention? The SDM bridge

Bricken and Pehlevan (2021) show that softmax attention with normalised keys equals the read of a Kanerva SDM in the “pattern view”, which abstracts the neurons by assuming  $r = 2^n$  address units. A KC layer *is* the neurons: a finite- $r$  Monte Carlo estimate of the attention weights. EXP-005 (pre-registered in commit a06064412) stored  $N$  (key, value) pairs in  $\mathbb{R}^{64}$ , queried with noise  $\sigma \in \{0, 0.25, 0.5, 1\}$ , in a random-key regime R and a correlated-key regime C, 5 seeds, and compared a hash table, 1-NN, attention at  $\beta \in \{10, 25, 50, 100\}$ , a no-KC ablation, a synthetic fly hash, the FlyWire KC layer ( $r = 5,177$ , measured APL weights), a centred readout, a dense random projection with the same  $k$ WTA and a canonical Kanerva SDM; nobody learns. The decisive hypothesis: is there any (regime,  $N, \sigma$ ) where the KC layer costs less than attention *and* stays within 0.10 of its fidelity? The encoder diagnostic answered first: the code is alive (0% dead cells, 259 active per input) but its kernel has  $\beta_{\text{eff}} = 2.6$  and floor  $\kappa(0)/\kappa(1) = 0.05$ , against  $\beta = 25$  and a floor of  $1.4 \times 10^{-11}$

Table 5: EXP-005 associative recall, fidelity  $\cos(\text{output}, \text{correct value})$ ,  $\sigma = 0$ , mean of 5 seeds. Hash table and 1-NN are 1.00 throughout.

| regime, $N$ | attention $\beta=25$ | no KC | synthetic KC | connectome KC | Kanerva SDM | centred KC |
|-------------|----------------------|-------|--------------|---------------|-------------|------------|
| R, 64       | 1.00                 | 0.71  | 0.90         | 0.90          | 0.91        | 0.96       |
| R, 512      | 1.00                 | 0.33  | 0.60         | 0.58          | 0.62        | 0.77       |
| R, 4096     | 1.00                 | 0.12  | 0.25         | 0.24          | 0.26        | 0.39       |
| C, 16       | 1.00                 | 0.55  | 0.88         | 0.89          | 0.88        | 0.91       |
| C, 4096     | 0.95                 | 0.04  | 0.11         | 0.11          | 0.11        | 0.13       |

Table 6: EXP-008, recall@1 of the source photo given the cutout, 114 test pairs against 3,994 images; fly configurations chosen on the calibration half.

| method                                           | recall@1 | 95% CI         |
|--------------------------------------------------|----------|----------------|
| exact content hash                               | 0.000    | —              |
| aHash 64 bit                                     | 0.840    | [0.784, 0.893] |
| pHash 64 bit                                     | 0.874    | [0.822, 0.920] |
| no-KC ablation on pixel (20 of 1,024 active)     | 0.865    | [0.815, 0.912] |
| SimHash, same 40 bits, on pixel                  | 0.899    | [0.851, 0.943] |
| colour histogram (RGB $8^3$ )                    | 0.921    | [0.877, 0.965] |
| fly hash on pixel (2,000 cells, 40 active)       | 0.923    | [0.876, 0.961] |
| dHash                                            | 0.924    | [0.882, 0.961] |
| MobileNet 576-d, exact 1-NN                      | 0.930    | [0.886, 0.969] |
| pixel $32 \times 32$ grey, exact 1-NN            | 0.939    | [0.899, 0.974] |
| fly hash on MobileNet (20,000 cells, 400 active) | 0.947    | [0.912, 0.978] |

for the attention [Bricken and Pehlevan \(2021\)](#) measure in trained models; four encoders give the same curve within 0.02. Table 5 shows the consequence: attention is exact across the range and the KC layer degrades from  $N = 16$ ; it becomes cheaper than attention at  $N \approx 410$ , where its fidelity is already  $\leq 0.64$  (R) or  $\leq 0.36$  (C), and the hypothesis has zero cells in either regime for all five SDM/KC arms. By identification (nearest stored value is the right one) the raw KC layer holds about 500 items ( $0.1r$ ) and the centred readout about 1,500, in 331k numbers, whereas attention holds 4,096 pairs exactly in 524k numbers: more numbers, but more pairs per number. Eight of ten pre-registered predictions held, and the two misses are explained with measured moments of the code. The bridge holds, and that is why the substitution fails: the circuit is the cheap sample estimate of what attention computes exactly. What it has is on the axis attention does not occupy (fixed cost, one-shot writing, novelty for free), where it competes with linear attention and Bloom filters and does not choose its kernel.

### 5.3 Application: near-duplicate photographs in a product catalogue

The purest fit to the envelope: the same handbag photographed twice, with different crop, background, lighting or resolution, among the images of an e-commerce catalogue (325 SKUs, 4,196 manifest entries, 3,994 downloadable thumbnails; the 83 missing were HTTP 404s in 42 SKUs). Ground truth came from a manifest field: 273 background-removed cutouts, each carrying the URL of its source photo; 228 pairs survived, split by SKU into 114 for calibration and 114 for test. Exact search has two forms here with opposite answers: the manifest `checksum` field pairs 228 of 228 (provenance), a content hash of the normalised thumbnail 0 of 228; the operational case, a supplier photo without provenance, is the second.

Table 6 gives the test scores. The best row is the fly hash on MobileNet features (0.947) and the trivial baseline, exact nearest neighbour on a  $32 \times 32$  grey thumbnail, is at 0.939; with 114 pairs the two are statistically indistinguishable, as is everything from 0.923 upward. The fly’s published claim reproduces at the level it is made: on raw pixels it beats SimHash at the same bit budget by +2.4 pp and its own ablation by +5.8 pp (isomorphism control  $\Delta = 0.000000$ ). The paired stress test EXP-008b (228 transformed queries, eight transformations, paired bootstrap against exact 1-NN on the fly’s own descriptor) decides it: the fly never wins, ties in four conditions and loses

significantly in four (crop 10%  $-0.068$  [ $-0.129, -0.007$ ]; brightness  $-0.094$ ; grey background  $-0.091$ ; textured background  $-0.150$  [ $-0.209, -0.093$ ]). What decides is the descriptor: swapping pixels for the CNN moves recall on cropped queries from 0.048 to 0.539, and putting the fly hash on top moves it between  $-15$  and  $+0.4$  points, mostly downward. The same-SKU control behaves as the envelope predicts (recall@1 of another photo of the same product 0.011–0.027 on 1,019 pairs). The recommendation shipped was pixel 1-NN at a calibrated threshold (precision 0.872, recall 0.895) with a colour histogram beside it; the fly was not switched on because 3,994 vectors of 4 KB are 16 MB.

## 5.4 Code clones: EXP-009

The last application asked whether the circuit’s envelope includes code-clone detection, a domain with a public benchmark and a literature of hash-based baselines. Pre-registered on 2026-09-23 (commit 508db1828; scripts committed before the first run in ffaa50c2e): the MB from Section 4.2, unchanged, on BigCloneBench (Svajlenko et al., 2014) in its CodeXGLUE form (Lu et al., 2021), `google/code_x_glue_cc_clone_detection_big_clone_bench`. That version has no clone type and no project, so the pre-registered Type-4/WT3 slice is untestable and the split is by function (60/20/20, seed 1789; the official splits are already function-disjoint). From a pool of 1,506,429 unique pairs over 9,126 functions,  $N = 50,000$  pairs were sampled (30,000 / 10,000 / 10,000; 22.31% positives); test  $n = 10,000$ ; F1 with paired bootstrap CI (2,000 resamples), MCC alongside. Two circuit forms share one KC layer: an LSH form (the fly hash, pair score = shared active cells) and an MBON form (pair vector  $[|a - b|, a \cdot b] \rightarrow \text{KC} \rightarrow \text{MBON rule}$ ); “the circuit” in the  $\Delta$ s is the form with the best validation F1, MBON on bge (20,000 KCs,  $\rho = 0.05$ , 6 claws). Baselines: normalised exact token hash, MinHash ( $k = 5$ , 128 permutations), SimHash (64 bit), exact cosine on the same dense vectors with a validation threshold, logistic regression on the pair’s hashed TF-IDF, majority, always-clone, random; extras outside the pre-registration: logistic regression on the dense pair vector, no-KC, code shuffled against items, and raw (un-standardised) cosine. One run, 82.5 min of CPU.

The scoreboard (F1, test): logistic regression on pair TF-IDF 0.865 [0.855, 0.875] (MCC +0.825; 23 s to train) > logistic regression on the dense pair 0.798 (SVD) / 0.635 (bge) > circuit-MBON bge 0.582 [0.567, 0.598] (MCC +0.447) > raw cosine 0.502 (SVD) / 0.512 (bge) > MinHash 0.395 [0.380, 0.409] > always-clone 0.365 > SimHash 0.364; the exact hash finds one pair in 10,000. The pre-registered gate, circuit minus its degree-preserving shuffle  $\geq +0.10$ , gave +0.007 [+0.000, +0.014] and **failed**, which by the pre-registration voids  $\Delta_1$ – $\Delta_3$  even though  $\Delta_1$  (vs MinHash, +0.188 [+0.171, +0.203]) and  $\Delta_2$  (vs exact cosine on the same bge vectors, +0.075 [+0.061, +0.089]) passed numerically; the LSH form, which is the direct analogue of the published fly hash, scores +0.003 [−0.006, +0.012] against exact cosine.  $\Delta_3$ : the KC code per item (1,600–3,200 bytes as int32 indices) costs more than the dense vector it would replace (1,024–1,536 bytes). Isomorphism control 0 of 10,000 predictions changed; code shuffled against items falls to 0.197; the no-KC ablation ranges from  $-0.015$  to  $+0.074$  depending on form and representation, and in the chosen form it *beats* the circuit (0.597 vs 0.582).

Two methodological findings belong to the record. First, a degree-preserving permutation of a projection that is itself random and untrained is a re-seeding: it cannot make the circuit worse, so the pre-registered gate was uninformative by construction (two reshuffle seeds sit inside the intact circuit’s CI). The controls that discriminate are the code shuffled against items (0.197) and the no-KC arm; the rule is that degree-preserving rewiring only tests wiring that is not random by construction (it did its job on the measured connectome in Section 4.4). Second, a per-dimension z-score inherited from EXP-006, where it feeds the KC layer, was applied before the cosine baseline as well and destroyed the SVD baseline (0.365, i.e. always-clone, against 0.502 without it); the SVD  $\Delta_2$  of +0.176 is therefore inflated and the honest value against raw cosine is +0.038 [+0.019, +0.058] (in bge nothing changes, 0.508 vs 0.512). In the decision rule’s language: the circuit’s envelope does not include code-clone detection at this scale (9,101 functions  $\times$  1.5 kB = 14 MB fits in memory), and a supervised linear model is the cheap, exact-enough method: on the same dense bge pair vector, logistic regression beats the circuit by +0.053 [+0.040, +0.066] (SVD: +0.258), and on the pair’s

Table 7: Angular integration, regime N1, percentage loss against the circKF optimum, median of 5 seeds on a fixed test set of 2,000 sequences. Brackets are the min–max across seeds, not a confidence interval. Optimum error 23.0°.

| estimator                            | parameters | error (°) | loss (%)   | [min; max]    |
|--------------------------------------|------------|-----------|------------|---------------|
| optimum (circKF)                     | —          | 23.0      | 0.0        |               |
| ring                                 | 9          | 23.7      | <b>0.9</b> | [0.3; 1.1]    |
| ring, relabelled (isomorphism)       | 9          | 23.7      | 0.9        | [0.3; 1.1]    |
| ring, 64 wedges                      | 9          | 23.8      | 1.0        | [0.5; 1.3]    |
| ring without attractor ( $g_e = 0$ ) | 8          | 24.7      | 2.2        | [1.9; 2.3]    |
| gru-h16                              | 1,090      | 50.5      | 39.0       | [36.1; 44.0]  |
| gru-h8                               | 354        | 56.0      | 47.4       | [43.0; 50.3]  |
| ring without rotation                | 9          | 59.6      | 53.1       | [39.6; 55.1]  |
| trivial: last landmark               | —          | 61.0      | 55.2       |               |
| ring, shuffled (degree-preserving)   | 9          | 60.9      | 55.1       | [10.1; 55.4]  |
| optimal-gain integrator              | —          | 63.0      | 57.5       |               |
| gru-h32                              | 3,714      | 89.2      | 98.7       | [27.0; 100.6] |
| trivial: zero                        | —          | 90.6      | 100.8      |               |

TF-IDF, a different input, by +0.282 [+0.266, +0.298].

## 6 The ring attractor and its spatial-grid derivative

### 6.1 Task, instrument and pre-registration

The task follows [Kutschireiter et al. \(2023\)](#): heading  $\phi_t = \phi_{t-1} + \Delta_t$ ,  $\Delta_t \sim \mathcal{N}(0, \sigma_\Delta^2)$ ; observed angular velocity  $u_t = \Delta_t + \mathcal{N}(0, \sigma_u^2)$ ; a von Mises landmark with probability  $p$  per step. Step 0.15 s, 200 steps, typical turn 153°/s,  $\sigma_u = 0.3 \sigma_\Delta$ , landmark every  $\approx 3$  s with  $\kappa_z = 16$ ; regimes N0 (pure integration), N1 (with landmark; the main experiment) and N2 (landmark blackout between steps 60 and 160, never seen in training). The score is percentage loss against the circular Kalman filter optimum (0 = optimum, 100 = chance) on the second half of each sequence. The instrument was validated before any model existed: seven checks, eleven assertions, including that circKF ties a 2,000-particle filter to +0.02% and that the best untrained baseline loses 55.9 pp to the optimum;<sup>1</sup> a first parametrisation was rejected by that gate. All models receive the same four inputs per step. The ring is the canonical ellipsoid-body/protocerebral-bridge topology: 16 wedges, local excitation and global inhibition ([Kim et al., 2017](#)), two P-EN shift populations ([Turner-Evans et al., 2017](#)), a fixed cosine landmark expansion and a parameter-free population-vector readout; only 9 gains train. Comparators: GRUs with  $h \in \{1, \dots, 32\}$  (25 to 3,714 parameters), RNNs, a “split” GRU that receives  $\text{ReLU}(\pm u)$  for free, the optimum, two untrained integrators and two trivial estimators; the GRUs received a second hyper-parameter search at full budget that the ring did not (amendment E4), and the ring’s gains were jittered per seed (E3). Five pre-registered effects (commit 793c5482f, 4 min 41 s before the first recorded training run):  $\Delta_1$  ring  $\geq 10$  pp better than gru-h2;  $\Delta_2$  smallest GRU matching the ring has  $\geq 350$  parameters;  $\Delta_3$  ring reaches 25% loss with  $\leq 1/3$  of the sequences the best GRU needs;  $\Delta_4$  ring  $\geq 10$  pp better than the best GRU in N2; blocking  $\Delta_5$ : a degree-preserving shuffle of the recurrent connectivity is  $\geq 10$  pp worse than the ring, with an isomorphism control within 1 pp.

### 6.2 The first result

Table 7 is what the first article led with: the ring at 0.9% with 9 parameters, the best GRU (1,090 parameters) at 39.0%, no GRU budget reaching the ring. All five  $\Delta$ s were declared met:  $\Delta_5$  +44.49 pp (two-level bootstrap 95% CI [+26.61, +54.84]; ring 0.89% vs shuffled 55.10%),  $\Delta_1$  +99.52 pp,  $\Delta_2$  “no budget reaches”,  $\Delta_3$  ring at 2,048 sequences vs GRU never within 12,800,  $\Delta_4$  +70.44 pp [+65.50, +75.50], isomorphism 0.0033 pp.  $\Delta_5$  is not a budget artefact: the shuffled ring’s

<sup>1</sup>The first article described 55.9 pp as the gap between best and worst untrained estimator; the verification found it is the best trivial baseline’s loss against the optimum.

Table 8: GRU retrained on the same 12,800 sequences for several epochs (percentage loss, N1, fixed test set; no early stopping, best epoch read on the test set). Reproduced identically on 2026-09-16.

| model                             | 50 steps (= article) | 150  | 250  | 400 | 500         | best to 600 (N2)    |
|-----------------------------------|----------------------|------|------|-----|-------------|---------------------|
| gru-h16, seed 0, lr 0.03          | 34.1                 | 9.5  | 5.4  | 3.3 | 2.4         | 2.4 (11.4)          |
| gru-h16, seed 1, lr 0.03          | 44.3                 | 13.4 | 10.8 | 9.3 | 6.3         | 4.4 (17.1)          |
| gru-h32, seed 1, lr 0.01          | 55.6                 | 10.2 | 5.4  | 2.0 | <b>0.95</b> | 0.95 ( <b>3.0</b> ) |
| ring, median of 5 seeds (article) | 0.9                  |      |      |     |             | 0.9 (4.4)           |

own sample-efficiency curve plateaued at 54.4–54.6% from 4,096 sequences onward, while the intact ring reached 0.5% on the same ladder at 12,800 (0.9% in the main table, a different run); one of the five permutations happened to preserve enough structure to reach 10.1%, and the CI contains that seed. Two facts survive independently of any GRU comparison. Untrained: freshly built models on 1,000 N1 sequences score 54.2% (ring; min–max [50.6; 71.4]), 60.3% (64 wedges), 68.7% (shuffled), 99.7% (gru-h32) and 100.4% (rnn-h32); the bias is in the wiring and the shuffle already costs before any gradient. Ablations localise function: without rotation the model falls to the “last landmark” level (53.1% vs 55.2%); without recurrent excitation it loses 2.4% with a landmark every 3 s but 87.1% with none (ring 10.3%, 64-wedge ring 3.9%, all GRUs > 100%), the persistence-without-anchor role the biology assigns to the attractor (Seelig and Jayaraman, 2015; Noorman et al., 2024).

### 6.3 What adversarial review found

**The GRU had not been trained.** The loop drew a fresh batch of 256 per step in one pass: “12,800 sequences” was 50 Adam steps, and the 2,048-sequence calibration that chose each GRU’s learning rate was 8 steps; the literature the article cited trains such networks for  $10^4$ – $10^5$  steps (Cueva et al., 2020; Schaeffer et al., 2022). Every GRU curve that had started to learn was still falling 10–20 pp between the last two ladder points. gru-h32 was the acute case: learning rate 0.1 chosen on one seed, two of five evaluation seeds trained (34%, 27%) and three diverged; “[27.0; 100.6]” was a mixture of two populations. The reviewer reran the same 12,800 sequences for 12 epochs with the same learning rates (Table 8): a GRU-h32 retrained for 500 steps (one seed, best epoch selected on the test set, no early stopping) reached 0.95% in N1 and 3.01% in N2, matching the ring in N1 and beating it in N2 (0.9% / 4.4%). The ring converges in 50 steps because it starts at 54% and tunes nine gains; giving the comparator “the same budget” gave it less.

**The right comparator has 3 parameters, and it beats the ring.** The circKF has an explicit recurrent form: a 2-d state rotated by  $gu$ , shrunk by  $d$ , summed with  $g_v(m \cos z, m \sin z)$ . That vector integrator has three free scalars and no ring, wedges, inhibition or P-EN populations; it carries only the geometry the ring gets for free. With the article’s loop and 50 steps, 3 seeds, it scores 0.11 / 0.20 / 0.20% in N1 and 0.10 / 0.62 / 0.16% in N2 against the ring’s per-seed 0.8 / 1.1 / 0.3 / 0.9 / 1.1 and 3.0 / 4.4 / 4.6 / 5.2 / 3.5; it learns the Bayesian gain explicitly ( $g = 0.919/0.897/0.943$  against 0.917); untrained it sits at 2.4–4.8% against 54% for the ring. The honest sentence: SO(2) geometry solves the task with 3 parameters, and the ring is a 16-wedge discretisation whose six extra parameters cost precision. The report also noted that the ring’s N0 per-seed losses are 7.0 / 7.8 / 43.6 / 52.8 / 11.0%, two of five above 40%: the catastrophic drift of Noorman et al. (2024).

**Consequences.** Both comparators were re-executed in the article’s own worktree under a pre-registered criterion and reproduced the reviewer’s numbers exactly.  $\Delta_1$ ,  $\Delta_2$  and  $\Delta_4$  fall;  $\Delta_3$  is a lower bound and, as a crossing, untested.<sup>2</sup> What survives because it never depended on the GRU:

<sup>2</sup>The superseded article’s header differs from the record used here in two places: it quotes the retrained GRU at 5.4% (h16) and 1.25%/3.5% (h32), which are the epoch-5 and epoch-9 rows of the reviewer’s log, whereas the reproduction of 2026-09-16 confirms 0.95%/3.01% at epoch 10; and it lists  $\Delta_3$  among the survivors, whereas the referee report classifies it as not tested because sample efficiency was never measured with training to convergence. The later sources are used.

$\Delta_5$ , the isomorphism control and the untrained inductive-bias curve. The rule that came out of this episode is the third row of Table 1.

#### 6.4 Where the learned ring disagrees with the fly

Training moved every one of the nine parameters: the shift per step went from  $19.3^\circ$  (about one  $22.5^\circ$  wedge) to  $90.2^\circ$ ,  $3.9\times$  the task’s typical turn; the bump narrowed from  $80.6^\circ$  to  $52.3^\circ$  full width at half maximum, against  $82.3 \pm 11.5^\circ$  measured with a stripe by Seelig and Jayaraman (2015); the landmark gain rose from 0.90 to 5.26. For this task the optimal tuning is not the fly’s; the biology fixed the class of solution, not the values.

#### 6.5 Dopaminergic modulation through the fan-shaped body

A whole-brain dopamine map (Appendix C) found that dopaminergic neurons reach the central complex mainly through the fan-shaped body (1,651 edges / 12,086 synapses DAN→FB and 3,881 / 43,995 FB→columnar neurons, against 1,173 direct synapses). A pre-registered one-parameter modulation of the landmark gain by a causal “surprise”, which reduces exactly to the ring at zero, gave paired N2 differences of  $-0.256$ ,  $+0.636$  and  $-1.193$  pp over 3 seeds: two of three worse, and the pre-registered criterion, written as a difference of separate medians, technically passes on one seed and was not used. The result is inconclusive, probably null.

#### 6.6 Chess with a spatial grid derived from the ring principle

GradeEspacial adapts the ring’s principle (fixed local excitation, global inhibition, divisive normalisation, parameter-free readout) to a static  $8 \times 8$  grid with Chebyshev distance, settling for  $K = 4$  steps; it is a derivative, not the central-complex circuit. The full Lichess puzzle database (6,100,952 puzzles) was streamed, 500,000 puzzles (1,000,000 examples) encoded under a pre-registered CPU ceiling, a fixed test set of 199,998 examples held out, and 800,002 examples left for training. All numbers in Table 9 come from a single training seed and split (seed 20260916) with a fixed epoch budget per arm; a pre-registered adversarial revision (2026-09-27) retrained the MLP to convergence at the two smallest sizes and identified the grid’s fixed point, and both corrections are folded in below. The pre-registered criteria for the 9-parameter v1 (improves with scale by  $\geq 0.05$ ; within 0.03 of the MLP; beats its ablation by  $\geq 0.02$ ) all failed: its MCC is 0.1300 at both 200,000 and 800,002 examples, identical to seven decimals, and it never separates from its no-excitation ablation. The MLP cells at 2,000 and 10,000 in Table 9 are undertrained: the library’s early stopping on validation accuracy cut the original MLP at epoch 10 at both sizes, with best validation accuracies of 0.470 and 0.474, below chance, the GRU failure of Section 6.3 again. Retrained to convergence (validation-loss early stopping, patience 20, 3 seeds, same fixed test set) the MLP reaches a median MCC of 0.0970 (seeds 0.1268 / 0.0970 / 0.0883) at 2,000 examples against 0.1185 for the grid and 0.1841 for logistic regression, and 0.1918 (0.1918 / 0.2010 / 0.1827) at 10,000 against 0.0769 and 0.2286: at 2,000 the grid still edges the converged MLP by 0.02 with one of three seeds reversing the sign, at 10,000 the MLP beats the grid in all three seeds, and logistic regression beats both at both sizes. The value 0.1300388 shared by v1 at 200,000 and by v2 and v3 at 2,000 is identified: the three models make 100.00% identical predictions on the test set, and the rule is  $\text{pred} = 1$  iff  $\max(v_{\text{from}}, 0) + \max(v_{\text{to}}, 0) > 2$ , i.e. a white non-pawn piece stands on the from- or the to-square; it reproduces the MCC to 16 digits with agreement 1.0000. The cause is the encoding: a signed value per square (+white, −black) followed by  $\text{ReLU}(g_v \cdot \text{input} + b)$  with  $g_v > 0$  and  $b \approx 0$  zeroes every black piece, so the grid is blind to half the pieces and the decision lives in the linear term, which is why the ablation ties. Post hoc, the same rule family with threshold 4.5 matches v1 at 2,000 (0.1185077) and threshold 3.25 the exact 0.07692550749 of v1 at 10,000 and of two ablation cells.

With 800,002 examples, four geometric kernels (king, rook, bishop, knight; 33 parameters) lift the grid from 0.130 to 0.313 [0.3092, 0.3167], above logistic regression (0.2431 at full data; 0.2443 at 200,000) and its own ablation (0.1646); pawn and queen kernels (v3, 49 parameters) add +0.0271 to 0.3401, below the pre-registered floor of 0.03; below 200,000 examples the kernels add nothing (v2

Table 9: Spatial grid on Lichess puzzle move classification, MCC on a fixed test set of 199,998 examples, by training size; one training seed and split. v1: 9 parameters, one Chebyshev kernel. v2: four geometric kernels (king, rook, bishop, knight), 33 parameters. v3: v2 plus pawn and queen kernels, 49 parameters. Ablations switch off local excitation. <sup>†</sup> undertrained (early stopping at epoch 10); converged median 0.097 / 0.192 (3 seeds; Section 6.6).

| train size                                                    | logistic | strong MLP         | v1 grid | v1 ablation | v2 grid | v2 ablation | v3 grid      | v3 ablation |
|---------------------------------------------------------------|----------|--------------------|---------|-------------|---------|-------------|--------------|-------------|
| 2,000                                                         | 0.184    | 0.003 <sup>†</sup> | 0.119   | 0.119       | 0.130   | 0.130       | 0.130        | 0.130       |
| 10,000                                                        | 0.229    | 0.019 <sup>†</sup> | 0.077   | 0.077       | 0.111   | 0.085       | 0.070        | 0.083       |
| 50,000                                                        | 0.243    | 0.388              | 0.096   | 0.124       | 0.076   | 0.077       | 0.078        | 0.077       |
| 200,000                                                       | 0.244    | 0.467              | 0.130   | 0.130       | 0.130   | 0.147       | 0.130        | 0.130       |
| 800,002                                                       | 0.243    | <b>0.564</b>       | 0.130   | 0.130       | 0.313   | 0.165       | <b>0.340</b> | 0.171       |
| any capture rule (0 parameters): 0.394 at every training size |          |                    |         |             |         |             |              |             |

sits on the same fixed rule at 2,000, v1 at 200,000; the other cells were not matched example by example). A zero-parameter baseline that none of the three rounds had measured, “the move is a capture”, scores MCC 0.3942 on the same test set: above every grid variant ( $\leq 0.3401$ ), above logistic regression (0.2443) and above the MLP up to 50,000 examples (0.3879); only the MLP at 200,000 and 800,002 (0.4674 / 0.5637) beats it. The comparison that matters is that the geometric-kernel grid at 800,002 examples beats logistic regression but loses to a zero-parameter capture rule and to a dense MLP; the earlier reading that geometry beat data does not survive, and the task carries a cheap shortcut (the puzzle’s move is a capture far more often than a sampled legal move) that any future comparison must declare. Classification skill did not become playing strength: in Elo tournaments (30 games per pair, Bradley–Terry with bootstrap CIs, random mover at 1000, move = argmax of the classifier) the grids, a greedy material heuristic and a REINFORCE self-play agent (400 games) were all indistinguishable from random; 2-ply alpha–beta search lifted only the material evaluator out of the tie (1466.6 [1084.6, 1578.9]), while search over the learned evaluators (v2 1019.8 [931.7, 1522.4], v3 1143.8 [983.3, 1563.6]) did not, because a move classifier read as a position evaluator is a noisier leaf signal than material.

## 7 What the protocol caught

Table 10 is the record that motivates the protocol: errors that produced a number that looked publishable or was believed for some time, what exposed each, and the rule it left; the remaining thirteen rows are in Table 14. Most detectors were cheap (an ablation, a lookup, a count of ties, a look at the raw per-seed curve); the errors that took longest to find were the ones where the subjective state was confidence rather than doubt, and adversarial review with a mandate to run experiments found the three worst. The rules are not new in isolation (single-quantity metrics being gameable is Goodhart’s law; continual-learning benchmarks have paired forgetting with intransigence since Lopez-Paz and Ranzato, 2017; Chaudhry et al., 2018); what is new is the record of each one being needed, in a small program, within two weeks.

## 8 Limitations

- **Two circuits and one derivative, one to six tasks each** (the MB: text, chess, the connectome task, photographs, recall and clones; the ring and the grid: one task each). No general law about “biological architecture” is licensed; the first article’s thesis (“structure buys efficiency where the geometry coincides and charges rigidity outside it”) was retired as a law stated on two cases. One task per circuit cannot separate a property of the circuit from a property of the task.
- **The ring is canonical topology, not measured wiring**; the comparison with the measured central complex was not made (Kutschireiter et al., 2023 report the two indistinguishable in their setting).

Table 10: Non-trivial errors caught during the program, the detector, and the rule left.

| error (with the number it produced)                                                                                                                                                                                                                            | detector                                                                                       | rule                                                                                                                         |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| Naive null flips the sign: +13.53 pp against a shuffle, -7.11 pp against a degree-preserving shuffle                                                                                                                                                           | running both nulls                                                                             | degree-preserving null is mandatory; the naive null is only the weak control                                                 |
| Degree-preserving permutation of an already-random projection: gate +0.007 [+0.000, +0.014], uninformative by construction                                                                                                                                     | two reshuffle seeds inside the intact circuit’s CI; code shuffled against items falls to 0.197 | such a null only tests wiring that is not random by construction; otherwise shuffle codes against items and ablate the layer |
| CSV rows treated as edges: the null received 83% more partners and 49,236 plastic parameters against 26,937                                                                                                                                                    | checking the premise “one row is one edge” against the file                                    | aggregate by neuron pair; assert equal parameter counts                                                                      |
| Strong baseline undertrained, twice: “12,800 sequences” was 50 Adam steps (8 in calibration) and gru-h32 went from 98.7% to 0.95% when retrained; the chess MLP was cut at epoch 10 by accuracy-based early stopping (validation accuracy 0.470, below chance) | adversarial review reran both with a convergence criterion; raw curves had not plateaued       | the strong baseline gets more budget than the proposed model; convergence check before any crossing claim                    |
| Missing minimal comparator, twice: a 3-parameter vector integrator scores 0.11–0.20% against the ring’s 0.9%; a zero-parameter rule (“the move is a capture”, MCC 0.3942) beats every chess grid and logistic regression                                       | reviewer’s comparator ladder; a sweep of fixed rules of the task                               | the minimal model with the same geometry, and the zero-parameter rules of the task, must be in the table                     |
| “Dense” format, sparse content: 897-d chess vector 2.3% non-zero, 86.5% of KCs with zero active claws                                                                                                                                                          | content-density measurement                                                                    | density means content, not format; measure before running a KC layer                                                         |
| Tie credit by array index favoured coarse methods (pHash 0.904 → 0.874; the ablation tied with 40 candidates per query)                                                                                                                                        | counting ties                                                                                  | tie credit by expected value                                                                                                 |
| Criterion written as a difference of medians passes while 2 of 3 paired seeds worsen                                                                                                                                                                           | reading the pairing                                                                            | criteria written as paired differences                                                                                       |

- **Seeds.** Three in EXP-002, EXP-007, the vector-integrator and dopamine comparators; five in the main ring, connectome and recall experiments; ten in EXP-003; one projection seed in the photograph experiment; one training seed and split in the whole chess grid study (Table 9; only the MLP retrain of the 2026-09-27 revision used 3 seeds); one run in the clone experiment. Intervals are bootstraps over seeds and test items, not over re-collection.
- **Retrained GRU.** The GRU that matches the ring was retrained for 500 steps on one seed with the best epoch selected on the test set and no early stopping; it shows that the original comparison was unfair, not what a tuned GRU achieves.
- **Robustness of the ring** to turn noise was swept only against the undertrained gru-h32 (3 seeds) without the trivial baseline in the table, so the inversion reported in the first article ( $-1.3$  pp at  $\sigma_u = \sigma_\Delta$ ) is not a comparison this paper relies on; the closed-form mechanism (one learned gain applied at all noise levels) was asserted, not tested.
- **Scale.** Everything ran on CPU. The MB’s published advantage is sublinear cost where exact search does not fit; no experiment reached that scale, so the decision rule’s third line is untested, not refuted.
- **Dense images.** Where the encoder is alive (EXP-003), the KC layer’s +26.4 pp is measured against a degenerate ablation that stops learning after task 1 and is not evidence of capability; no strong adversary was run and MNIST was not run.
- **Photographs.** One transformation (background removal) on a largely white-background

catalogue (152 of 160 audited primary photos).

- **Chess and code.** The grid’s signed-value encoding made it blind to black pieces, so every cell matched example by example (v1 at 2,000 and 200,000, v2 and v3 at 2,000, three ablation cells) measures a two-square rule, not geometry; the remaining small-data cells were not matched, and the v2/v3 results at full data are the only regime where the grid demonstrably left that rule, and they were not investigated further. The grid evaluator was trained as a move classifier and read as a position evaluator; tournament intervals span hundreds of Elo. The clone benchmark version lacks clone types, so the pre-registered Type-4 slice was untestable, and the split is by function rather than by project.
- **Provenance.** The pre-registration-before-training order is verifiable only with the original branches and the ledger, because the work entered the main line by squash; the ledger was per-worktree on the first day; the clone run’s tree was dirty by two non-numeric edits. The first day compressed pre-registration, execution, drafting and review into hours.

## 9 Disclosure of agentic execution

The experiments, code, analyses and the LaTeX manuscript were produced by LLM-based agents operating under the author’s direction inside a version-controlled protocol; this manuscript was drafted by an LLM agent from the result files listed in Appendix A. The direction of the programme, the choice of the two circuits and of the domains to test them in, the request for each new experiment, and the acceptance of every result reported here are the author’s. The repository does not record which individual thresholds, amendments or retractions were set by the author rather than proposed by an agent and accepted, and no such attribution is claimed; the pre-registration files themselves are written in the agents’ voice. The model family recorded for the experiments, reviews and this text is Claude; the wider repository also employed Codex, Qwen, Grok and Gemini agents, and the programme’s files do not record whether any of those touched this work. Exact version identifiers are not recorded for every run and are not asserted here. The adversarial reviewers and the numeric verifiers were LLM agents of the same family as the experimenters, working in separate checkouts with separate mandates. Every reference was resolved against its primary source via the arXiv, Crossref and PubMed APIs on 2026-09-27. `fontes.md` (number-to-source map), the referee and verification reports on the first article (`PARECER.md`, `VERIFICACAO.md`) and on this manuscript (`PARECER-ARXIV-2026-09-27.md`, `VERIFICACAO-ARXIV-2026-09-27.md`), the pre-registrations, the result files and the run ledger are packaged as ancillary files of this submission by `montar_anc.py`, which lists every file it includes. The author reviewed the manuscript before submission and is responsible for its conclusions.

## 10 Conclusion

Of the transpositions tested with a strong control, none beat the corresponding trivial or minimal adversary: the MB loses to logistic regression and to a dictionary on text, to an MLP on chess and on a synthetic conjunctive task, to a linear model on code clones, and to its own degree-preserving shuffle when its measured wiring is used; the ring loses to a 3-parameter integrator that carries only its geometry and to a GRU that was allowed to train; the KC layer is the sample estimate of the attention it was meant to replace. What the MB is, by measurement, is a near-duplicate recogniser whose value begins where exact search stops fitting; what the ring keeps is that topology matters relative to itself shuffled (+44.49 pp) and that the bias is present before any gradient; what the grid shows is that a 33-parameter geometric grid beats a linear model only at full data, and loses there to a zero-parameter capture rule and to a dense network. The contribution offered to whoever repeats this kind of work is the apparatus and the record of why each part of it exists, above all a strong baseline that receives more budget than the model being proposed.

## A Reproducibility

All paths are relative to the repository root and all listed files are included in the ancillary archive. The ledger `MOSCA/pesquisa/ledger.jsonl` (versioned since commit `c3b98aee3`) holds one entry per run (25 entries covering EXP-001 to EXP-009 and the chess revision, some experiments with more than one run, two hosts) with commit, dirty-tree flag, command, machine, dataset SHA, seeds, wall and CPU time and peak memory. String hashing is CRC32 and `PYTHONHASHSEED` is fixed throughout. The FlyWire file `connections_princeton.csv.gz` is pinned by SHA-256 (`445f996b...5bd48b`) and downloaded from FlyWire Codex (FAFB v783); text, chess and clone data are external and rebuilt by the listed scripts.

Table 11: Files, registration and seeds of each experiment (paths under `MOSCA/`).

| experiment               | files ( <code>MOSCA/...</code> )                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | registration, seeds                                                                                                           |
|--------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| MB text, EXP-002         | <code>comum.py</code> ; <code>pesquisa/exp002_memoria_instancia.py</code> ; <code>pesquisa/CORRECOES.md</code> ; 9 tests in <code>tests/test_comum.py</code>                                                                                                                                                                                                                                                                                                                                                                                                  | 3 seeds; group split by instance                                                                                              |
| Fair retest, EXP-006/007 | <code>pesquisa/swe-justo/</code> : <code>PRE-REGISTRO.md</code> , <code>codificar.py</code> , <code>exp006_swe_justo.py</code> , <code>exp007_prequencial_denso.py</code> , <code>mosca_densa.py</code> ; 7 tests in <code>tests/test_mosca_densa.py</code>                                                                                                                                                                                                                                                                                                   | commit <code>4c1388b24</code> ; data SHA <code>c250424ff4a56b5d</code> ; splits 1789/1790/1791; runs at 23f1d0a66, clean tree |
| Continual, EXP-003/004   | <code>pesquisa/exp003_continuo.py</code> , <code>exp004_graduado.py</code>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 10 seeds; -kcs 500,2000, seeds 100 + $s$                                                                                      |
| Chess, MB                | <code>pesquisa/xadrez/</code> : two pre-registrations, <code>experimento.py</code> , <code>experimento_v2.py</code> , two result JSONs                                                                                                                                                                                                                                                                                                                                                                                                                        | split by puzzle                                                                                                               |
| Connectome vs null       | <code>pesquisa/conectoma/</code> : <code>PRE-REGISTRO.md</code> , <code>extrai.py</code> , <code>circuito.py</code> , <code>experimento.py</code> , <code>resultado.json</code> , <code>resultado-linha-como-aresta.json</code>                                                                                                                                                                                                                                                                                                                               | commit <code>f93c9f04b</code> ; seeds 0–4; 53 min                                                                             |
| Attention, EXP-005       | <code>pesquisa/atencao/</code> : <code>PRE-REGISTRO.md</code> , <code>experimento.py</code> , <code>resultados/exp005.json</code> ; 5 tests in <code>tests/test_atencao.py</code>                                                                                                                                                                                                                                                                                                                                                                             | commit <code>a06064412</code> ; 5 seeds                                                                                       |
| Photographs, EXP-008     | <code>aplicacoes/fotos/</code> : <code>dados.py</code> , <code>experimento.py</code> , <code>estresse.py</code> , <code>resultados.json</code> , <code>resultados-estresse.json</code> ; 12 tests in <code>tests/test_fotos.py</code> ; thumbnails rebuilt by <code>dados.py</code> from the manifest, not shipped                                                                                                                                                                                                                                            | projection seed 1789; not pre-registered                                                                                      |
| Code clones, EXP-009     | <code>pesquisa/clones/</code> : <code>PRE-REGISTRO.md</code> , <code>RESULTADO-2026-09-27.md</code> , <code>extrair.py</code> , <code>codificar.py</code> , <code>experimento.py</code> , <code>resultados/exp009.json</code> , <code>exp009.predicoes.npz</code>                                                                                                                                                                                                                                                                                             | pre-registered 2026-09-23; scripts at <code>ffaa50c2e</code> ; seed 1789; data SHA <code>c1dd266ba4f25744</code>              |
| Ring attractor           | <code>pesquisa/anel/</code> : <code>PRE-REGISTRO.md</code> , <code>verificacao.py</code> , <code>experimento.py</code> , <code>relatorio.py</code> , <code>parecer_comparadores.py</code> , <code>testar_fb_mod.py</code> ; <code>resultados/registro.json</code> , <code>R_por_sequencia.npz</code> (all $\Delta s$ recomputable in seconds)                                                                                                                                                                                                                 | commit <code>793c5482f</code> ; seeds 0–4, calibration 100, test 424242, bootstrap $31337 \times 2,000$ ; 116 runs            |
| Spatial grid             | <code>pesquisa/xadrez-espacial/</code> : six pre-registrations (one, fluxo-contínuo, never run) and five results, plus the revision pair <code>PRE-REGISTRO-REVISAO-2026-09-27.md</code> / <code>RESULTADO-REVISAO-2026-09-27.md</code> ; <code>modelo_espacial{,_v2,_v3}.py</code> , <code>experimento_escala{,_v2}.py</code> , <code>torneio_elo.py</code> , <code>rl_autojogo.py</code> , <code>busca_alfabeta.py</code> , <code>revisao_2026_09_27.py</code> ; <code>resultados/curva_escala{,_v2,_v3}.json</code> , <code>revisao-2026-09-27.json</code> | split seed 20260916; MLP retrain seeds 0–2                                                                                    |
| Review                   | <code>pesquisa/PARECER.md</code> , <code>VERIFICACAO.md</code> (first article); <code>pesquisa/arxiv/PARECER-ARXIV-2026-09-27.md</code> , <code>VERIFICACAO-ARXIV-2026-09-27.md</code> (this manuscript); <code>LEDGER.md</code> , <code>ARTIGO.md</code> (superseded, kept), <code>artigo/artigo.html</code> ; <code>RESULTADO-SWE-SMITH.md</code> (invalid, kept)                                                                                                                                                                                           |                                                                                                                               |

## B Additional tables and the continual-learning experiment

Table 12: EXP-002, SWE-smith, prequential protocol, 49,773 items grouped by instance, 3 seeds, MCC (standard errors  $\leq 0.03$  omitted). LOO oracle +0.841.

| exposure band | fly (MB) | no KC  | hash table    | 1-NN   | logistic      |
|---------------|----------|--------|---------------|--------|---------------|
| never seen    | +0.212   | +0.240 | -0.012        | +0.202 | <b>+0.268</b> |
| seen once     | +0.249   | +0.288 | <b>+0.936</b> | +0.789 | +0.368        |
| seen 1-3×     | +0.257   | +0.294 | <b>+0.874</b> | +0.767 | +0.379        |
| all           | +0.267   | +0.290 | +0.590        | +0.587 | +0.371        |

EXP-003 (**digits**,  $8 \times 8$ , 10 classes, five two-class tasks in sequence, dense 64-d input, 10 seeds) was pre-registered with two thresholds: freezing wins if it reduces forgetting by  $\geq 10$  pp, and the KC layer is useful if the no-KC arm is  $\geq 5$  pp worse. Both passed (forgetting reduction +86.4 pp; KC contribution +26.4 pp, CI  $\pm 3.8$ ; encoder alive, Jaccard 0.389 within vs 0.258 between classes) and the interpretation fell when the diagonal of the task matrix was inspected (Table 13): the frozen model stops learning, and the no-KC arm freezes 87.5% of its units on task 1 and scores zero on every later task with “0.0% forgetting”. Because the ablation is degenerate, the +26.4 pp is not evidence that the KC layer helps. Graded protection ( $\eta/(1 + \lambda \cdot \text{use})$ ) retains plasticity (diagonal 97.9%) and halves forgetting (36.5%) at 2,000 KCs,  $\lambda = 10$ , but no configuration met the joint criterion (diagonal  $\geq 70\%$  and forgetting  $\leq 20\%$ ). The verification found this table had no ledger record (re-execution reproduced 18 of 18 cells) and that the freezing rule implemented (protect the KCs that participated in a task) differs from the FlyModel rule as summarised in the program’s review (freeze all KC→MBON weights except those of active KCs; Shen et al., 2023); the discrepancy is recorded, not resolved.

Table 13: EXP-003, class-incremental continual learning, 10 seeds; “diagonal” is accuracy on each task right after training it (seed 0).

| arm                     | final accuracy   | forgetting       | diagonal (%)                                                             |
|-------------------------|------------------|------------------|--------------------------------------------------------------------------|
| MB with binary freezing | $46.8 \pm 3.2\%$ | $0.5 \pm 0.2\%$  | $99 \cdot 61 \cdot 38 \cdot \mathbf{16} \cdot 18$                        |
| MB without freezing     | $29.6 \pm 5.0\%$ | $86.9 \pm 6.4\%$ | $99 \cdot 100 \cdot 100 \cdot 99 \cdot 98$                               |
| no KC (ablation)        | $20.3 \pm 0.6\%$ | $0.0 \pm 0.1\%$  | $98 \cdot \mathbf{0} \cdot \mathbf{0} \cdot \mathbf{0} \cdot \mathbf{0}$ |
| online SGD              | $23.2 \pm 2.5\%$ | $94.5 \pm 3.2\%$ | —                                                                        |

## C Whole-brain dopamine

The MB extract of FlyWire v783 labels 331 neurons as DANs. The complete connectome’s neurotransmitter prediction gives 584 dopaminergic neurons of 139,255 classified (**da\_avg** 0.34–0.98, median 0.72); 285 of the 331 extract DANs (86%) are among them, and 299 of the 584 (51%) lie outside the MB extract. Of their 495,583 synapses, 115,297 (23.3%) are in MB compartments; the largest single regions are the right anterior ventrolateral protocerebrum (41,421; 8.4%) and the fan-shaped body (39,687; 8.0%), and the ellipsoid body is not in the top fifteen. The 584 span 114 named types (PAM 261, PPL 24, PPM 12; the rest include central-complex columnar types, PFNp 45 and PFR 17, and 60 **eye\_bristle** mechanoreceptors that are a probable false positive of the prediction, reported rather than pruned). The single global dopamine of the MB models is therefore a strong simplification: at least 114 DAN types, three destinations of comparable weight, and more than half of the fly’s dopaminergic neurons outside the modelled circuit.

## Data availability

Code, the run ledger, pre-registrations, result files, **fontes.md**, and the four review files (PARECER.md and VERIFICACAO.md on the first article; PARECER-ARXIV-2026-09-27.md and

Table 14: Remaining errors caught during the program (continuation of Table 10).

| error (with the number it produced)                                                                                                                                    | detector                                                                      | rule                                                                                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| Row split with 3.7 trajectories per instance: a dictionary on <code>instance_id</code> scored +0.374; 59.6% of test instances were in training                         | the trivial baseline; the LOO oracle (+0.841)                                 | group split; lookup on the key as leakage sentinel                                       |
| Dead encoder: 93.5% of active KCs had one active claw (0.09 active inputs per KC)                                                                                      | ablation: removing the KC layer                                               | encoder diagnostic before the learner; blocking ablation                                 |
| Process-salted <code>hash()</code> : +0.296, +0.297, +0.282, +0.274 were four networks                                                                                 | changed MCC by 0.013                                                          | CRC32; explicit seeds; PYTHONHASHSEED                                                    |
| Stream-shift test with a zero ceiling by construction ( $P(y' x) = \frac{1}{2}$ )                                                                                      | non-reproducible reruns                                                       | print the ceiling; script deleted                                                        |
| Three single-quantity metrics gamed in one day: F1 (F1 0.550 at 38% accuracy), “recover 80% of own level”, forgetting reduction (+86.4pp with diagonal 99/61/38/16/18) | analytic check of the trigger (167 of 200 hashed characters were boilerplate) | joint criteria with a floor on each quantity                                             |
| Subset preview: KC contribution +0.081 on 6,000 lines, −0.023 on the full file                                                                                         | MCC; the task-matrix diagonal                                                 |                                                                                          |
| Six configurations (18 cells) with no ledger record (EXP-004)                                                                                                          | running the full file                                                         | never conclude from a subset of a blocked file                                           |
| Column labelled “IC95” was the min–max across seeds                                                                                                                    | numeric verification; re-execution reproduced 18 of 18                        | one versioned ledger (five divergent copies and two unrelated “EXP-005” had accumulated) |
| Literature mechanism paraphrased inverted (freeze recruited vs freeze non-active KCs)                                                                                  | numeric verification                                                          | label ranges as ranges                                                                   |
| Front-end search with fixed regulariser measured regularisation, not capacity (+0.124 vs +0.203)                                                                       | noticing that $n_{KC}$ is the readout’s parameter count                       | flagged; unresolved                                                                      |
| Optimum on the grid edge                                                                                                                                               | exploratory extension on validation only                                      | joint search; run discarded whole                                                        |
| z-score inherited from an upstream stage applied before a cosine baseline (SVD baseline 0.365 vs 0.502 without it)                                                     | raw-cosine extra arm                                                          | reported as exploratory; post-hoc wins recorded, not counted                             |
| Calibration on puzzle positions                                                                                                                                        | running it                                                                    | the reference cosine is on the raw vector; standardise only upstream of the KC layer     |
| under-predicted tournament time by 2× (17,301 s)                                                                                                                       |                                                                               | calibrate on the distribution that will be run                                           |

VERIFICACAO-ARXIV-2026-09-27.md on this manuscript) ship as ancillary files (`anc/`) of this submission. The FlyWire connectome is available from FlyWire Codex; SWE-smith trajectories from the SWE-bench project; Lichess puzzles from the Lichess open database; BigCloneBench in its CodeXGLUE form from the Hugging Face hub; the product photographs are the author’s catalogue, and the thumbnails are rebuilt by `dados.py` from the manifest and are not shipped.

## References

- Hoger Amin, Anthi A. Apostolopoulou, Raquel Suárez-Grimalt, Eleftheria Vrontou, and Andrew C. Lin. Localized inhibition in the *Drosophila* mushroom body. *eLife*, 9:e56954, 2020. doi: 10.7554/eLife.56954.
- Simran Arora, Sabri Eyuboglu, Aman Timalsina, Isys Johnson, Michael Poli, James Zou, Atri Rudra, and Christopher Ré. Zoology: Measuring and improving recall in efficient language models. arXiv preprint arXiv:2312.04927, 2023.
- Simran Arora, Sabri Eyuboglu, Michael Zhang, Aman Timalsina, Silas Alberti, Dylan Zinsley, James Zou, Atri Rudra, and Christopher Ré. Simple linear attention language models balance the recall-throughput tradeoff. arXiv preprint arXiv:2402.18668, 2024. Based.
- Trenton Bricken and Cengiz Pehlevan. Attention approximates sparse distributed memory. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, 2021.
- Trenton Bricken, Xander Davies, Deepak Singh, Dmitry Krotov, and Gabriel Kreiman. Sparse distributed memory is a continual learner. In *International Conference on Learning Representations (ICLR)*, 2023.
- Sophie J. C. Caron, Vanessa Ruta, L. F. Abbott, and Richard Axel. Random convergence of olfactory inputs in the *Drosophila* mushroom body. *Nature*, 497(7447):113–117, 2013. doi: 10.1038/nature12063.
- N. Alex Cayco-Gajic and R. Angus Silver. Re-evaluating circuit mechanisms underlying pattern separation. *Neuron*, 101(4):584–602, 2019. doi: 10.1016/j.neuron.2019.01.044.
- Arslan Chaudhry, Puneet K. Dokania, Thalaiyasingam Ajanthan, and Philip H. S. Torr. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In *Computer Vision – ECCV 2018*, Lecture Notes in Computer Science, pages 556–572. Springer, 2018. doi: 10.1007/978-3-030-01252-6\_33.
- Krzysztof Choromanski, Valerii Likhoshesterov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins, Jared Davis, Afroz Mohiuddin, Lukasz Kaiser, David Belanger, Lucy Colwell, and Adrian Weller. Rethinking attention with performers. In *International Conference on Learning Representations (ICLR)*, 2021.
- Christopher J. Cueva, Peter Y. Wang, Matthew Chin, and Xue-Xin Wei. Emergence of functional and structural properties of the head direction system by optimization of recurrent neural networks. In *International Conference on Learning Representations (ICLR)*, 2020.
- Sanjoy Dasgupta, Charles F. Stevens, and Saket Navlakha. A neural algorithm for a fundamental computing problem. *Science*, 358(6364):793–796, 2017. doi: 10.1126/science.aam9868.
- Sanjoy Dasgupta, Timothy C. Sheehan, Charles F. Stevens, and Saket Navlakha. A neural data structure for novelty detection. *Proceedings of the National Academy of Sciences*, 115(51):13093–13098, 2018. doi: 10.1073/pnas.1814448115.
- Charles B. Delahunt and J. Nathan Kutz. Putting a bug in ML: The moth olfactory network learns to read MNIST. *Neural Networks*, 118:54–64, 2019. doi: 10.1016/j.neunet.2019.05.012. The “insect cyborg” front-end configuration is Delahunt and Kutz, “Insect cyborgs: Bio-mimetic feature generators improve machine learning accuracy on limited data”, arXiv:1808.08124.
- Nalin Dhiman. Topological sensitivity in connectome-constrained neural networks. arXiv preprint arXiv:2604.04033, 2026.
- Sven Dorkenwald, Arie Matsliah, Amy R. Sterling, Philipp Schlegel, Szi-chieh Yu, Claire E. McKellar, et al. Neuronal wiring diagram of an adult brain. *Nature*, 634(8032):124–138, 2024. doi: 10.1038/s41586-024-07558-y.
- Toshihide Hige, Yoshinori Aso, Mehrab N. Modi, Gerald M. Rubin, and Glenn C. Turner. Heterosynaptic plasticity underlies aversive olfactory learning in *Drosophila*. *Neuron*, 88(5):985–998, 2015. doi: 10.1016/j.neuron.2015.11.003.
- Brad K. Hulse, Hannah Haberkern, Romain Franconville, Daniel Turner-Evans, Shin-ya Takemura, Tanya Wolff, Marcella Noorman, Marisa Dreher, Chuntao Dan, Ruchi Parekh, Ann M. Hermundstad, Gerald M. Rubin, and Vivek Jayaraman. A connectome of the *Drosophila* central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife*, 10:e66039, 2021. doi: 10.7554/eLife.66039.
- Pentti Kanerva. *Sparse Distributed Memory*. MIT Press, Cambridge, MA, 1988. ISBN 9780262111324.
- Angelos Katharopoulos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. Transformers are RNNs: Fast autoregressive transformers with linear attention. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *Proceedings of Machine Learning Research*, pages 5156–5165, 2020.
- Sung Soo Kim, Hervé Rouault, Shaul Druckmann, and Vivek Jayaraman. Ring attractor dynamics in the *Drosophila* central brain. *Science*, 356(6340):849–853, 2017. doi: 10.1126/science.aal4835.

- Denis Kleyko and Dmitri A. Rachkovskij. On design choices in similarity-preserving sparse randomized embeddings. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2024. doi: 10.1109/IJCNN60899.2024.10651277.
- Dmitry Krotov and John J. Hopfield. Dense associative memory for pattern recognition. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, pages 1172–1180, 2016.
- Anna Kutschireiter, Melanie A. Basnak, Rachel I. Wilson, and Jan Drugowitsch. Bayesian inference in ring attractor networks. *Proceedings of the National Academy of Sciences*, 120(9):e2210622120, 2023. doi: 10.1073/pnas.2210622120.
- Janne K. Lappalainen, Fabian D. Tschopp, Sridhama Prakhya, Mason McGill, Aljoscha Nern, Kazunori Shinomiya, Shin-ya Takemura, Eyal Gruntman, Jakob H. Macke, and Srinivas C. Turaga. Connectome-constrained networks predict neural activity across the fly visual system. *Nature*, 634(8036): 1132–1140, 2024. doi: 10.1038/s41586-024-07939-3.
- Feng Li, Jack W. Lindsey, Elizabeth C. Marin, Nils Otto, Marisa Dreher, Georgia Dempsey, Ildiko Stark, Alexander S. Bates, Markus William Pleijzier, Philipp Schlegel, Aljoscha Nern, Shin-ya Takemura, Nils Eckstein, Tansy Yang, Audrey Francis, Amalia Braun, Ruchi Parekh, Marta Costa, Louis K. Scheffer, Yoshinori Aso, Gregory S. X. E. Jefferis, Larry F. Abbott, Ashok Litwin-Kumar, Scott Waddell, and Gerald M. Rubin. The connectome of the adult *Drosophila* mushroom body provides insights into function. *eLife*, 9:e62576, 2020. doi: 10.7554/eLife.62576.
- Yuchen Liang, Chaitanya K. Ryali, Benjamin Hoover, Leopold Grinberg, Saket Navlakha, Mohammed J. Zaki, and Dmitry Krotov. Can a fruit fly learn word embeddings? In *International Conference on Learning Representations (ICLR)*, 2021.
- Andrew C. Lin, Alexei M. Bygrave, Alix de Calignon, Tzumin Lee, and Gero Miesenböck. Sparse, decorrelated odor coding in the mushroom body enhances learned odor discrimination. *Nature Neuroscience*, 17(4):559–568, 2014. doi: 10.1038/nn.3660.
- Ashok Litwin-Kumar, Kameron Decker Harris, Richard Axel, Haim Sompolinsky, and L. F. Abbott. Optimal degrees of synaptic connectivity. *Neuron*, 93(5):1153–1164.e7, 2017. doi: 10.1016/j.neuron.2017.01.030.
- David Lopez-Paz and Marc’Aurelio Ranzato. Gradient episodic memory for continual learning. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017.
- Shuai Lu, Daya Guo, Shuo Ren, Junjie Huang, Alexey Svyatkovskiy, Ambrosio Blanco, Colin Clement, Dawn Drain, Daxin Jiang, Duyu Tang, Ge Li, Lidong Zhou, Linjun Shou, Long Zhou, Michele Tufano, Ming Gong, Ming Zhou, Nan Duan, Neel Sundaresan, Shao Kun Deng, Shengyu Fu, and Shujie Liu. CodeXGLUE: A machine learning benchmark dataset for code understanding and generation. arXiv preprint arXiv:2102.04664, 2021.
- Sergei Maslov and Kim Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569): 910–913, 2002. doi: 10.1126/science.1065103.
- Mehrab N. Modi, Yichun Shuai, and Glenn C. Turner. The *Drosophila* mushroom body: from architecture to algorithm in a learning circuit. *Annual Review of Neuroscience*, 43:465–484, 2020. doi: 10.1146/annurev-neuro-080317-0621333.
- Marcos Negre Saura, Richard Allmendinger, Wei Pan, and Theodore Papamarkou. Spatial-aware decision-making with ring attractors in reinforcement learning systems. In *Advances in Neural Information Processing Systems 38 (NeurIPS 2025)*, 2025.
- Marcella Noorman, Brad K. Hulse, Vivek Jayaraman, Sandro Romani, and Ann M. Hermundstad. Maintaining and updating accurate internal representations of continuous variables with a handful of neurons. *Nature Neuroscience*, 27(11):2207–2217, 2024. doi: 10.1038/s41593-024-01766-5.
- Parikshit Ram and Kaushik Sinha. Federated nearest neighbor classification with a colony of fruit-flies. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8036–8044, 2022. doi: 10.1609/aaai.v36i7.20775. FlyNN; AAAI-22 Technical Tracks, vol. 36 no. 7.
- Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Lukas Gruber, Markus Holzleitner, Thomas Adler, David Kreil, Michael K. Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. Hopfield networks is all you need. In *International Conference on Learning Representations (ICLR)*, 2021.
- Alfonso Renart, Pengcheng Song, and Xiao-Jing Wang. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron*, 38(3):473–485, 2003. doi: 10.1016/S0896-6273(03)00255-1.
- Chaitanya K. Ryali, John J. Hopfield, Leopold Grinberg, and Dmitry Krotov. Bio-inspired hashing for unsupervised similarity search. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *Proceedings of Machine Learning Research*, pages 8739–8750, 2020.

- Rylan Schaeffer, Mikail Khona, and Ila R. Fiete. No free lunch from deep learning in neuroscience: A case study through models of the entorhinal-hippocampal circuit. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, pages 16052–16067, 2022. Preprint: bioRxiv, doi:10.1101/2022.08.07.503109.
- Philipp Schlegel, Yijie Yin, Alexander S. Bates, Sven Dorkenwald, Katharina Eichler, Paul Brooks, et al. Whole-brain annotation and multi-connectome cell typing of *Drosophila*. *Nature*, 634(8032):139–152, 2024. doi: 10.1038/s41586-024-07686-5.
- Johannes D. Seelig and Vivek Jayaraman. Neural dynamics for landmark orientation and angular path integration. *Nature*, 521(7551):186–191, 2015. doi: 10.1038/nature14446.
- Jaiyam Sharma and Saket Navlakha. Improving similarity search with high-dimensional locality-sensitive hashing. arXiv preprint arXiv:1812.01844, 2018. DenseFly.
- Yang Shen, Sanjoy Dasgupta, and Saket Navlakha. Reducing catastrophic forgetting with associative learning: A lesson from fruit flies. *Neural Computation*, 35(11):1797–1819, 2023. doi: 10.1162/neco\_a\_01615. Journal version of the 2021 preprint “Algorithmic insights on continual learning from fruit flies”, arXiv:2107.07617.
- William E. Skaggs, James J. Knierim, Hemant S. Kudrimoti, and Bruce L. McNaughton. A model of the neural basis of the rat’s sense of direction. In *Advances in Neural Information Processing Systems 7*. MIT Press, 1995.
- Jeffrey Svajlenko, Judith F. Islam, Iman Keivanloo, Chanchal K. Roy, and Mohammad Mamun Mia. Towards a big data curated benchmark of inter-project code clones. In *2014 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pages 476–480. IEEE, 2014. doi: 10.1109/ICSME.2014.77.
- Daniel Turner-Evans, Stephanie Wegener, Hervé Rouault, Romain Franconville, Tanya Wolff, Johannes D. Seelig, Shaul Druckmann, and Vivek Jayaraman. Angular velocity integration in a fly heading circuit. *eLife*, 6:e23496, 2017. doi: 10.7554/eLife.23496.
- Pantelis Vafidis, David Oswald, Tiziano D’Albis, and Richard Kempter. Learning accurate path integration in ring attractor models of the head direction system. *eLife*, 11:e69841, 2022. doi: 10.7554/eLife.69841.
- Xiaohui Xie, Richard H. R. Hahnloser, and H. Sebastian Seung. Double-ring network model of the head-direction system. *Physical Review E*, 66(4):041902, 2002. doi: 10.1103/PhysRevE.66.041902.
- Kechen Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, 16(6):2112–2126, 1996. doi: 10.1523/JNEUROSCI.16-06-02112.1996.
- Zhihao Zheng, Feng Li, Corey Fisher, Iqbal J. Ali, Nadiya Sharifi, Steven Calle-Schuler, Joseph Hsu, Najla Masoodpanah, Lucia Kmecova, Tom Kazimiers, Eric Perlman, Matthew Nichols, Peter H. Li, Viren Jain, and Davi D. Bock. Structured sampling of olfactory input by the fly mushroom body. *Current Biology*, 32(15):3334–3349.e6, 2022. doi: 10.1016/j.cub.2022.06.031.