Temporal Counting in MaleCNS-Derived Artificial Recurrent Networks: An Independent Replication, Memory-Capacity Extension, and Synchronized Blackjack Simulation

Ogdn Ames

Arizona State University

September 13, 2026

Author Note

Ogdn Ames is the author of this manuscript. This revised version was prepared with AI-assisted editing, literature integration, quantitative cross-checking, and APA formatting from completed local computational experiments and archived results. The reported experimental outcomes are preserved from the source study rather than generated from external literature. The manuscript integrates an original study and a separately prospectively specified independent follow-up conducted September 13, 2026 UTC with distinct prospective locks. No peer review or journal acceptance of this manuscript is claimed. The modeled systems are artificial networks whose topology is derived from a connectome; they are not physiological fly brains, and the blackjack analyses are simulations, not evidence of real-world gambling profitability.

Abstract

Structural connectomes constrain anatomical topology but do not specify functional neural dynamics. An initial MaleCNS v1.0 study found that a primary fixed reservoir lost to random graphs, whereas a postsynaptic-normalized comparison favored native topology on average across only three computational seeds. We conducted a prospectively specified independent replication using ten new input/readout/null seeds on the same 5,000-neuron anatomical subset, 2,048 fresh training shoes, 512 validation shoes, and 1,024 locked test shoes. Native, random, and degree/raw-strength-preserving families received identical training budgets. Fixed baseline true-count root-mean-square error (RMSE) averaged 0.8085 for native models and 0.9990 for random models. The primary seed-generalized comparison favored native topology. A validation-only search over longer-memory linear dynamics and 64, 128, or 256 readout states yielded mean RMSEs of 0.0097, 0.0150, and 0.0190 for native, random, and raw-strength-null families. An original GRU transferred to fresh shoes achieved 0.0928; exact accumulation remained error-free. A separate paired simulation covered 4,096 fresh shoes and 173,440 initial rounds. No estimator-versus-exact paired interval excluded zero. A synchronized video maps frozen artificial states to real soma coordinates while showing exposed cards, count errors, and same-stream settled equity. Results concern constrained artificial computation, not biological counting or profitability. Single-specimen anatomy, effective signed-strength confounds, and restricted dynamics limit generalization.

Keywords: connectomics, reservoir computing, computational replication, temporal memory, blackjack simulation

Temporal Counting in MaleCNS-Derived Artificial Recurrent Networks: An Independent Replication, Memory-Capacity Extension, and Synchronized Blackjack Simulation

A structural connectome describes anatomical connectivity, but it does not, by itself, determine the functional dynamics of a nervous system. The MaleCNS v1.0 release extends adult Drosophila connectomics to a fully proofread male central nervous system spanning brain and ventral nerve cord (Berg et al., 2026; HHMI Janelia FlyEM, 2026). Related whole-brain analyses show that synapse-level graphs contain rich network organization, while causal and dynamical modeling work emphasizes that anatomical paths do not uniquely specify effective influence in vivo (Lin et al., 2024; Pospisil et al., 2024). Constructing artificial dynamics on the MaleCNS graph therefore permits controlled questions about topology while avoiding the stronger and unsupported claim that the resulting network reproduces a living fly brain.

Hi-Lo card counting provides a deliberately simple sequential benchmark. The system assigns +1 to ranks 2-6, 0 to ranks 7-9, and -1 to aces and ten-valued ranks; its practical appeal is that the running statistic can be updated by scalar addition (Thorp, 1966). An exact computational solution therefore requires only one persistent scalar state, making the task useful for testing temporal memory rather than intelligence. A model can obtain nonzero predictive performance without accurately accumulating a long sequence, so meaningful evaluation requires comparisons with no-memory predictors, exact computation, trainable recurrent networks, and matched graph null models.

The reservoir-computing framework is particularly suitable for this comparison because it separates fixed recurrent dynamics from a trained readout. Echo-state and liquid-state approaches showed that high-dimensional recurrent systems can retain fading information about prior inputs while learning is concentrated in the output mapping (Jaeger, 2001; Maass et al., 2002). A

connectome-derived reservoir is thus a controlled engineering abstraction: the anatomical graph supplies topology, but memory capacity and prediction quality still depend on weight transforms, dynamical timescales, input placement, and readout capacity.

The present study addressed three hypotheses. First, a fixed MaleCNS-derived recurrent system would retain information sufficient to improve count estimation over no-memory baselines. Second, native topology would produce different performance from matched randomized topologies. Third, performance would depend on the assumed dynamics and transformation of anatomical counts into functional weights. The corresponding null interpretation was that native topology would offer no distinctive advantage once the relevant modeling choices and training budgets were controlled. The study emphasized pure sequential counting; simulated table play and offline decision value were secondary evaluations.

The independent follow-up directly addressed the original study's limited computational replication and restricted readout. It separated confirmatory replication of the fixed normalized baseline from a prospectively bounded engineering extension. This distinction matters: increasing the number of trainable readout coefficients and changing the artificial memory timescale can improve a decoder without implying that anatomy alone supplies an optimal counting algorithm. All original and follow-up experiments were conducted on September 13, 2026 (UTC), in separately versioned analyses. The original checkpoint freeze preceded its test; the new protocol and fresh test-array lock preceded follow-up training. No old or new test outcome was used to select the new models.

The visualization was designed as a data audit as well as an explanation. Its brain panel displays artificial neuronal states at canonical soma coordinates. Its numerical panel displays the exact and estimated counts from precisely the same exposure sequence. Its equity panel advances

only when that sequence reaches a settled round. Thus, the animation does not combine a favorable card example with an unrelated profit trajectory.

Method

Original Study: Design and Research Transparency

The study used a preregistration-style local protocol, including dated pretest clarifications and amendments, following the general principle that confirmatory hypotheses and analysis decisions should be fixed before outcome inspection (Nosek et al., 2018). This document was not an externally registered preregistration. Training, validation, and final-test seed ranges were disjoint and committed before model selection. Selected checkpoints and fitted probes were frozen before final outcomes were inspected. Final arrays were generated after freezing from the previously committed seeds, rather than stored in advance. The validation-motivated normalization analysis was identified as secondary and frozen before final testing; it was not selected from final-test results.

The original primary sample comprised 2,048 training shoes, 256 validation shoes, and 512 final-test shoes. Each shoe contained six decks, with 75% penetration producing 234 observations. Three stochastic model/input/control seeds were used. The anatomical sampling permutation was fixed, so these seeds did not represent three biological specimens or three independently sampled anatomical subgraphs.

Data Source and Quality Control

The canonical MaleCNS v1.0 annotation, neurotransmitter, and connection-weight Feather files were obtained from the official distribution (HHMI Janelia FlyEM, 2026). The three files totaled 1,109,008,094 bytes. Acquisition records included URLs, release version, download times, SHA-256 checksums, and the CC BY 4.0 license. Official neuPrint

documentation supplied relevant infrastructure documentation (neuPrint Development Team, n.d.). The release is associated with the peer-reviewed description of the complete male Drosophila central nervous system connectome (Berg et al., 2026). The present computational study uses the released graph and annotations; connectome-acquisition or biological claims are attributed only where supported by the publication or official data documentation.

Quality control distinguished raw segmentation identifiers from annotated neurons. The complete connection table contained 151,856,684 directed rows and 88,384,522 segment identifiers, with total anatomical weight 311,833,243, 123 self-connections, and no duplicate rows. These segment counts must not be interpreted as neuron counts. The annotation table contained 211,577 rows; retaining status Traced yielded 165,122 neurons and a neuron-only graph with 25,563,197 edges.

Primary modeling used a nested, uniformly sampled 5,000-neuron induced subgraph selected using a fixed seed of 1729, independently of counting performance. This subgraph contained 23,255 raw edges. Its largest strongly connected component comprised 3,757 neurons before signed exclusions and 2,495 afterward. Induced sampling therefore changed the recurrent structure substantially relative to the complete graph.

Counting Task and Information Restrictions

Card ranks were represented by 13-dimensional one-hot vectors. Hi-Lo assigned +1 to ranks 2–6, 0 to ranks 7–9, and −1 to aces and ten-valued ranks. The oracle running count was the cumulative sum of these values. Unrounded true count equaled running count divided by remaining cards/52. Neural readouts predicted running count; true count was then calculated using the public remaining-card denominator. Thus, the network was not required to learn division or independently infer shoe depth.

Only the currently exposed rank entered the recurrent state update. Future cards, oracle counts, hidden ranks, shoe permutations, and random-number-generator state were excluded. States reset at shuffling. Running-count integer accuracy used nearest-integer rounding with ties to even; true count remained continuous. Separate current-card-only and no-information controls evaluated performance without cumulative rank history.

Reservoir Construction and Dynamics

Connectivity matrices used the orientation W[post, pre]. The update was $x(t + 1) = (1 - a)x(t) + a\varphi(gWx(t) + Bu(t))$, where a was leak, g was recurrent gain, B was a fixed input projection, and φ was linear or hyperbolic tangent. Recurrent connectivity remained fixed in the primary regime. A ridge-regression readout (Hoerl & Kennard, 1970) used 64 randomly sampled states and an intercept, giving 65 trained coefficients. A 128-state output ablation increased this count to 129.

Anatomical synaptic counts were not treated as measured conductances. Evaluated transforms included raw counts, square roots, log1p, normalized presynaptic output, normalized postsynaptic input, and a binary ablation. Matrices were subsequently scaled by their maximum absolute row sum, giving a conservative contraction bound for gains below one. This engineering stabilization could suppress native recurrent effects.

The signed approximation assigned positive outgoing influence to acetylcholine and negative influence to GABA. Glutamate was varied among excluded, positive, and negative alternatives. Histamine, dopamine, serotonin, octopamine, unclear, and unknown labels remained distinct but had zero outgoing influence in the primary approximation. This was an explicit modeling exclusion, not a physiological claim that those neurons lack functional effects. Unsigned and transmitter-label-shuffled variants were also tested.

Inputs used fixed artificial populations. A structured-input ablation restricted selection to annotated sensory, visual, or optic-lobe-related populations, including optic-lobe intrinsic neurons. Neither input scheme modeled visual recognition of cards or established stimulation of biological sensory receptors.

Training and Comparison Models

Each primary graph family received the same four dynamics/gain/leak candidates and three ridge penalties, selected using validation true-count RMSE. Comparators included uniform random directed graphs matched for node and edge counts, binary-degree-preserving rewiring, equal-weight directed swaps preserving raw degrees and strengths, shuffled weights, shuffled transmitter labels, zero recurrent connectivity, and directed acyclic masks. Equal-weight swaps were constrained nulls rather than uniform samples of all compatible graphs. Preserving raw strengths did not guarantee preservation of effective signed strengths after signing and normalization.

Zeroing W removed between-neuron recurrence but retained local exponential memory when leak was below one. It was therefore distinguished from the genuinely no-memory feedforward predictor. Additional baselines were a zero predictor, a no-information constant, an engineered scalar recurrence, and the exact accumulator.

Vanilla recurrent neural networks and gated recurrent units (GRUs; Cho et al., 2014) each used 16 hidden units, Adam optimization (Kingma & Ba, 2015) with learning rate .003, and 40 epochs, retaining the best validation checkpoint. The networks had 513 and 1,505 trainable parameters, respectively. They were engineering baselines with matched neural training budgets, not exact parameter- or state-capacity matches to the reservoirs.

A limited global-gain search constituted the implemented scaling regime; it was mathematically part of primary hyperparameter selection rather than a distinct rich synaptic-learning experiment. An exploratory local-plasticity variant used bounded pre/post activity-correlation scaling without count labels, followed by engineering normalization and readout fitting. It was not spike-timing-dependent plasticity.

Secondary Normalization Analysis

Validation results motivated a separately identified matched postsynaptic-normalization comparison. Native, random, degree-preserving, and degree/raw-strength-preserving families used linear dynamics, gain .99, leak .05, identical ridge choices, and three seeds. These models used the same training, validation, and locked final-shoe sample sizes as the primary study. The analysis was frozen before final evaluation, but remained exploratory because its inclusion was motivated by development results.

Blackjack Environment and Offline Value

The deterministic simulator supported multiple deck counts, hard and soft totals, multiple aces, natural blackjacks, configurable payouts, doubling, splitting, double after split, resplitting, split-ace restrictions, late surrender, multiple players, and dealer hit/stand-on-soft-17 rules. Dealer naturals followed a U.S.-peek-style convention. The dealer always revealed and completed play, including after all players busted; this non-universal convention affects exposure sequences. A conservative reserve could also cause early reshuffling. The playing policy was not claimed to be optimal basic strategy.

Task B assessed counting in actual game exposure order over 64 independent table shoes per estimator. Dealer hole-card rank was unavailable until revelation. True-count conversion

used physical cards remaining, including the effect of already-dealt hidden cards, but did not supply their ranks. Zero-input time gaps were evaluated separately from game-event ordering.

Task C used 1,024 paired shoes and 43,382 initial rounds. A fixed policy wagered one unit, increasing to two units when estimated pre-round true count was at least 2. Playing actions were identical across estimators. Outcomes were generated once and reweighted by estimator-specific stakes, with split and double payouts scaled accordingly. Profit was measured per initial round rather than per split hand.

Outcomes and Statistical Analysis

The primary endpoint was per-shoe true-count RMSE, averaged across the three model seeds before paired shoe comparisons. Bootstrap intervals used 2,000 resamples of independent shoes, not individual card observations, consistent with resampling the independent sampling unit rather than within-shoe observations (Efron & Tibshirani, 1993). Mean pooled RMSE across model seeds was also summarized; this is a different statistic from mean per-shoe RMSE and cannot be substituted for the paired endpoint. Additional outcomes included running-count MAE, RMSE, rounded accuracy, terminal error, calibration, and correlations.

The 95% shoe-bootstrap intervals condition on the three fitted model seeds and do not fully capture new-seed or biological uncertainty. Secondary contrasts were exploratory and not multiplicity adjusted. Narrow intervals were therefore interpreted alongside effect magnitudes, seed variation, and modeling assumptions rather than as sufficient evidence of general superiority.

Follow-Up: Independent Replication and Prospective Boundaries

The follow-up protocol was committed at 88b0f86 before training. It fixed ten new computational seeds: 101, 202, 303, 404, 505, 606, 707, 808, 909, 1010. Anatomical sampling

remained the original nested seed-1729 5,000-neuron subset. The experiment therefore replicates stochastic artificial input/readout placement and null generation, not biological sampling. The three families were native MaleCNS, uniform random matched for node count, edge count, and raw weight multiset, and directed equal-weight swaps preserving raw in/out degrees and strengths. The existing signed transmitter exclusions and maximum-row-norm stabilization were unchanged. Effective signed strengths were not matched; the null must not be called a fully functionally strength-matched network.

Follow-up training used seeds 2,100,000–2,102,047; validation used 2,200,000–2,200,511; the fresh primary test used 2,300,000–2,301,023; offline table evaluation used 2,400,000–2,404,095. A separate 64-shoe smoke set began at 2,500,000. Every ordinary pure shoe comprised 234 exposures from a six-deck unbiased permutation. The final card array and its SHA-256 hash were committed before training solely to establish the prospective test lock. No fitted-model inference or outcome-based analysis occurred before checkpoint freezing. Unlike the original study's seed-only pretraining lock, the follow-up also stored the concrete final array before selection.

Fixed Replication Versus Validation-Selected Extension

The confirmatory baseline used linear dynamics, postsynaptic normalization, gain .99, leak .05, and 64 readout states. Ridge penalty was chosen using validation RMSE from 10^{-5}, .001, and .1. This reproduces the original secondary family's hyperparameters on independent shoes and computational seeds. The two primary contrasts were native minus random and native minus degree/raw-strength-null mean per-shoe true-count RMSE.

For the engineering extension, every family and seed received the same 27 validation candidates: three gain/leak pairs (.99/.05, .999/.02, .9999/.01), three readout sizes (64, 128, 256),

and three ridge penalties. The standard 64-state output selection was retained exactly, and independently chosen additional states extended it into nested 128- and 256-state sets. Input projection and readout indices were identical across corresponding graph families. This avoids replacing all sampled outputs when increasing decoder size. The baseline is nested within the extension's candidate set; consequently, validation improvement is guaranteed or tied, whereas improvement on the unseen final set is empirical. All weight matrices and input projections remained fixed. The only fitted coefficients were the 65, 129, or 257 readout parameters including the intercept. Selection minimized pooled continuous validation true-count RMSE, not the final primary endpoint.

The update matrix for linear dynamics is $(1 - a)I + agW$. Given the unit row-norm bound, the conservative per-step contraction bound is $q \leq 1 - a(1 - g)$, yielding q bounds .9995, .99998, and .999999 for the three settings. These are upper bounds rather than measured decay constants or physiological time constants. Their closeness to one motivates the longer-memory manipulation but does not establish that the actual effective graph has a near-unit eigenvalue. The study did not train recurrent weights, fit measured conductances, or simulate spikes. No comprehensive factorial final-test evaluation of all 27 candidates was conducted: only the frozen baseline and validation winner were evaluated, preventing post hoc test selection.

Streaming Estimation, Transfer Baselines, and Numerical Scope

Readout fitting accumulated sufficient statistics $X'X$ and $X'y$ over bounded batches in float64, while reservoir states and sparse matrices used float32. This bounded memory consumption without storing all training states. Direct and streamed ridge solutions were compared in small tests. Every state update checked finite values. The archived checkpoints contain the complete sparse weight matrix, input map, output indices, readout coefficients,

candidate scores, code/data provenance, and training-state extrema. The shared-subset and computational-seed matching were audited from saved models before final evaluation.

The original primary native seed-11, post-normalized native seed-11, and GRU seed-11 checkpoints were evaluated unchanged on the fresh primary shoes. They are explicitly transfer baselines: they were trained on the original data and were not retrained with follow-up sample allocation. The GRU has 1,505 trainable parameters, compared with up to 257 readout parameters for a frozen reservoir, and it jointly trains recurrent and input weights. This comparison establishes measured transfer accuracy, not exact capacity-matched architectural superiority. The scalar exact accumulator supplies the zero-error engineering reference. The new follow-up did not repeat every original no-memory, transmitter, biological-dynamics, and lesion ablation.

Seed-Generalized Statistical Analysis

For each fitted model and final shoe, the error endpoint was continuous true-count RMSE across its 234 exposures. Paired family differences were retained in a 10-by-1,024 seed-by-shoe matrix. Computational seeds and shoes were independently resampled with replacement in a crossed bootstrap, preserving paired families and shared shoe indices. Five thousand replicates generated percentile intervals. Seed-mean paired effects, between-seed standard deviations, and Student t intervals with nine degrees of freedom were also reported. These analyses incorporate stochastic model uncertainty that the original shoe-only intervals omitted. They do not establish independent biological replication, and the native graph and shared training set remain fixed.

The two confirmatory replication contrasts received 97.5% intervals, a conservative Bonferroni familywise-95% convention. Ordinary 95% intervals remain available for transparency. Claims of a robust primary advantage require compatible interval direction in both

crossed-bootstrap and seed-t summaries; near-zero or conflicting intervals are reported as unresolved. Extension contrasts are prespecified secondary engineering analyses with unadjusted 95% intervals. Pooled RMSE averaged across model seeds is reported separately from the paired mean-per-shoe endpoint. Their numerical differences are not contradictions. Figures show all ten seed outcomes rather than replacing uncertainty with a single average.

Follow-Up Table Simulation, Equity, and Exposure Accounting

The equity comparison held the original Rules, basic playing policy, and one-versus-two-unit pre-round stake rule fixed. Across 4,096 new shoes, the same card draws, actions, and unit payouts were reused for all estimators; only initial stake differed. The model seed was prospectively fixed at 101, not selected for accuracy or profit. Conditions were no-count unit stake, exact Hi-Lo, old primary native, old post-normalized native, new fixed native, new selected native, selected random, and selected raw-strength null. Counts and neuronal states reset between shoes, while chronological cumulative profit did not. Exposure streams ended at the simulator's existing penetration/reserve boundary rather than silently reshuffling mid-shoe.

The archive stores each exposed rank, physical remaining-card count, exact running count, shoe offset, round identifier, initial-round start and end indices, unit payout, pre-round position, model prediction, stake, and profit. Hole ranks are unavailable before their reveal event. Pre-round stakes use the last prior-round prediction, or zero at a new shoe, never the current round's outcome. The simulator's split replacement cards are both dealt before playing the first split child hand, and its dealer always reveals and completes play; these are disclosed simulation conventions rather than assertions of universal casino procedure. The strategy does not use count to alter playing actions. Larger stakes scale existing double/split payouts.

Five thousand paired shoe-cluster resamples estimated pooled profit per initial round, absolute intervals, and profit differences versus exact counting and no-count stakes. Each bootstrap ratio divided resampled shoe profits by resampled round counts; it did not weight every shoe equally despite varying length. Round profit variance, mean stake, and stake disagreement were also recorded. These are one-prespecified-model-seed comparisons, not seed-generalized equity estimates. Unadjusted intervals and unresolved differences must not be called proof of equivalence, optimal policy performance, or profitability.

Synchronized Anatomical Visualization

The animation uses the first two prespecified equity shoes (seeds 2,400,000 and 2,400,001), without outcome-based selection. The selected native seed-101 model is replayed sequentially from its frozen checkpoint. Each exposure's full 5,000-dimensional artificial state, exact/model counts, and remaining-card denominator are retained. Scalar replay is checked against the independent batched equity inference. The display maps states only to available canonical somaLocation coordinates; 4,261 of 5,000 modeled neurons have valid three-dimensional positions. Missing coordinates are excluded, never invented. Two-dimensional projection axes retain raw dataset orientation and are divided by 1,000 for readability without assigning unsupported anatomical axes or physical units.

Color uses a fixed symmetric-log signed-state scale (linear threshold equal to one hundredth of the fixed maximum) derived from the maximum absolute state seen during training across the candidate dynamics, not an individually rescaled test frame. Glow opacity uses a fixed monotonic mapping, proportional within each layer to clip(10|state|/training maximum, 0, 1)^0.6. This makes small activity visible without per-frame normalization. It is a visualization transform, not measured light, firing rate, spikes, calcium, or biological excitation. A separate high-

resolution three-dimensional soma image retains the same recorded locations. Neither image reconstructs neurites or a fly's body. The video shows the current exposed card, running/true counts, signed estimation error, current initial stake, and cumulative equity. Equity updates only at the final exposure of a settled round, and persists across the visible shuffle reset. All displayed equity comes from exactly the same card stream.

Results: Original Experiment

Primary Counting Performance

The exact accumulator achieved zero error. The seed-11 native model achieved running-count RMSE 4.8912, MAE 3.4259, and rounded accuracy 13.83%. Its terminal running-count MAE was 5.3313. In comparison, the zero predictor's running-count RMSE was 6.6576. The native model therefore retained predictive information without reliably implementing exact cumulative counting.

Table 1 summarizes mean pooled true-count errors across three seeds. Under the primary mapping, native topology performed worse than the uniform random graph and approximately matched the degree-preserving null. The mean paired native-minus-random per-shoe difference was 0.2745, 95% CI [0.2393, 0.3097], where positive values favored the comparator. The degree-null difference was −0.0002, 95% CI [−0.0012, 0.0009]. The raw-strength-null difference was −0.0046, 95% CI [−0.0058, −0.0034]. Although the latter interval excluded zero, its small magnitude did not establish a practically substantial advantage.

Engineering Baselines

Across seeds, GRU true-count RMSE averaged 0.0983 (SD = 0.0196), compared with 1.3987 (SD = 0.2671) for the vanilla recurrent network. Corresponding running-count RMSEs were 0.2548 (SD = 0.0589) and 3.4430 (SD = 0.6757). The seed-11 GRU achieved 97.14%

rounded running-count accuracy. Thus, the trainable GRU was substantially more accurate than the tested fixed reservoirs, although unequal parameterization prevents attributing this difference solely to architecture.

Postsynaptic Normalization

The secondary matched analysis changed the ranking (Table 1). Native mean pooled true-count RMSE was 0.8528, compared with 0.9769 for random networks and 1.1604 for the degree/raw-strength null. Native per-seed errors were 1.0584, 0.8483, and 0.6516; random errors were 0.9128, 1.0737, and 0.9440. Native topology therefore lost to random topology in one of three seeds despite its favorable average.

The paired mean-shoe native-minus-random difference was −0.0845, 95% CI [−0.1110, −0.0564]. Differences versus degree and degree/raw-strength nulls were −0.2653, 95% CI [−0.3024, −0.2300], and −0.2272, 95% CI [−0.2624, −0.1909], respectively. These are conditional contrasts within the secondary normalized family, not comparisons of secondary native models against primary controls.

Ablations, Memory, and Generalization

Removing graph recurrence increased primary mean pooled true-count RMSE to 2.2311, but substantial performance remained possible through local leaky state and a trained readout. Seed-11 signed and unsigned errors were 2.0156 and 2.0243. The exploratory local Hebbian variant achieved 2.0146, providing negligible improvement over the corresponding primary model.

Hub removal illustrated a normalization confound. With each lesioned graph renormalized, true-count RMSE was 1.7998; retaining the original normalization yielded 2.0265.

An apparent benefit from lesioning therefore could not be attributed simply to removing harmful neurons, because the intervention also changed effective global scaling.

Independent identically distributed rank probes showed weak long-lag decoding. For the primary native model, R^2 was .1026 at lag 1, .0197 at lag 16, and .0001 at lag 64. Postsynaptic normalization yielded .1793, .0386, and .0158, respectively. These probes measured recoverability of individual past Hi-Lo values, not sufficiency of the cumulative statistic: an exact accumulator need not reconstruct each earlier card.

Out-of-distribution robustness was limited. At 90% penetration, primary native true-count RMSE was 3.6358 and post-normalized native error was 2.3706. Positive-first adversarial ordering yielded errors of 21.6089 and 10.7535, respectively. Actual table exposure sequences produced errors of 2.1950 for primary native, 1.1957 for post-normalized native, and 1.0710 for post-normalized random models. These findings do not support uniformly accurate generalization across sequence conditions.

Offline Decision Value and Computational Scale

All absolute simulated profit intervals included zero (Table 2), as did all paired profit-difference intervals comparing tested estimators with exact counting. Stake choices nevertheless differed: mean stake was 1.162 for the oracle and 1.142 for the post-normalized native estimator. The experiment detected differences in decisions but did not resolve their expected-profit consequences. Failure to resolve a difference is not evidence of equivalence.

The CPU-only host had four virtual Intel Xeon Platinum 8272CL cores and approximately 16 GiB RAM. Full raw quality control reached approximately 11.01 GiB high-water resident memory, exceeding the original 6 GiB target. Sparse full-neuron pilots were completed using 64 training, 16 validation, and 32 test shoes. Native and random true-count

RMSEs were 2.5095 and 2.4787, respectively. These small pilots established executable full-graph modeling, not a powered full-scale comparison. The source delivery record documented 62 passing tests and verification of 217 frozen artifact hashes.

Results: Independent Follow-Up

Fixed-Configuration Independent Replication

Mean pooled true-count RMSE over ten seeds was 0.8085 (SD = 0.1508) for native models, 0.9990 (SD = 0.0824) for random models, and 1.1572 (SD = 0.1048) for the raw-strength null. These are new fits on fresh train/validation shoes, not rescoring of the original three normalized models. Table 3 and Figure 1 separate fixed replication from validation-selected extensions.

The native-minus-random primary paired mean per-shoe effect was -0.1578; the 97.5% crossed interval was [-0.2542, -0.0435], and the corresponding seed-t interval was [-0.2927, -0.0229]. Native had lower mean-shoe error in 9 of ten computational seeds. The native-minus-raw-strength-null effect was -0.2922, crossed 97.5% interval [-0.3922, -0.1709], seed-t 97.5% interval [-0.4311, -0.1534]; native won in 9 of ten seeds. Under the prospectively stated conservative criterion, the native–random analysis favored native topology. Table 4 and Figures 2–3 retain uncertainty rather than extrapolating a favorable average to all input placements.

Longer-Memory Dynamics and Larger Readouts

The validation-selected native models achieved mean pooled true-count RMSE 0.0097 (SD = 0.0019), running-count RMSE 0.0186, running-count MAE 0.0120, rounded integer accuracy 100.00%, and final-position MAE 0.0506. The matched selected random and raw-strength-null families achieved true-count RMSEs 0.0150 and 0.0190. All three selected graph families achieved 100% nearest-integer running-count accuracy on the fresh pure-shoe

observations in every computational seed. Thus, the remaining native advantage concerns small continuous errors, not uniquely successful rounded counting. This finite-sample result does not prove exact unrounded or indefinite accumulation. No model was chosen by these final scores. Selected native readouts had mean coefficient L2 norm 8665.4, with maximum absolute coefficient 3210.3 across seeds. These are substantial engineering gains applied to small artificial states; they are not calibrated biological synaptic conductances. Noise injection and perturbation robustness were not evaluated, so low deterministic error must not be equated with physiological reliability.

For native selected-minus-baseline, the paired per-shoe effect was -0.7011, crossed 95% interval [-0.7813, -0.6381], and seed-t interval [-0.7882, -0.6141]. The selected native-minus-random effect was -0.0046, crossed 95% interval [-0.0055, -0.0036]; selected native-minus-strength effect was -0.0079, crossed interval [-0.0094, -0.0063]. These secondary engineering contrasts are not multiplicity-adjusted discoveries. All seed differences are archived. An extremely small numerical advantage, even if precise, must be evaluated relative to an exact zero-error accumulator.

Across all 30 family-by-seed searches, selected output sizes were {'64': 0, '128': 0, '256': 30}; selected gain/leak frequencies were {'0.99/0.05': 0, '0.999/0.02': 0, '0.9999/0.01': 30}. Figure 4 shows the complete validation capacity/timescale summaries, with no final-set retuning. Increasing readout size changes both observable state coverage and trainable capacity, so the experiment does not separately identify which produces the gain. The final selected comparison combines longer memory and output sampling; it is not an isolated causal estimate for each factor. Figure 5 shows position-dependent held-out error. Strong performance over 234 legal-

shoe exposures does not prove indefinite exact accumulation, robustness to adversarial streams, or transfer to arbitrary timescales.

Transfer and Exact Baselines

The original frozen GRU seed-11 achieved fresh-shoe true-count RMSE 0.0928, running-count RMSE 0.2233, and rounded integer accuracy 97.16%. The original primary native and post-normalized native transfer errors were 2.0299 and 1.1073, respectively. This separates new fits from unchanged historical checkpoints. The follow-up exact accumulator had running- and true-count RMSE zero. It needs one scalar state and no learned coefficients; high reservoir accuracy does not make it the preferable engineering solution for counting.

Fresh Table Exposure and Paired Equity

The fresh game simulation produced 173,440 initial rounds and 968,238 visible card events across 4,096 independent shoes. Selected native seed-101 table running-count RMSE was 0.0180; true-count RMSE was 0.0097. Its mean stake was 1.1463 units, versus 1.1467 for exact counting, and stake disagreement occurred on 0.05% of initial rounds. Accurate near-threshold estimates can still change discrete stake choices. In particular, a small negative error when the exact true count equals the inclusive threshold of 2 can prevent the two-unit stake. The unchanged policy uses continuous estimates, not rounded running counts; it was not altered after observing this effect.

Selected native pooled profit was -0.00382 units per initial round, absolute 95% interval [-0.0102, 0.0027]. Its paired difference versus exact was -0.00007, 95% interval [-0.0002, 0.0001]. No estimator-versus-exact paired interval excluded zero. Table 5 and Figures 6–7 give every condition, including unfavorable trajectories. Chronological cumulative curves are

descriptive realized paths, not expected growth forecasts. They begin at zero cumulative profit and include all predetermined shoes without stopping at a favorable outcome.

Replay Fidelity, Compute, and Software Evidence

The video contains 470 exposed-card frames from the two prospectively selected shoes. Its maximum scalar-versus-batched prediction discrepancy was 4.97e-14, below the declared 10^{-4} tolerance. It displays real coordinate locations for 4,261 neurons and artificial-state illumination on a fixed training-derived scale. Figure 8 exposes the corresponding exact/model trace and signed error. The video ends with a separately labeled eight-second full-study summary: the two prespecified illustration shoes happened to finish +16 native units, whereas all 4,096 shoes finished -662 native units. The positive clip was retained because its seeds were fixed before outcomes, not chosen for profit; the complete chronological curve and uncertainty are shown in the closing panel. Full decoding and sampled-frame inspection are documented in the media validation record; these validate encoding and correspondence, not biological fidelity.

The 30 family searches consumed 37.25 summed worker-minutes, including candidate features, readout fitting, and validation, under two concurrent single-thread numerical workers. Largest training-process high-water RSS was 0.838 GiB; this is process-level high-water memory, not isolated incremental model allocation. The earlier smoke benchmark took 0.588 seconds for 64 shoes and a 256-state readout. Actual completed timings, not benchmark extrapolation, are the authoritative compute evidence (Figure 9). An initial equity run was interrupted when repeated NPZ decompression inside card loops caused avoidable I/O overhead. A separate wrapper cached the same immutable arrays; fixture predictions were bitwise identical, and frozen inference arithmetic, models, rules, and test sets were unchanged. The interruption and equivalence evidence are retained in the research log; the summed training time above does

not include that interrupted equity I/O time. No GPU, energy measurement, purchase, or public deployment was used. Original and follow-up tests and original/new frozen hash verification are recorded in the delivery audit. The archive preserves the original paper/video unchanged and places new results in followup/v2.

Discussion

The independent study changes the evidence base by distinguishing replication of a fixed normalized reservoir from improvements enabled by a bounded engineering search. The original work established that topology rankings could reverse with normalization and that three-seed shoe-only intervals understated model uncertainty. The follow-up adds ten fresh computational seeds, disjoint training and validation shoes, concrete pretraining test-array locking, and uncertainty over both model seeds and shoes. Its primary native–random result favored native topology. This is a more defensible statement than treating the original favorable normalized mean as a general property of fly anatomy.

The larger-readout/longer-memory extension demonstrates what the original 64-state decoder could not establish. Selected native mean running-count RMSE was 0.0186, compared with 1.7095 for newly replicated fixed native models. This is measured held-out improvement, not just lower validation error. Yet it remains a result of the entire artificial system: fixed anatomical graph, excluded transmitter classes, chosen normalization, rank stimulation, leak/gain settings, sampled decoder, ridge fitting, and the finite legal-shoe distribution. The matched controls also benefit from the same engineering choices. A favorable native contrast cannot be credited solely to biological circuitry without addressing residual effective signed-strength and timescale differences.

The model's computational meaning must remain modest. Hi-Lo is an exact scalar algorithm. Even very low RMSE or high rounded accuracy does not demonstrate intelligence, numerical concepts, consciousness, blackjack understanding, or biological fly competence. No actual fly saw a card or learned in this experiment. The count network received abstract rank encodings, while a separate deterministic program implemented legal cards, game rules, payouts, and the exact target. Improved performance relative to a transferred GRU would not establish a universal architectural advantage: the GRU was an unchanged original checkpoint, not a newly tuned, parameter-matched competitor. Conversely, a GRU advantage would not establish incapacity of all connectome-derived dynamics.

Scientific and Engineering Value

The main value of the study is not a claim that a fly connectome solves blackjack. It is a controlled test of how anatomical topology, artificial dynamics, and decoder capacity interact under prospective evaluation. On the fresh ten-seed replication, fixed native models had 19.1% lower pooled true-count RMSE than matched random models (0.8085 vs. 0.9990). Within the bounded engineering extension, the selected native configuration reduced pooled RMSE by 98.8% relative to the fixed native baseline (0.0097 vs. 0.8085). The selected native family also had 35.3% lower pooled RMSE than selected random networks and 48.9% lower pooled RMSE than the selected degree/raw-strength null. These percentage comparisons are derived directly from the reported held-out RMSEs and should be read alongside the much smaller absolute selected-family differences and the exact accumulator's zero error.

The more general contribution is methodological separation. Native topology produced a reproducible advantage in the fixed normalized regime, but the dominant performance gain came from longer-memory dynamics and a larger readout applied to every graph family. That

distinction prevents a topology effect from being confused with a hyperparameter effect. Prospective test locking, matched null families, crossed seed-by-shoe uncertainty, artifact hashes, and synchronized replay further make the result auditable. For connectome-derived machine learning, this is more informative than a single favorable benchmark score because it identifies which conclusions survive changes in stochastic initialization and which depend on engineering choices.

Causal and Statistical Limitations

The ten computational seeds improve inference over input/output/null randomness but do not add biological specimens. All native models share one induced subset; induced selection omits outside connections and changes recurrence. The raw-strength control preserves anatomical weighted sums before signing, not the functional signed matrix after exclusions and normalization. Equal-weight swaps have constrained mixing and do not uniformly sample every admissible graph. Replication across many seeds therefore does not remove these structural confounds.

Readout enlargement simultaneously changes the number of observed states and fitted coefficients. The present selection design holds the candidate budget equal across graph families and uses nested output sets, but it cannot distinguish every explanatory component of an improved decoder. The conservative contraction bounds are not empirically matched timescale spectra. Low finite-shoe error is not proof of an exact internal accumulator or long-horizon numerical stability. Stronger tests would include perturbation sensitivity, readout conditioning, wider data-generating regimes, independently sampled anatomical subsets, and matched effective signed dynamics. These should use new prospective data rather than selecting on this now-observed final set.

Bootstrap shoe resampling and computational-seed resampling answer different sampling questions. The crossed procedure respects both, but ten seed effects still provide limited information about distribution tails. The conservative primary intervals address the two named fixed-replication contrasts; they do not make every secondary comparison confirmatory. Single-seed equity estimates are especially conditional. The table environment's non-universal reveal/split/reserve conventions and fixed nonoptimal playing policy further constrain applicability. No effect, whether positive or unresolved, establishes casino profitability.

Visualization and Reproducibility

The revised animation is not a biological simulation movie in the physiological sense. It is a spatial display of actual frozen artificial model states at real recorded soma positions. Missing coordinates remain absent; glow is a declared graphic transform. It does not present a stylized fly body or arbitrary neuron cloud as reconstructed anatomy. Showing exact counts alongside estimates makes error visible rather than hiding it with smooth activity. Tying profit updates to actual settlement events prevents future payouts from appearing early. A reset changes counts and neuronal state, not accumulated equity, and the first two predetermined shoes were retained regardless of their result.

The original study's wider biological ambitions remain incomplete: no spiking/conductance models, STDP, morphology-based delays, receptor-specific sign inference, full-scale multi-seed training, or anatomical localization experiment was added. This follow-up addresses its specifically authorized replication, memory/readout, equity, paper, and visualization objectives rather than claiming to exhaust that larger program. Hash freezes, recorded failure states, raw per-shoe/per-round outputs, and versioned deliverables make the actual evidence inspectable.

Highest-Value Next Experiment

The most informative next experiment is a new prospective multi-subgraph replication with controls matched for effective signed in/out strength and measured dynamical decay, while independently crossing readout size and leak/gain. It should evaluate long-horizon numerical and perturbation robustness and a freshly trained capacity-matched neural baseline. This would test whether any remaining native advantage survives the confounds still present here, rather than extending a favorable finite-shoe result to biological computation in general.

Conclusion

The original primary MaleCNS reservoir lacked a broad topology advantage; its favorable secondary normalization result required independent replication. The present ten-seed replication favored native topology, with fixed native pooled true-count RMSE 19.1% below the matched random family. Validation-selected longer-memory dynamics and larger readouts reduced native mean true-count RMSE to 0.0097 on 1,024 untouched shoes, compared with 0.0150 for selected random networks and 0.0190 for the selected raw-strength null. The exact accumulator remained perfect, so the result is not an argument for replacing a known scalar algorithm with a 5,000-node reservoir. Its value is instead as a reproducible systems experiment showing that native topology can matter under a fixed regime while dynamical timescale and readout capacity can dominate absolute error. A new 4,096-shoe equity study measured decisions and profits under fixed rules, and the synchronized visualization ties artificial-state replay to the same exposed-card and settlement stream. These findings describe constrained artificial computation on an anatomical substrate, not a fly that understands or plays blackjack and not evidence of casino profitability.

Data and Code Availability

The local repository preserves original artifacts and a separately versioned followup/v2 study. The latter includes PREREGISTRATION.md, configs/protocol.json, a pretraining locked card array and hash, FROZEN.json, complete selected and baseline checkpoints, all final predictions, per-shoe losses, every equity exposure and settled-round payout/stake, chart-source CSVs, replay states and soma coordinates, and reproduction commands. Chart PNG/PDFs, the full APA-style manuscript in Markdown/DOCX/PDF, video MP4, high-resolution frames, and validation manifests accompany the source bundle. Original raw canonical data remain governed by the documented CC BY 4.0 provenance. This manuscript version is authored by Ogdn Ames. No journal acceptance or independent peer review of this manuscript is claimed, and no public repository accession is asserted beyond the cited external MaleCNS resources. Existing final sets must not be reused for new selection.

References

Berg, S., Beckett, I. R., Costa, M., Schlegel, P., Januszewski, M., Marin, E. C., Nern, A., Preibisch, S., Qiu, W., Takemura, S., Fragniere, A. M. C., Champion, A. S., Adjavon, D.-Y., Cook, M., Gkantia, M., Hayworth, K. J., Huang, G. B., Kampf, F., Katz, W. T., ... Jefferis, G. S. X. E. (2026). Sexual dimorphism in the complete Drosophila male central nervous system connectome. *Cell, 189*(18), 5504-5526.e15. https://doi.org/10.1016/j.cell.2026.08.015

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724-1734). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1179

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall.

HHMI Janelia FlyEM. (2026). *MaleCNS (Version 1.0)* [Data set]. https://janelia-flyem.github.io/male-cns/download/

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55-67. https://doi.org/10.1080/00401706.1970.10488634

Jaeger, H. (2001). *The "echo state" approach to analysing and training recurrent neural networks* (GMD Report No. 148). GMD - German National Research Center for Information Technology. https://doi.org/10.24406/publica-fhg-291111

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1412.6980

Lin, A., Yang, R., Dorkenwald, S., Matsliah, A., Sterling, A. R., Schlegel, P., Yu, S.-C., McKellar, C. E., Costa, M., Eichler, K., Bates, A. S., Eckstein, N., Funke, J., Jefferis, G. S. X. E., & Murthy, M. (2024). Network statistics of the whole-brain connectome of Drosophila. *Nature, 634*(8032), 153-165. https://doi.org/10.1038/s41586-024-07968-y

Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation, 14*(11), 2531-2560. https://doi.org/10.1162/089976602760407955

neuPrint Development Team. (n.d.). *neuprint-python documentation* [Software documentation]. https://connectome-neuprint.github.io/neuprint-python/docs/index.html

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America, 115*(11), 2600-2606. https://doi.org/10.1073/pnas.1708274114

Pospisil, D. A., Aragon, M. J., Dorkenwald, S., Matsliah, A., Sterling, A. R., Schlegel, P., Yu, S.-C., McKellar, C. E., Costa, M., Eichler, K., Jefferis, G. S. X. E., Murthy, M., & Pillow, J. W. (2024). The fly connectome reveals a path to the effectome. *Nature, 634*(8032), 201-209. https://doi.org/10.1038/s41586-024-07982-0

Thorp, E. O. (1966). *Beat the dealer: A winning strategy for the game of twenty-one* (Rev. ed.). Random House.

Table 1

Mean Pooled True-Count RMSE Across Three Stochastic Seeds

Model	Primary mapping	Postsynaptic normalization
MaleCNS	2.0154	0.8528
Uniform random	1.7188	0.9769
Degree-preserving null	2.0157	1.1929
Degree/raw-strength-preserving null	2.0200	1.1604
Vanilla RNN	1.3987	—
GRU	0.0983	—
Exact accumulator	0.0000	—

Note. Lower values indicate better performance. Values are averages of seed-specific pooled RMSEs, not the mean per-shoe RMSE used for paired inference. Neural and exact baselines are listed once; the normalization manipulation does not apply to them. The postsynaptic family is a validation-motivated secondary analysis. RMSE = root-mean-square error; RNN = recurrent neural network; GRU = gated recurrent unit.

Table 2

Offline Simulated Profit per Initial Round

Estimator	Pooled profit	95% shoe-bootstrap CI
No-count unit stake	0.0008	[−0.0092, 0.0111]
Exact Hi-Lo	0.0027	[−0.0096, 0.0150]
Primary MaleCNS, seed 11	0.0030	[−0.0085, 0.0141]
Primary random, seed 11	0.0010	[−0.0110, 0.0128]
Post-normalized MaleCNS, seed 11	0.0027	[−0.0099, 0.0147]
Post-normalized random, seed 11	0.0021	[−0.0107, 0.0144]

Note. Results cover 1,024 paired shoes and 43,382 initial rounds under a fixed one- or two-unit stake policy. Values are simulated stake units, not real-currency returns. All paired estimator-versus-exact profit intervals also included zero. CI = confidence interval. These results neither demonstrate equivalence nor establish real-world profitability.

Table 3

Independent Follow-Up Counting Performance Across Ten Computational Seeds

Family	Fixed TC RMSE, mean (SD)	Selected TC RMSE, mean (SD)	Selected RC RMSE
MaleCNS	0.8085 (0.1508)	0.0097 (0.0019)	0.0186
Uniform random	0.9990 (0.0824)	0.0150 (0.0008)	0.0283
Degree/raw-strength null	1.1572 (0.1048)	0.0190 (0.0015)	0.0352

Note. Pooled RMSEs are calculated per model and summarized across ten seeds. RC = running count; TC = continuous true count. Fixed replication and selected extension are different estimands. All use the same 1,024 fresh test shoes.

Table 4

Paired Follow-Up Contrasts With Seed-Generalized Uncertainty

Contrast	Mean shoe-TC-RMSE difference	Crossed interval	Seed-t interval
Fixed native − random	-0.1578	[-0.2542, -0.0435]	[-0.2927, -0.0229]
Fixed native − raw-strength null	-0.2922	[-0.3922, -0.1709]	[-0.4311, -0.1534]
Selected native − random	-0.0046	[-0.0055, -0.0036]	[-0.0057, -0.0035]
Selected native − raw-strength null	-0.0079	[-0.0094, -0.0063]	[-0.0097, -0.0060]
Native selected − fixed	-0.7011	[-0.7813, -0.6381]	[-0.7882, -0.6141]

Note. First two rows use 97.5% intervals for the two primary contrasts; remaining rows use unadjusted exploratory 95% intervals. Negative favors the first-named model. Crossed intervals independently resample computational seeds and shared shoes; seed-t intervals use ten seed means, df = 9.

Table 5

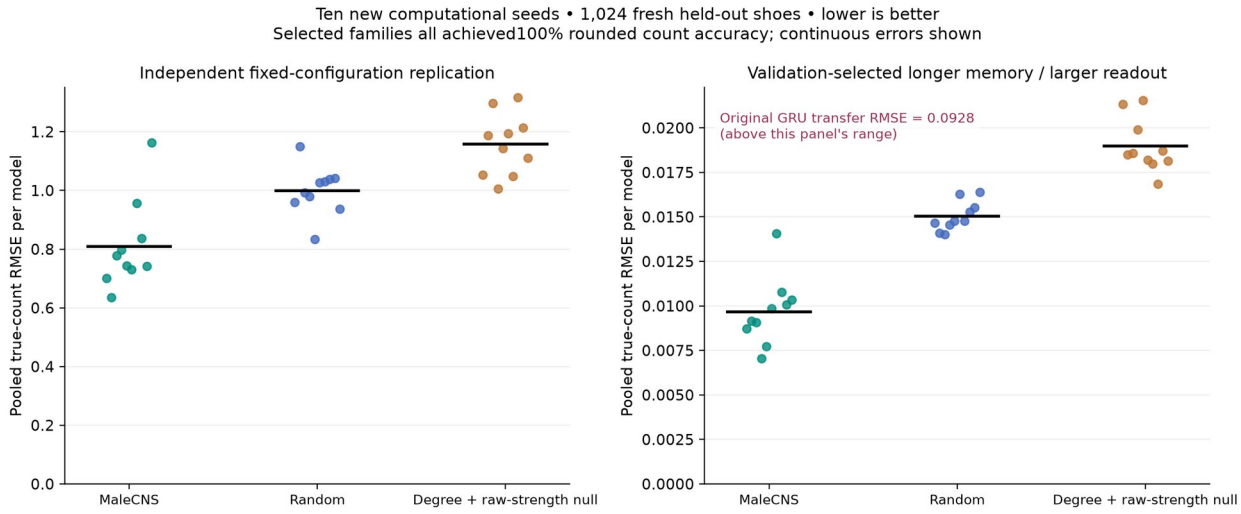
Fresh Paired Offline Equity Evaluation

Estimator	Profit / initial round	Absolute 95% CI	Paired vs exact 95% CI
no count	-0.00532	[-0.0107, 0.0001]	[-0.0036, 0.0004]
exact	-0.00375	[-0.0101, 0.0028]	[0.0000, 0.0000]
old primary	-0.00462	[-0.0106, 0.0013]	[-0.0028, 0.0010]
old post	-0.00404	[-0.0104, 0.0023]	[-0.0016, 0.0010]
new baseline	-0.00319	[-0.0095, 0.0032]	[-0.0005, 0.0016]
new selected	-0.00382	[-0.0102, 0.0027]	[-0.0002, 0.0001]
random selected	-0.00377	[-0.0102, 0.0027]	[-0.0002, 0.0001]
strength selected	-0.00382	[-0.0103, 0.0027]	[-0.0002, 0.0001]

Note. 4,096 independent shoe clusters, 173,440 initial rounds; values in simulated stake units. One prespecified model seed101, original seed11 transfer models. All playing/stake rules held fixed. Intervals are not proof of casino profitability or equivalence.

Figure 1

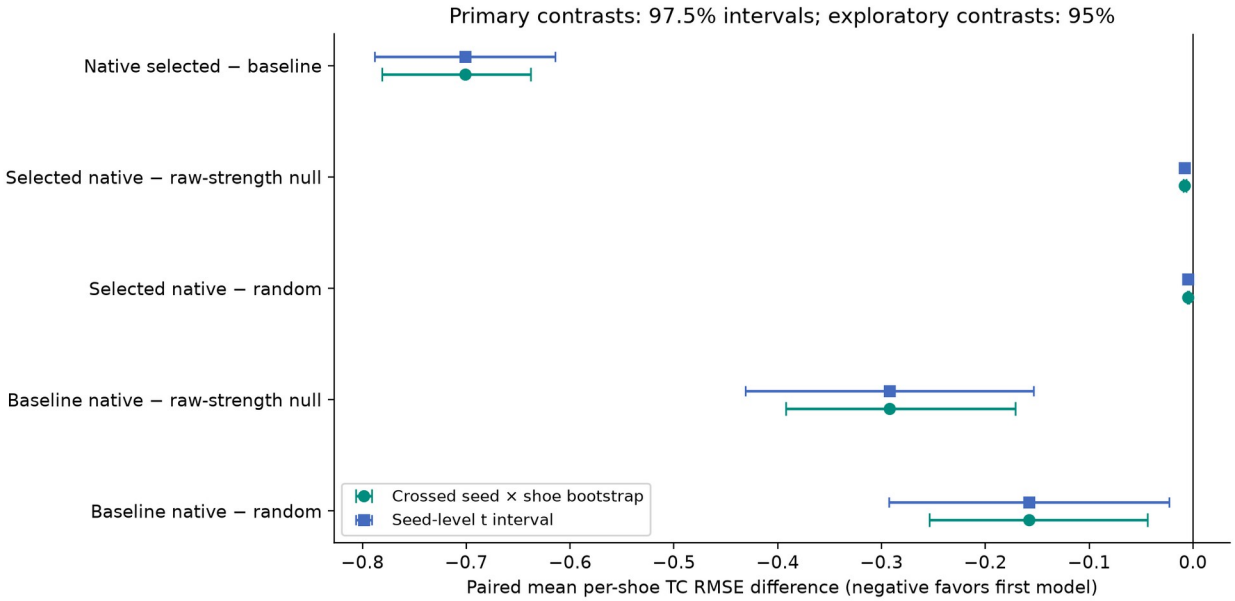
Fresh-seed counting performance



Note. Source numerical data and raw artifact paths are archived in charts/MANIFEST.json. No figure was manually fitted to a preferred outcome.

Figure 2

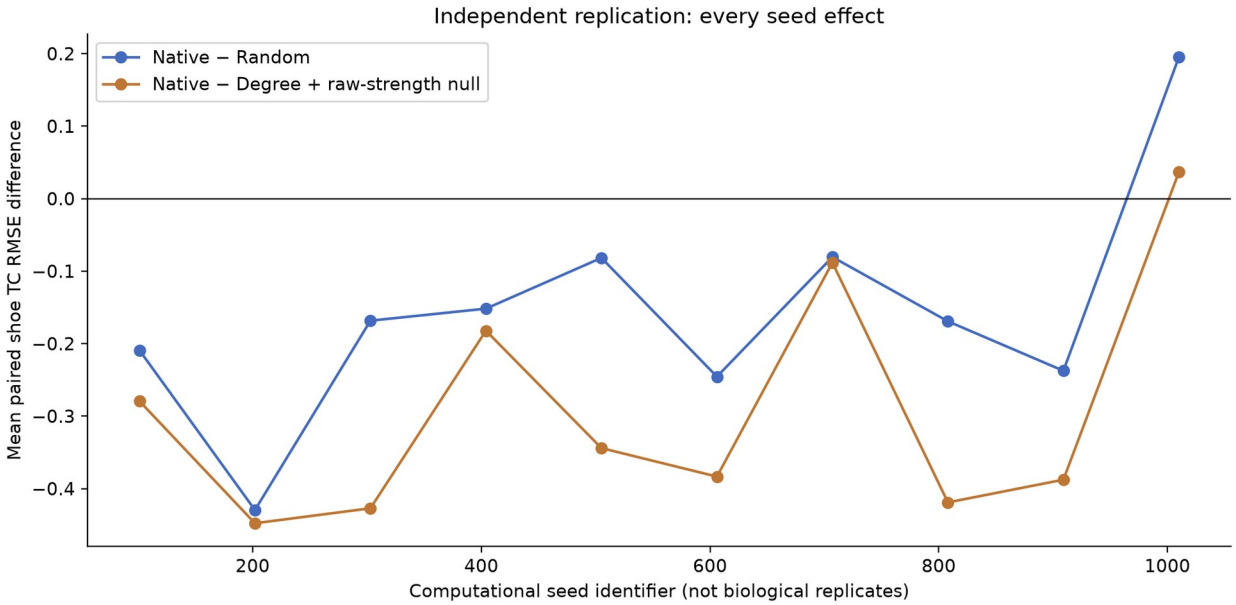
Paired uncertainty incorporating computational seeds



Note. Source numerical data and raw artifact paths are archived in charts/MANIFEST.json. No figure was manually fitted to a preferred outcome.

Figure 3

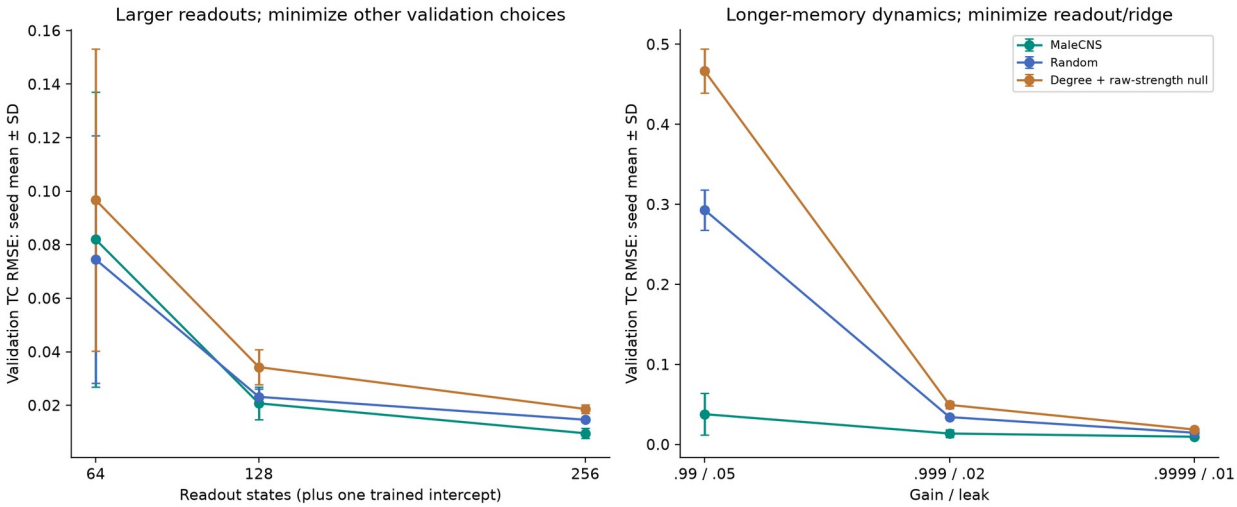
Every fixed-replication seed effect



Note. Source numerical data and raw artifact paths are archived in charts/MANIFEST.json. No figure was manually fitted to a preferred outcome.

Figure 4

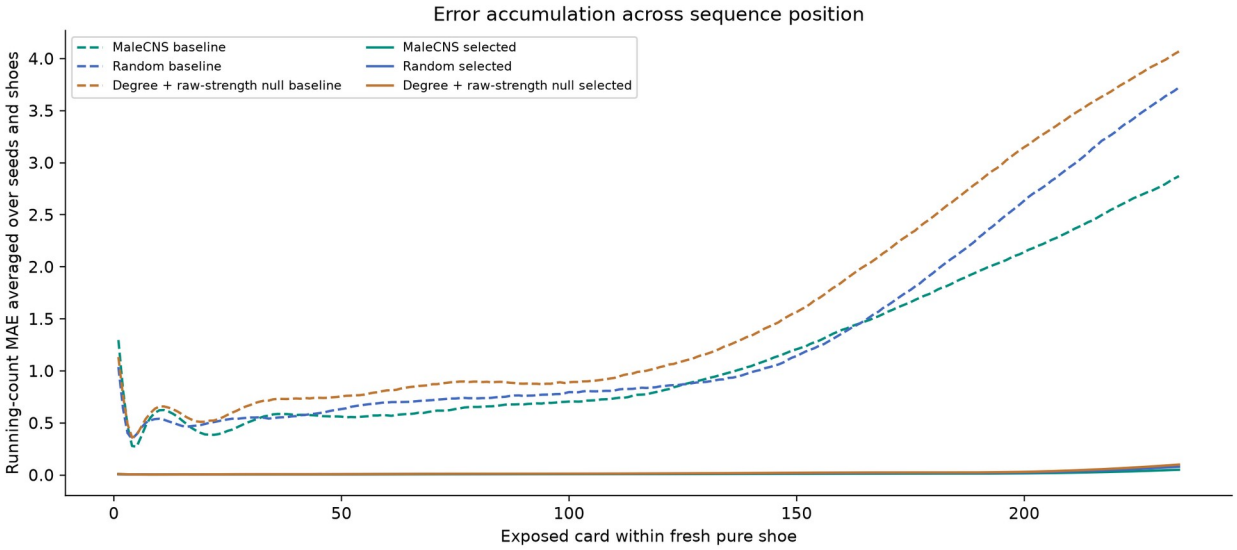
Validation-only capacity and timescale comparisons



Note. Source numerical data and raw artifact paths are archived in charts/MANIFEST.json. No figure was manually fitted to a preferred outcome.

Figure 5

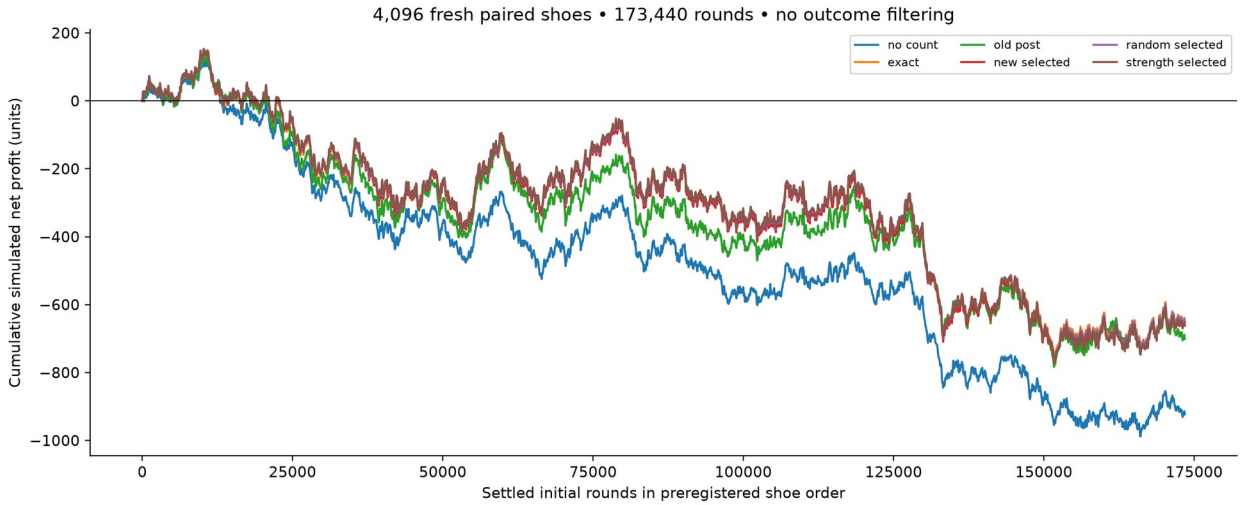
Held-out count error by shoe position



Note. Source numerical data and raw artifact paths are archived in charts/MANIFEST.json. No figure was manually fitted to a preferred outcome.

Figure 6

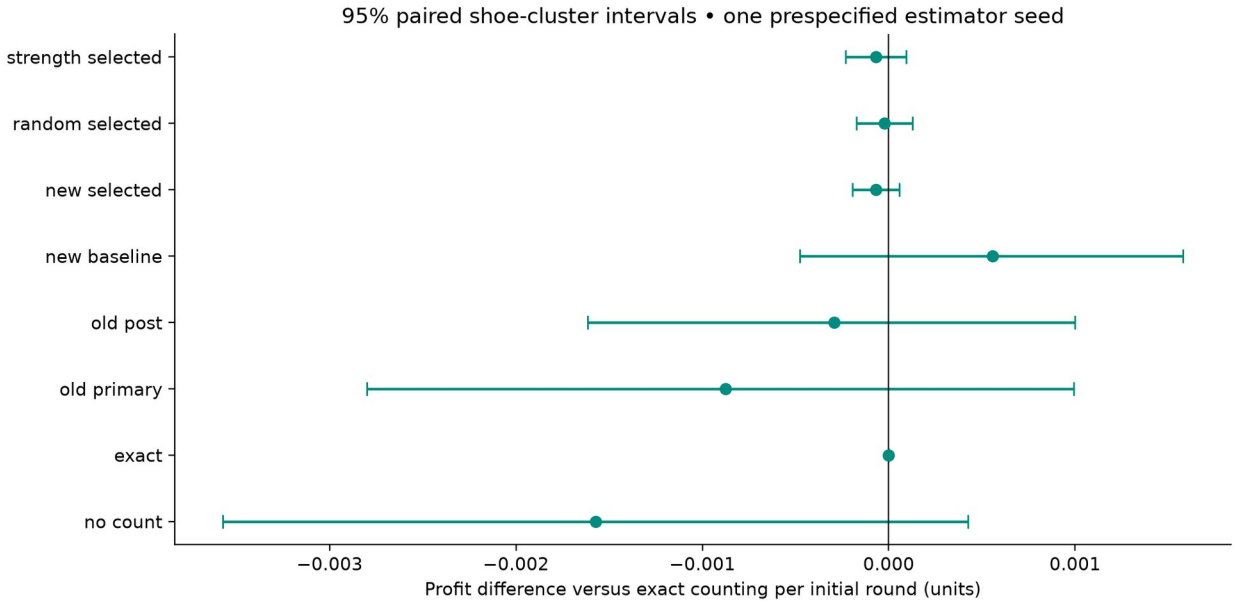
Chronological cumulative simulated profit



Note. Source numerical data and raw artifact paths are archived in charts/MANIFEST.json. No figure was manually fitted to a preferred outcome.

Figure 7

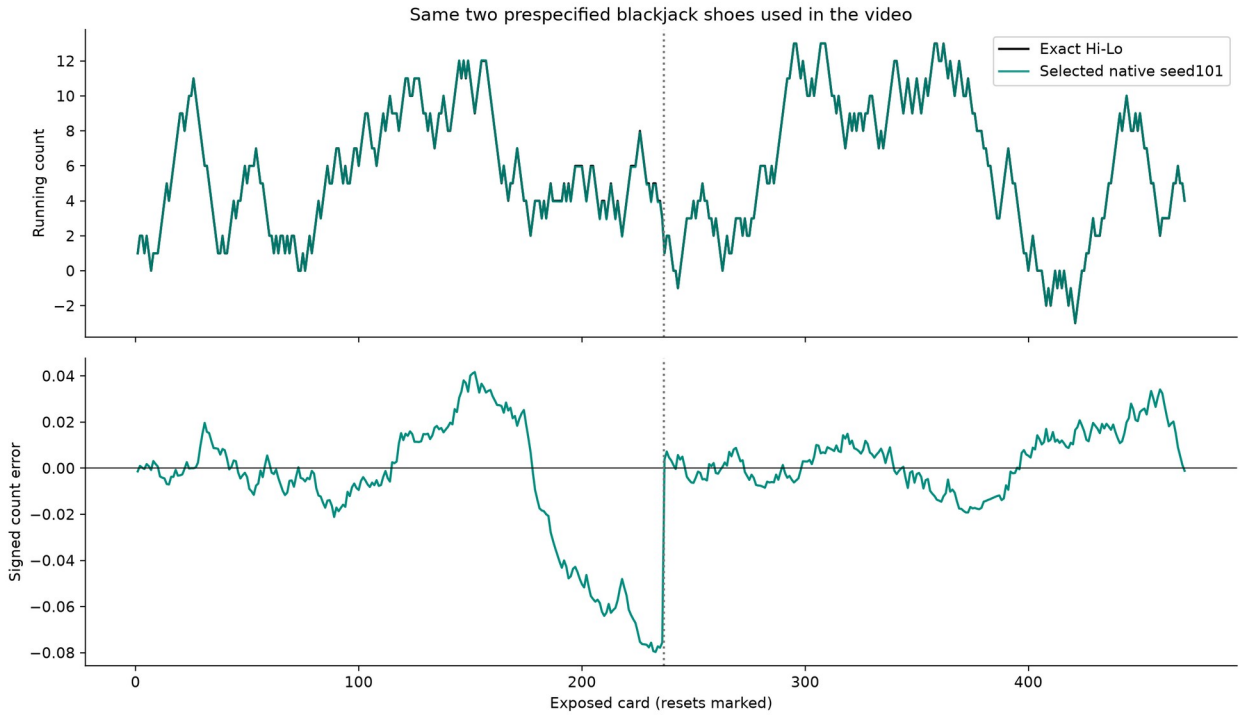
Paired profit uncertainty versus exact counting



Note. Source numerical data and raw artifact paths are archived in charts/MANIFEST.json. No figure was manually fitted to a preferred outcome.

Figure 8

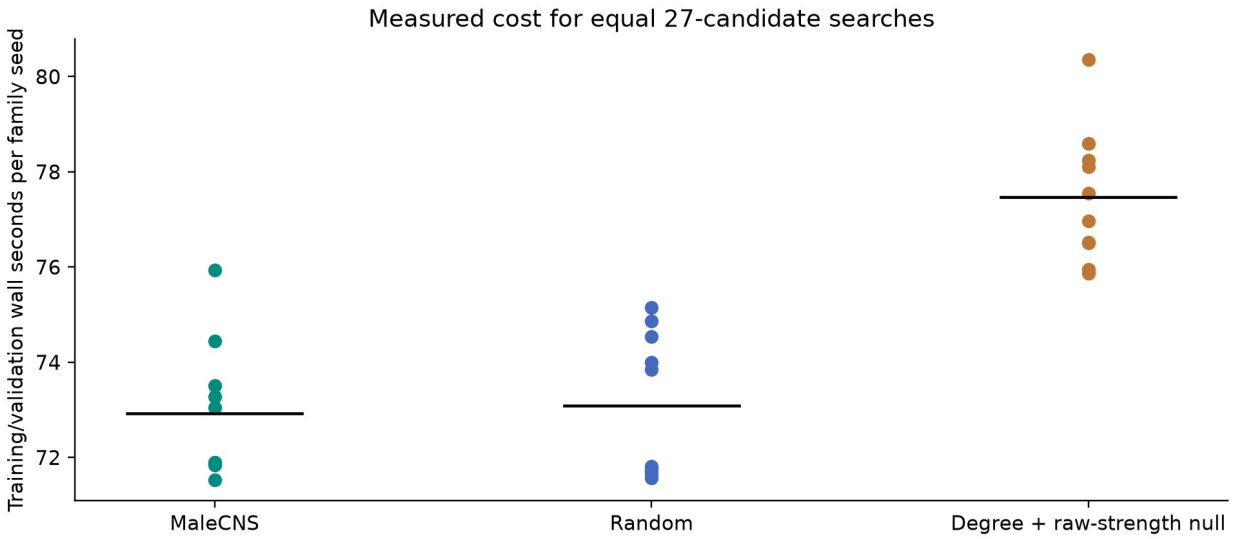
The actual video stream: exact and predicted counts



Note. Source numerical data and raw artifact paths are archived in charts/MANIFEST.json. No figure was manually fitted to a preferred outcome.

Figure 9

Measured training/validation compute cost



Note. Source numerical data and raw artifact paths are archived in charts/MANIFEST.json. No figure was manually fitted to a preferred outcome.