Fly-Derived Reservoir Computing for Momentum Trading: Retrospective Evidence, Engineering Strengths, and Empirical Failures

Ogdn Ames

Ames Investment Systems

September 13, 2026

Author Note. This paper reformats and expands the supplied retrospective research report into an APA 7-style research document. The trading evidence remains simulated and retrospective; no live orders, broker fills, or genuinely prospective market outcomes are claimed.

Abstract

This paper evaluates a deliberately unusual quantitative-finance hypothesis: whether a fixed recurrent reservoir derived from the synaptic topology of the male Drosophila central nervous system can improve momentum-based allocation decisions. The underlying study executes 2,520 scenario/model runs, 7,560 fold-level metric rows, and 240 exploratory paired comparisons across four daily assets—SPY, GLD, QQQ, and AAPL—while 20 of the originally requested 24 asset-frequency cases remain unavailable because authorized source data were not present. The principal finding is negative but informative. In the common period, the fly-derived family trails buy-and-hold on the assigned mean fold objective for SPY, GLD, and QQQ; AAPL shows a small mean-of-ratios advantage that reverses under an equal-capital ensemble comparison. No multiplicity-adjusted comparison establishes superiority, and calibrated strength-preserving graph nulls are generally similar to or slightly better than the anatomical topology. The study nevertheless demonstrates several methodological strengths: causal feature construction, expanding-prefix selection, cost-aware retraining, multiple topology and conventional controls, deterministic accounting reconciliation, explicit data-provenance manifests, and a frozen prospective protocol. Its most important failures are equally clear: no untouched prospective outcome, incomplete market coverage, reduced biological fidelity, simplified execution assumptions, sensitivity to aggregation and cost assumptions, and persistent exposure to data-snooping risk. The correct conclusion is therefore not that connectome-inspired computation is useless, but that this particular retrospective experiment does not establish a trading advantage attributable to fly anatomy.

Keywords: reservoir computing, Drosophila connectome, momentum, quantitative trading, backtesting, data snooping, Sharpe ratio, Sortino ratio

Fly-Derived Reservoir Computing for Momentum Trading: Retrospective Evidence, Engineering Strengths, and Empirical Failures

Quantitative finance regularly borrows ideas from physics, statistics, information theory, and machine learning. This study pushes that interdisciplinary impulse further by asking whether a recurrent network whose topology is derived from a biological connectome can serve as a useful computational reservoir for market timing. The hypothesis is interesting precisely because it is easy to overstate. A fruit-fly connectome is a map of biological connectivity, not a financial model, and a backtest is an empirical screen, not a proof of causal economic value. The proper standard is therefore comparative: the fly-derived topology should improve the assigned objective after modeled costs, outperform conventional and topology-matched controls, and eventually survive genuinely unseen data.

The supplied retrospective report reaches a disciplined conclusion: the engineering pipeline works, but the market evidence does not demonstrate that fly-derived wiring improves the assigned objectives or outperforms appropriate controls (OpenClaw, 2026a). That distinction—engineering success versus scientific or economic confirmation—is the central theme of this paper. The negative result is not a defect to be hidden. It is the most useful result in the package because it identifies exactly where the idea remains plausible, where it fails, and what a credible next test must do.

Research Question and Evidentiary Standard

The original research scope covered six assets (SPY, GLD, QQQ, AAPL, RSP, and SOXL) at four frequencies (15-minute, hourly, four-hour, and daily), for 24 requested asset-frequency cases. Objectives were asset specific: Sharpe ratio for SPY and GLD, Sortino ratio for QQQ and AAPL, and profit for RSP and SOXL. Only four daily cases were executable from the

authorized cached data, leaving 20 of 24 cases source-blocked (OpenClaw, 2026a). Missing evidence must not be recoded as failure or success. It simply limits what can be concluded.

A meaningful success claim requires three layers of evidence. First, the strategy should improve the assigned objective after realistic costs and financing assumptions. Second, that improvement should be attributable to the fly-derived topology rather than to generic momentum, recurrent memory, risk reduction, or decoder flexibility. Third, the advantage should persist in genuinely unseen outcomes. The present package addresses the first two questions retrospectively and prepares, but does not execute, the third (OpenClaw, 2026a).

Why the Hypothesis Is Scientifically Interesting

From Connectome to Computational Reservoir

Reservoir computing separates a recurrent dynamical system—the reservoir—from a simpler readout or decision layer. The reservoir transforms current and past inputs into a high-dimensional state, allowing temporal information to persist without fully training every recurrent weight. Lukoševičius and Jaeger (2009) describe this as a practical way to exploit recurrent nonlinear dynamics while keeping training relatively simple. That architecture makes a connectome-derived graph conceptually relevant: a fixed biological topology can be treated as an alternative reservoir rather than as a claim that the financial model literally reproduces an animal brain.

The biological source is the male Drosophila central nervous system connectome reported by Berg et al. (2026). The full resource contains 166,691 neurons spanning the brain and nerve cord. The financial implementation, however, samples 5,000 nodes for each of two anatomy seeds—about 3% of the available neurons—and applies artificial leak, tanh, sign, scaling, and readout rules (Berg et al., 2026; OpenClaw, 2026b). Acetylcholine is mapped to

positive weights, GABA to negative weights, and other transmitter classes are made computationally silent. These are engineering choices. They are not a physiological simulation and should not be marketed as one.

This difference is crucial. A connectome gives structural information about neurons and synaptic relationships, but a trading reservoir also requires a state-update law, input encoding, scaling, leak dynamics, and a readout mechanism. The original anatomy does not specify that market returns should enter as volatility-normalized momentum features, that synaptic weights should be log-transformed, or that a tanh gate should control exposure. The experiment therefore tests a computational artifact inspired by anatomy—not "the fly brain trading stocks." That framing makes the work more defensible and, frankly, more interesting.

Momentum as the Financial Signal

The financial side of the experiment is built around momentum and trend persistence rather than arbitrary neural-network inputs. Momentum has a substantial empirical literature. Jegadeesh and Titman (1993) documented cross-sectional continuation in stock returns, while Moskowitz, Ooi, and Pedersen (2012) reported time-series momentum across liquid equity-index, currency, commodity, and bond futures. These findings do not validate any particular fly-derived model, but they provide an economically recognizable signal family against which the reservoir can be tested.

Here, momentum is computed over 5-, 20-, and 60-bar horizons after a 61-bar warmup and scaled by recent volatility. Inputs are clipped to control extremes. The reservoir transforms those causal price-only features, and a finite policy bank converts the resulting gate into market exposure. The design is therefore better understood as a nonlinear momentum-and-risk allocation

system whose recurrent state is partly shaped by a connectome-derived graph (OpenClaw, 2026a).

Data, Scope, and Provenance

The market source field is Bloomberg TOT_RETURN_INDEX_GROSS_DVDS from licensed local caches. SPY, QQQ, and AAPL cover September 1, 2023 through September 11, 2026, while GLD extends back to January 4, 2010. The common-date intersection used for aligned comparisons is September 1, 2023 through September 10, 2026 (OpenClaw, 2026b). No public fallback was inserted to fill the missing intraday, RSP, or SOXL cases.

The use of total-return indices has an important accounting implication. These index levels incorporate distributions and are not the same thing as historical quoted security prices. Consequently, reconstructed positions are fractional index units and model dollar notionals rather than literal ETF or stock share quantities. The report does not claim independent corporate-action reconstruction, exchange-level fills, queue position, or broker confirmation (OpenClaw, 2026a). This is a limitation, but it is also a strength of the disclosure: the document does not fabricate precision it does not possess.

Table 1

Executed Evidence Versus Original Scope

Dimension	Requested	Executed	Interpretation
Assets	SPY, GLD, QQQ, AAPL, RSP, SOXL	SPY, GLD, QQQ, AAPL	RSP and SOXL daily histories were source-blocked.
Frequencies	15-min, 1-hour, 4-hour, daily	Daily only	Intraday comparisons cannot be ranked or declared inferior.
Asset-frequency cases	24	4	20 cases remain missing evidence, not failed tests.
Prospective outcomes	Required for confirmation	0	A future batch was prepared but not evaluated.

Note. Source scope and coverage are from OpenClaw (2026a).

Model Architecture and Experimental Design

Reservoir Dynamics and Policy Bank

The fly reservoir uses a leaky recurrent update of the form $h_t = (1 - l)h_{(t-1)} + l \tanh(0.9Wh_{(t-1)} + Bx_t)$, with leak $l = 0.2$. A seeded 32-node readout produces a gate from the reservoir state. Twelve momentum experts—three horizons, two thresholds, and two sizes—are modulated by gate factors of −1, 0, and 1 to create 36 clipped actions, with cash and full exposure added for a 38-action bank. Graph weights remain fixed online (OpenClaw, 2026a).

This architecture has one particularly strong scientific feature: it makes topology testable. The fly network is not evaluated in isolation. The protocol includes calibrated strength-preserving nulls, raw-strength nulls, random recurrent controls, no-recurrence controls, conventional and adaptive policies, fixed exposure, buy-and-hold, and cash. If the fly wiring really contributes something unique, it should separate from controls that preserve simpler network statistics or eliminate recurrence altogether.

Selection Timeline and Folds

Each scope places its initial training boundary halfway through the available rows and then divides the remainder into three disjoint test blocks. For each refit, the last quarter of the expanding historical prefix acts as inner validation and chooses between 63- and 126-day objective windows. The selected causal window trajectory is then evaluated over the next block (OpenClaw, 2026a). This is substantially better than selecting a strategy with knowledge of the immediately following test block.

However, causal execution inside a historical replay is not equivalent to a pristine research holdout. The researchers had already seen the broad historical periods while developing earlier versions, which means model family, thresholds, controls, and scenario choices could

have been influenced by prior outcomes. White (2000) formalized the danger of repeated data reuse in model selection, and Bailey et al. (2017) showed why ordinary holdouts can still be unreliable in investment backtesting when many alternatives are tried. The report appropriately labels this as retrospective contamination rather than pretending that prefix causality erases research history.

Risk Metrics and What They Can—and Cannot—Say

SPY and GLD are evaluated with Sharpe ratio; QQQ and AAPL use Sortino ratio against the assumed cash hurdle. Sharpe ratio summarizes mean excess return per unit of total volatility, while Sortino focuses the denominator on downside deviations relative to a target. The latter is useful when upside volatility is not considered harmful, although it becomes statistically unstable when negative observations are sparse. Lo (2002) further demonstrates that naive square-root annualization of Sharpe ratios can be distorted by serial dependence. The report acknowledges these caveats and flags short windows rather than presenting the ratios as exact population quantities.

The study also reports profit, maximum drawdown, exposure, turnover, cost dollars, worst-day and multi-day losses, underwater duration, and an empirical expected-shortfall approximation. This broader risk panel is important because a lower drawdown can be created simply by holding less market exposure. A model that earns less, trades less, and spends half its time in cash can look safer without being a better timing model.

Primary Results

The central result is straightforward. On the common period, arithmetic mean fold objectives for fly / calibrated null / buy-and-hold are 0.9212 / 0.9260 / 1.0630 for SPY Sharpe; 0.5654 / 0.5624 / 1.0521 for GLD Sharpe; 0.8100 / 0.8337 / 1.5656 for QQQ Sortino; and 1.2769

/ 1.2844 / 1.2564 for AAPL Sortino (OpenClaw, 2026a). Relative to buy-and-hold, the fly mean is about 13.3% lower on SPY, 46.3% lower on GLD, and 48.3% lower on QQQ; AAPL is only about 1.6% higher on the mean-of-ratios calculation. Those percentages are descriptive, not inferential statistics.

Table 2

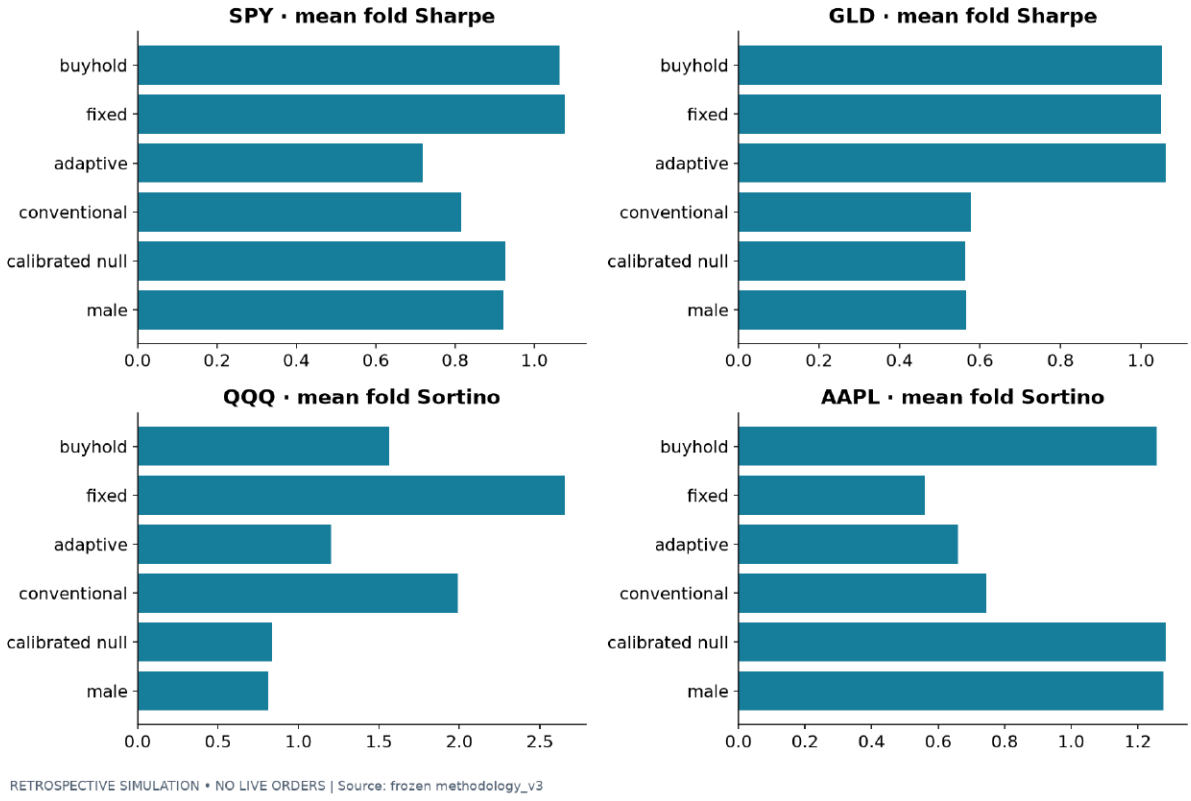
Common-Period Assigned Objectives

Asset	Objective	Fly	Calibrated null	Buy-and-hold	Fly − buy-and-hold
SPY	Sharpe	0.9212	0.9260	1.0630	−0.1418
GLD	Sharpe	0.5654	0.5624	1.0521	−0.4867
QQQ	Sortino	0.8100	0.8337	1.5656	−0.7556
AAPL	Sortino	1.2769	1.2844	1.2564	+0.0205

Note. Values are arithmetic means of fold ratios; recurrent-family values additionally average six anatomy/computational runs. They are not ratios of pooled ensemble returns. Source: OpenClaw (2026a).

Figure 1

Mean Fold Objectives Across Model Families



Note. The original report labels the fly-derived family "male," referring to the male Drosophila connectome source. The figure displays arithmetic mean fold ratios, not a single investable return stream. Source: OpenClaw (2026a).

The AAPL Aggregation Reversal

AAPL is the most instructive case because it demonstrates how a favorable headline can disappear when the aggregation rule changes. The mean fold Sortino is 1.2769 for the fly family versus 1.2564 for buy-and-hold, suggesting a small advantage. Yet when equal capital is assigned to the model-family members and the ensemble return stream is compared directly with buy-and-hold, the paired difference is negative on average across folds (OpenClaw, 2026a). Both calculations are legitimate, but they answer different questions because Sharpe and Sortino are

nonlinear functionals. Selecting whichever aggregation happens to look favorable would be a classic presentation error.

This reversal is a strength of the report because it is not hidden. It also illustrates why model evaluation should be specified before results are seen. The more alternative summaries a researcher computes, the easier it becomes to discover one that flatters the hypothesis—even if the underlying economic effect is weak.

Asset-Level Risk and Trading Behavior

Table 3

Common-Period Fly-Family Risk and Trading Summary

Asset	Mean fold profit / $100k	Mean max drawdown	Mean exposure	Mean turnover	Assigned result vs. buy-and-hold
SPY	$3,976	−3.84%	45.77%	12.46×	Lower Sharpe
GLD	$13,468	−10.32%	61.81%	9.85×	Lower Sharpe
QQQ	$2,441	−6.89%	46.39%	11.97×	Lower Sortino
AAPL	$6,299	−9.15%	41.33%	15.96×	Slightly higher mean ratio; ensemble reverses

Note. These are averages across independent folds and recurrent-family runs. They do not form a continuously compounded live account. Source: OpenClaw (2026a).

SPY

SPY provides a clean example of the difference between absolute profit and benchmark-relative efficiency. The fly family produces positive average fold profit of roughly $3,976 per normalized $100,000, but its mean Sharpe of 0.9212 trails buy-and-hold at 1.0630. Mean exposure is only 45.77%, so some reduction in drawdown is mechanically associated with reduced participation. A positive simulated dollar result therefore does not establish superiority.

GLD

GLD is valuable because a much longer cache history exists. On the common period, fly Sharpe is 0.5654 versus 1.0521 for buy-and-hold; on the maximum-cache scope, fly is 0.2820

versus 0.6451. The calibrated null is slightly behind fly on the common period but ahead on the longer scope (OpenClaw, 2026a). That sign change weakens any argument that the exact anatomical wiring is responsible for performance.

QQQ

QQQ is the weakest benchmark-relative result in the common period. Fly Sortino is 0.8100 versus 1.5656 for buy-and-hold, and the maximum-cache comparison remains unfavorable at 0.8824 versus 1.6033. Mean exposure is 46.39%, which again warns against interpreting lower drawdown as better timing. The worst fold drawdown across the full common-period fly family is −14.34%, materially worse than the mean fold drawdown statistic alone suggests (OpenClaw, 2026a).

AAPL

AAPL has the highest fly-family turnover at 15.96× and a higher base one-way modeled cost of 5.5 basis points, versus 3.5 basis points for the ETFs. The favorable mean-of-ratios result is therefore not only small but also obtained in the asset with the most active trading and the highest modeled unit cost. The calibrated null's mean Sortino of 1.2844 is slightly higher than the fly family's 1.2769 (OpenClaw, 2026a).

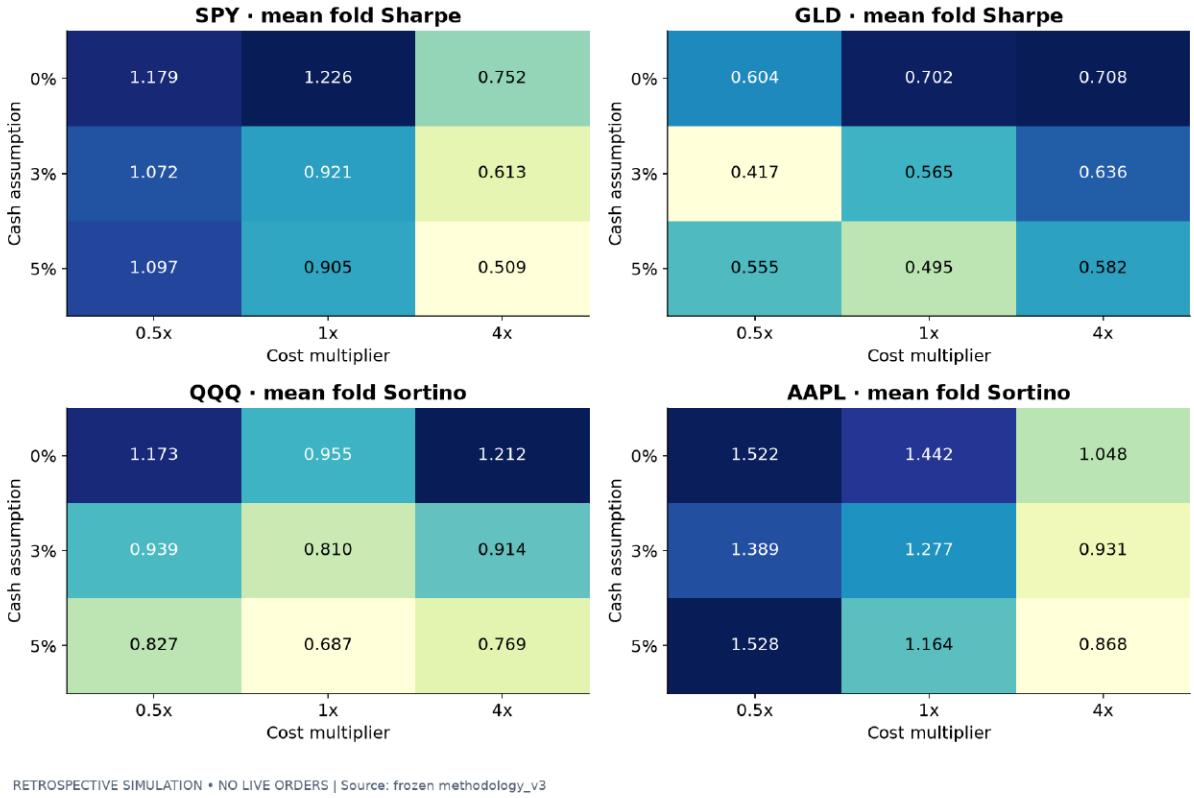
Transaction Costs, Cash, and Behavioral Retraining

The study does more than subtract a fixed fee from a frozen trade list. It crosses cash assumptions of 0%, 3%, and 5% with cost multipliers of 0.5×, 1×, and 4×, then recalculates candidate net histories, online policy choices, validation-window selection, and portfolio accounting. Scenario changes alter signals in 2,048 of the 2,520 runs relative to the base case (OpenClaw, 2026a). This matters because transaction costs can change which strategy appears optimal, not merely how much profit is left after trading.

The broader finance literature likewise treats implementation costs as strategy-dependent rather than cosmetic. Momentum can remain economically meaningful after costs in some settings, but realized capacity and net returns depend on turnover, liquidity, market impact, portfolio construction, and institutional execution (Frazzini et al., 2012). The current report is appropriately narrower: its proportional-cost assumptions are modeling scenarios, not observed fills.

Figure 2

Sensitivity of the Fly Family to Cash and Cost Scenarios



Note. Each cell retrains policy selection under the scenario rather than subtracting a fee from a fixed trade path. Cell values are means across folds and six fly runs. Source: OpenClaw (2026a).

Statistical Uncertainty and Multiplicity

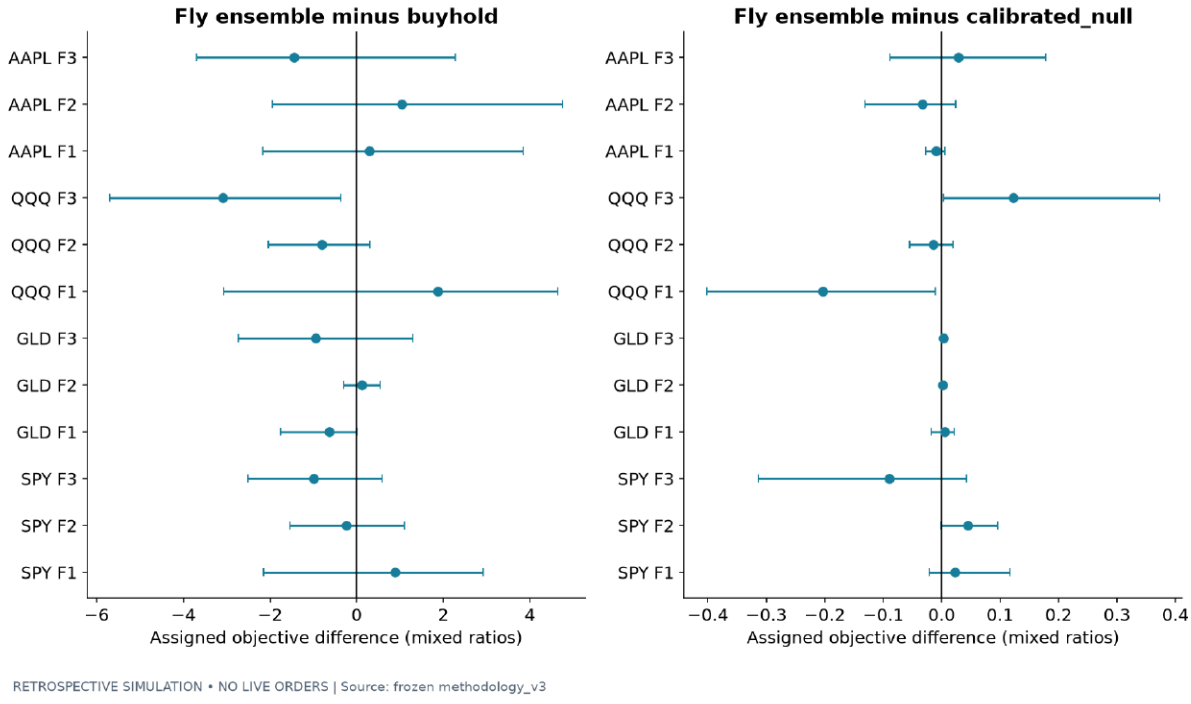
The review uses 2,000 circular block-bootstrap replicates with a 21-day block to compare equal-capital model-family ensemble equity on common dates. The primary exploratory family

contains 240 comparisons: two scopes × four assets × three folds × ten non-cash comparators. A more conservative Bonferroni adjustment is also supplied for a 4,320-family count covering windows and scenarios. The smallest descriptive adjusted p-value in the 240-comparison family is approximately .2399; no adjusted comparison establishes superiority (OpenClaw, 2026a).

This result is consistent with the larger literature on backtest selection. White (2000) emphasizes that repeated use of the same time series for model search can make the best observed rule look better than it truly is. Bailey and López de Prado (2014) propose a deflated Sharpe ratio specifically to address selection bias and non-normal returns, while Bailey et al. (2017) show how many tested configurations can create overfit winners even when standard holdout practices are used. The present study does not claim to solve the entire research-history problem with a single adjustment, and that restraint is appropriate.

Figure 3

Exploratory Paired Objective Differences



Note. Intervals are exploratory 95% circular block-bootstrap intervals and are not simultaneous confidence bands. Mixed Sharpe and Sortino objectives are displayed but not pooled into a universal score. No multiplicity-adjusted superiority is established. Source: OpenClaw (2026a).

Strengths of the Research Program

The empirical conclusion is negative, but the research process has real strengths. These strengths matter because a failed hypothesis tested cleanly is more valuable than a weakly tested hypothesis with attractive charts.

Strong Control Architecture

The most important strength is the breadth of controls. The fly reservoir is compared not only with buy-and-hold but also with calibrated strength-preserving null graphs, raw-strength nulls, random recurrent networks, no-recurrence variants, and conventional policy families. This structure asks whether any observed benefit comes from exact topology, gross graph statistics, recurrence itself, or merely the momentum decoder. The result—that fly and calibrated-null performance are small and inconsistent relative to one another—directly attacks the claimed mechanism instead of only testing profitability.

Causal Execution Discipline

Features use only information available through the relevant date, signals are executed on subsequent available closes, and validation choices come from the expanding prefix rather than the next test block. This does not create a pristine untouched holdout, but it sharply reduces ordinary look-ahead bias in the mechanical simulation (OpenClaw, 2026a).

Reproducible Accounting

The report reconstructs complete daily ledgers, distinguishes tiny nonzero rebalances from the formal turnover-based trade-count threshold, and reconciles equity, fees, turnover, drawdown, Sharpe, and Sortino against saved metrics to an absolute tolerance of 10^−6. It also

records file hashes and source manifests. Hashes do not prove economic truth, but they are useful for proving that a reviewer is working from the same computational evidence.

Prospective Falsifiability

A 140-configuration daily prospective batch is frozen across the four available assets, with known dates excluded from future evaluation. The loader checks source metadata, code, graph, and prefix integrity and is designed to reject repeat evaluation (OpenClaw, 2026a). That does not create future data, but it creates the conditions for a more credible next test. A hypothesis becomes scientifically stronger when the protocol makes future failure easy to observe rather than easy to explain away.

Failures and Limitations

Failure 1: No Demonstrated Trading Advantage

The primary failure is empirical. The fly family does not beat buy-and-hold on the assigned common-period mean fold objective for three of four assets, and the one apparent AAPL advantage reverses under a different, defensible ensemble calculation. Calibrated nulls are usually similar or slightly better. If the research question is “does the male fly topology produce a robust, benchmark-beating trading edge?”, the present evidence says no.

Failure 2: No Untouched Prospective Test

All reported market outcomes were historically available during the broader research process. Expanding-prefix evaluation reduces local look-ahead but cannot undo researcher exposure to the historical record. This is the most consequential inferential weakness. The project needs genuinely future observations evaluated once under frozen rules.

Failure 3: Partial Scope

Twenty of 24 requested asset-frequency cells remain untested because authorized data were unavailable. That prevents any claim about whether intraday frequencies are better or worse than daily and leaves RSP and SOXL profit objectives unresolved. A dataset shortage is not evidence against the model, but it is a direct limit on generalization.

Failure 4: Reduced Biological Fidelity

The experiment samples about 3% of the 166,691-neuron male CNS resource and replaces biological dynamics with a deterministic artificial state equation. Many transmitter classes are computationally silent; leak, scaling, tanh saturation, and readout selection are imposed by the model designer. This is acceptable for a topology-inspired engineering experiment, but it blocks stronger claims about “biological intelligence” or neural computation in the organism.

Failure 5: Simplified Execution and Financing

The ledger uses fractional total-return-index units and proportional transaction costs. It does not reconstruct quoted historical shares, bid-ask dynamics, market impact, order queues, venue routing, or real broker cash credit. The constant annual cash assumptions are scenario inputs rather than observed historical financing rates. These simplifications are reasonable for early research but material for any claim of deployable profitability.

Failure 6: Ratio and Aggregation Sensitivity

Short folds, downside-only denominators, serial dependence, and nonlinear aggregation make Sharpe and Sortino comparisons noisy. AAPL’s reversal is the clearest example. Arithmetic means of fold ratios, pooled returns, and equal-capital ensemble ratios are not

interchangeable. A future protocol should specify a single primary aggregation rule and a small set of secondary diagnostics in advance.

Table 4

Strengths and Failures in One View

Component	Strength	Remaining failure / risk
Controls	Multiple topology, recurrence, conventional, buy-and-hold, and cash comparators.	No topology-specific advantage is established.
Temporal discipline	Prefix-causal features, validation, and next-close accounting.	Broader research history was already exposed to the data.
Costs	Policies are retrained under cost/cash scenarios.	Costs are modeled, not observed microstructure fills.
Risk reporting	Profit, drawdown, exposure, turnover, tails, and ratios are reported.	Short folds and nonlinear ratios remain noisy.
Reproducibility	Hashes, manifests, deterministic ledgers, and reconciliation checks.	Reproducibility confirms computation, not economic causality.
Biological source	Connectome lineage is documented.	5,000-node artificial reservoir is not a whole-brain physiological model.
Prospective design	Future batch is frozen and excludes known dates.	No future market outcome has yet been evaluated.
Coverage	Four daily assets are analyzed.	20 of 24 requested asset-frequency cases are still missing.

Interpretation: What the Negative Result Actually Means

The strongest defensible statement is “no demonstrated fly-model advantage.” That is narrower than “the model loses money,” because several folds and assets show positive dollar profits. It is also narrower than “connectome-inspired computation cannot work,” because only one reduced implementation and a partial market scope were tested. The result rejects a stronger claim: the available retrospective evidence does not show that this specific fly-derived wiring adds economic value beyond appropriate controls.

This distinction is important for research culture. Quantitative projects often treat a backtest as a product demo, which creates pressure to foreground whichever statistic looks best. Here, the scientifically interesting result is the opposite: a technically elaborate system failed to create clear incremental value. That failure identifies where complexity is not paying rent. The reservoir may be generating rich internal dynamics, but richness alone is not an edge.

A Better Prospective Research Program

The next stage should be narrower, preregistered, and harder to game. The goal should not be to “improve the backtest” on the same historical period. It should be to determine whether the frozen hypothesis survives future evidence.

Primary Prospective Test

The 140 prepared daily configurations should be evaluated only after a meaningful amount of genuinely new data accumulates. The primary endpoint should be specified in advance for each asset, including the exact aggregation rule, cost scenario, benchmark, and multiplicity family. Results should be reported for every configured model and every available asset, including failures and undefined ratios. Repeated peeking should either be prohibited or explicitly incorporated into the error-control procedure.

Add an Ablation Ladder

A useful ablation sequence would compare: (a) price-only momentum without recurrence, (b) random recurrent topology, (c) degree/strength-preserving null topology, (d) the fly topology, and (e) alternative non-biological structured graphs matched for spectral and memory properties. The critical question is not whether a complicated network can trade, but whether the anatomical topology produces incremental predictive or risk-allocation value after matching generic dynamical properties.

Measure the Mechanism, Not Only the P&L

If the hypothesis is that biological topology provides useful memory, future work should predefine graph-memory diagnostics under market-driven inputs and test whether those diagnostics mediate performance. A topology claim is stronger when it links graph structure to measurable state dynamics and then to economic outcomes. Without that chain, “connectome-

derived” risks becoming a decorative label attached to an otherwise conventional momentum system.

Upgrade Execution Realism Only After Signal Validation

It would be premature to build a high-fidelity broker simulator before establishing predictive value. If the prospective signal survives, then execution assumptions should be upgraded: quoted prices rather than total-return-index notionals, historically varying cash rates, spreads, slippage, market impact, and capacity. This sequencing avoids spending engineering effort on the plumbing of a strategy that has not yet demonstrated an edge.

Conclusion

The fly-derived momentum project is a serious engineering experiment with an intentionally unconventional source of recurrent topology. Its strongest achievement is not a trading result but a research framework capable of producing a credible negative finding. The study documents data lineage, executes thousands of model/scenario combinations, applies multiple topology and conventional controls, reconstructs accounting deterministically, explores uncertainty, and prepares a frozen prospective protocol. Those are meaningful strengths.

The economic evidence, however, is not favorable. SPY, GLD, and QQQ trail buy-and-hold on their assigned common-period mean fold objectives. AAPL’s small favorable mean-of-ratios result disappears under an equal-capital ensemble comparison, and the calibrated null remains slightly ahead on the mean Sortino. No multiplicity-adjusted comparison establishes superiority. The reduced reservoir is not a physiological model, most requested asset-frequency cells remain untested, and all current performance is retrospective.

The appropriate scientific response is therefore not to rescue the narrative with a better-looking seed, chart, timeframe, or aggregation rule. It is to keep the model frozen and let new

data decide. If the prospective test succeeds across controls, the project will have stronger evidence than another round of historical tuning could provide. If it fails, that failure will still be useful: it will show that biological topology can be computationally interesting without being economically predictive. In quantitative research, that is a result worth knowing.

References

Bailey, D. H., Borwein, J. M., López de Prado, M., & Zhu, Q. J. (2017). The probability of backtest overfitting. Journal of Computational Finance, 20(4), 39–69. https://doi.org/10.21314/JCF.2016.322

Bailey, D. H., & López de Prado, M. (2014). The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting, and non-normality. The Journal of Portfolio Management, 40(5), 94–107. https://doi.org/10.3905/jpm.2014.40.5.094

Berg, S., Beckett, I. R., Costa, M., Schlegel, P., Januszewski, M., Marin, E. C., Nern, A., Preibisch, S., Qiu, W., Takemura, S.-y., Fragniere, A. M. C., Champion, A. S., Adjavon, D. Y., Cook, M., Gkantia, M., Hayworth, K. J., Huang, G. B., Katz, W. T., Kämpf, F., … 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

Frazzini, A., Israel, R., & Moskowitz, T. J. (2012). Trading costs of asset pricing anomalies [Working paper]. AQR Capital Management.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of Finance, 48(1), 65–91. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Lo, A. W. (2002). The statistics of Sharpe ratios. Financial Analysts Journal, 58(4), 36–52. https://doi.org/10.2469/faj.v58.n4.2453

Lukoševičius, M., & Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. Computer Science Review, 3(3), 127–149. https://doi.org/10.1016/j.cosrev.2009.03.005

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. Journal of Financial Economics, 104(2), 228–250. https://doi.org/10.1016/j.jfineco.2011.11.003

OpenClaw. (2026a). v3 final feasible-scope report, protocol, paths, metrics, review and prospective preparation [Unpublished internal research artifacts]. methodology_v3/.

OpenClaw. (2026b). MaleCNS online momentum study: Data and anatomy provenance manifests [Unpublished internal research artifacts]. malecns_momentum_20260913_scoped/data/.

OpenClaw. (2026c). Memory calibration and methodology v2 evidence [Unpublished internal research artifacts]. methodology_v2/.

Sortino, F. A., & van der Meer, R. (1991). Downside risk. The Journal of Portfolio Management, 17(4), 27–31.

White, H. (2000). A reality check for data snooping. Econometrica, 68(5), 1097–1126. https://doi.org/10.1111/1468-0262.00152

Appendix A

Technical Definitions and Reproducibility Notes

Daily return is defined as $r_t = P_t / P_{(t-1)} - 1$. Momentum inputs use 5-, 20-, and 60-bar horizons scaled by a 20-day volatility estimate with a floor of 0.001 and clipped to $[-5, 5]$. A fourth input scales rolling volatility by 0.02. The reservoir leak is 0.2 for the fly model, recurrent scaling is 0.9, and the gate is computed from a seeded 32-node readout (OpenClaw, 2026a).

Cash assumptions are 0%, 3%, and 5% annualized with calendar-day compounding on the cash balance. Cost multipliers are 0.5×, 1×, and 4× the base one-way proportional cost. The selected-fold order reconstruction treats the preceding available close as the signal date and the next available close as the execution date. Nonzero modeled transactions are exported; broker order IDs, venue status, and market fills are not synthesized (OpenClaw, 2026a).

The frozen protocol enumerates 35 configurations per scope: six fly runs, six calibrated-strength nulls, six raw-strength nulls, six random recurrent controls, six no-recurrence controls, and one each of conventional, adaptive, fixed, buy-and-hold, and cash. Two anatomical samples and three computational seeds generate the six recurrent-family runs. Across two scopes, four available assets, and nine cash/cost scenarios, this produces 2,520 runs and 7,560 fold metric rows (OpenClaw, 2026a).