

Skeptical review: 3.2. Behavioral and neuroimaging data quantification

Summary

This manuscript proposes a residual-based “neuro-cognitive decoupling” framework to operationalize resilience/vulnerability in cognitive aging in Egyptian fruit bats (Introduction; Sec. 2). The intended workflow is: (i) fit “normative” models for brain microstructure (DTI mean diffusivity; $Global_{MD}$ and 24 atlas ROIs) and for behavioral metrics from a three-phase spatial memory task using epigenetic age plus covariates (sex, origin colony) (Sec. 2.1, Sec. 2.2, Sec. 2.4.1); (ii) compute individual decoupling indices as residuals (observed – predicted) (Sec. 2.4.2); and (iii) associate brain residuals with behavioral residuals to localize regions where microstructural deviation relates to better/worse-than-expected cognition (Sec. 2.4.3, Sec. 3.3). The imaging pipeline and ROI extraction appear to run and yield plausible MD distributions and inter-ROI correlation structure (Sec. 2.3; Sec. 3.2.2–3.3.1). However, the behavioral parsing/metric extraction fails (NaNs/zero variance), and the manuscript replaces behavioral outcomes with synthetic values drawn from normal distributions for the final decoupling demonstrations (Sec. 2.2; Sec. 3.2.1; Sec. 3.3.2). As a result, the manuscript currently functions primarily as a methods/pipeline proof-of-concept rather than an empirical neuro-cognitive aging study; several core claims (brain–behavior associations, ROI “findings”) are necessarily illustrative and require either (a) repaired behavioral extraction or (b) a more rigorous simulation validation and clearer reframing.

Strengths

- Clear, intuitive conceptual framing: defining resilience/vulnerability via deviations from age-expected brain and cognition aligns with cognitive reserve/brain maintenance traditions and is easy to communicate (Introduction; Sec. 2.4).
- Biologically motivated use of epigenetic age (plus sex and colony) to define normative expectations, which goes beyond simple chronological-age adjustment (Sec. 2.1; Sec. 2.4.1; Sec. 3.1).
- Practical multimodal integration work (ID harmonization, data merging, atlas-based ROI extraction) in a non-traditional model species is valuable and reusable (Sec. 2.1.1–2.3; Sec. 3.2.2).
- Imaging-derived outputs appear internally consistent (units, plausible ranges; strong ROI correlation structure), suggesting the DTI side of the pipeline is close to publishable as a methods contribution (Sec. 3.2.2–3.3.1).
- The manuscript is generally transparent that behavioral results are synthetic/illustrative, reducing the risk of outright biological over-claiming (Abstract; Sec. 3.2.1; Conclusions).

- The three-step algebraic logic (normative fit \rightarrow residuals \rightarrow residual–residual association) is coherent and, if carefully implemented with appropriate inference safeguards, could support a useful resilience/vulnerability index (Sec. 2.4).

Major issues

1. **The central empirical aim (identifying genuine neuro-cognitive decoupling/resilience in aging bats) is not achieved because all cognitive outcomes used for decoupling analyses are synthetic (normal draws), so reported ROI associations, p-values, and R^2 cannot be interpreted biologically and may be misread as real discoveries (Abstract; Sec. 2.2; Sec. 3.2.1; Sec. 3.3.2; Sec. 3.4; Sec. 4.3–4.4).**

Recommendation: Decide and implement one of two coherent paper identities: (i) Methods/pipeline + simulation validation: explicitly reframe the manuscript as a technical paper and restructure Results so that any brain–behavior “findings” are presented only as simulation-based validation with known ground truth (see next issue). (ii) Empirical bat aging study: repair behavioral extraction and rerun the full pipeline on real behavioral metrics, then (and only then) interpret ROI patterns. In either case, remove or heavily qualify inferential language in Sec. 3.3.2/Sec. 4 that reads like biological discovery, and add prominent labeling in-text and in figure captions wherever synthetic behavior is used.

2. **Behavioral extraction failure is the practical bottleneck, but the manuscript provides only a high-level statement (NaNs/zero variance) without a technical post-mortem, making it hard for others (or the authors) to reproduce, diagnose, or fix the pipeline (Sec. 2.2.1–2.2.2; Sec. 3.2.1).**

Recommendation: Expand Sec. 2.2 and Sec. 3.2.1 with an actionable failure analysis: include representative raw log excerpts (rows/columns), enumerate parsing assumptions (timestamp formats, event codes, sheet names, delimiters, missing values), specify which assumptions were violated, and show intermediate sanity checks (e.g., number of visits per trial, time ordering). Provide a corrected or more robust parsing approach (e.g., schema validation, flexible timestamp parsing, explicit trial boundary detection) and a minimal manual-validation protocol (spot-check N trials vs. video/hand labels if available). If permitted, share anonymized example logs and the parsing script.

3. **Synthetic behavioral data are under-specified and internally inconsistent: the manuscript alternates between simulating “behavioral residuals” vs. simulating raw behavioral metrics and then residualizing them, which changes the mathematics and interpretation (Sec. 3.2.1 vs. Sec. 3.3.1; Sec. 2.4.1–2.4.2). The current approach is not a meaningful validation because random normal draws can still yield nominally significant results under multiple testing.**

Recommendation: Add a dedicated Methods subsection (e.g., Sec. 2.2.3) specifying exactly what is simulated (raw metrics vs residuals), distributions (means/SDs), constraints (non-negativity, time caps), correlation structure across behavioral metrics, and random seeds. Replace ad hoc normal draws with a formal simulation study aligned to the paper’s goal: • Null simulations (no planted brain–behavior link) to quantify type-I error under the full 6×25 testing regime. • Signal simulations with planted effects in selected ROIs (and with correlated ROI predictors reflecting the empirical ROI correlation matrix from Sec. 3.3.1) to show sensitivity/specificity and effect recovery. Report performance with and without multiple-comparisons correction (see next issue).

4. **Multiplicity and correlated predictors are not handled: the framework implies ~ 150 ROI-by-metric brain–behavior tests (6 behavior metrics \times (Global+24 ROIs)), with ROIs strongly correlated (Sec. 3.3.1), yet p-values are presented without an explicit correction strategy (Sec. 2.4.3; Sec. 3.3.2). With $N \approx 30\text{--}33$, uncorrected significance is not interpretable.**

Recommendation: In Sec. 2.4.3, specify a primary inferential plan for real data (e.g., FDR across all ROI tests per behavioral metric, or across all tests; or permutation-based max- T controlling family-wise error). In Sec. 3.3.2, either (a) report corrected results (even if illustrative) or (b) label all p-values as uncorrected and non-inferential. Consider adding a complementary multivariate strategy that reduces the multiple-testing burden and handles correlated ROIs (e.g., PCA/PLS on ROI MD residuals; ridge/elastic-net with cross-validation) while keeping the residual-based “decoupling index” as the interpretability layer.

5. **Two-stage residualization (brain residuals and behavior residuals computed from fitted normative models, then regressed residual-on-residual) raises inference/overfitting concerns in small samples if uncertainty from stage-1 fits is ignored (“generated regressor” / potential optimism). The manuscript does not clarify equivalence to single-stage covariate adjustment or how uncertainty is propagated (Sec. 2.4.1–2.4.3; Sec. 3.3).**

Recommendation: Clarify in Sec. 2.4 that the residual–residual regression is (under standard OLS conditions) related to a single-stage model (e.g., $Behavior \sim ROI_{MD} + Age + Sex + Colony$) but that two-stage procedures can complicate uncertainty accounting. For robust inference, add one of: • Cross-fitting: fit normative models in training folds and compute residuals in held-out folds before testing associations. • Full-pipeline bootstrap: resample animals, refit normative models, recompute residuals, and refit decoupling models to obtain confidence intervals for β and R^2 . Also state whether interactions/nonlinear age terms were considered (e.g., splines/quadratic), since mis-specified normative models can distort residuals.

- 6. Behavioral task protocol and MRI acquisition/preprocessing are under-described, limiting reproducibility and interpretability (Sec. 2.2; Sec. 2.3). Missing details include key spatial-memory task design/definitions and core DTI acquisition parameters and preprocessing/QC steps (scanner/sequence, b-values, directions, voxel size, motion/eddy correction, tensor fitting, atlas registration, exclusion criteria).**

Recommendation: Substantially expand Sec. 2.2 with a precise task description: apparatus geometry, number/identity of boxes, definition of trial/session, phase structure/timing, reward contingencies, criteria for progression, and handling of non-compliance/aborts; define edge cases for each metric (e.g., if correct box never visited). Expand Sec. 2.3 into (i) acquisition (scanner, field strength, sequence, b-values, #directions, TR/TE, resolution, anesthesia/handling) and (ii) preprocessing/QC (denoising, motion/eddy/susceptibility correction, tensor estimation, software versions, registration method/metrics, QC thresholds, exclusions). Provide atlas provenance (Sec. 2.3.1) sufficient for ROI extraction reproduction.

- 7. Cohort size/composition is inconsistent across sections (e.g., $N = 30$ in Sec. 2.1 vs $N = 33$ in Sec. 3.1/Sec. 3.2.2; sex/colony counts differ; epigenetic age max differs 13.84 vs 15.07). This makes it unclear which animals contribute to which models and figures.**

Recommendation: Create one authoritative cohort accounting table (in Sec. 2.1 or as Table 1) listing N for: (a) DTI available, (b) behavior logs available, (c) behavior metrics successfully extracted, (d) multimodal intersection used in each analysis. Include sex/colony breakdown and epigenetic-age range per subset. Update all text and figure captions in Sec. 3.1–3.3 to match, and briefly state exclusion reasons (missingness, QC failures).

- 8. ROI interpretability is blocked because ROIs are referred to only as ROI_1...ROI_24 without anatomical names; atlas details are minimal (Sec. 2.3.1; Sec. 3.3.2; Sec. 4.3). This limits both biological interpretation and future reuse.**

Recommendation: Provide an ROI lookup table (main text or Appendix) mapping ROI indices to anatomical labels (and laterality), plus voxel counts/volumes. In Sec. 3.3.2 and Discussion, refer to ROIs as “ ROI_k (RegionName)”. Expand Sec. 2.3.1 with atlas origin (species-specific vs adapted), resolution, and registration QC examples (e.g., overlay snapshots).

- 9. Positioning/novelty is somewhat overstated relative to established human-literature residual frameworks (cognitive reserve, brain maintenance, brain-age/cognitive-age residuals). The manuscript currently under-engages with this lineage, making it hard to see what is fundamentally new beyond application to bats and epigenetic age (Sec. 1; Sec. 4.1).**

Recommendation: In Sec. 1 and Sec. 4.1–4.4, add a concise related-work paragraph explicitly connecting to residual-based resilience/cognitive reserve frameworks and clarifying what is novel here (e.g., epigenetic-age normative model; bat model; atlas-based DTI pipeline). Moderate claims in Sec. 3.4/Sec. 4.4 to reflect proof-of-concept status until real behavioral metrics are available.

Minor issues

1. Regression/model specification is incomplete: coding of Sex/Origin, handling of missingness/outliers, centering/scaling of epigenetic age, and consideration of nonlinearity/interactions are unclear (Sec. 2.4.1–2.4.3). Binary outcomes (STM/LTM first-choice correct) appear treated with OLS without justification (Sec. 2.2.2; Sec. 2.4.1).

Recommendation: In Sec. 2.4.1, present models in standard notation and specify categorical coding (reference levels), missing-data strategy, and any transformations. For binary outcomes, either justify linear probability models (and interpret residuals carefully) or switch to GLMs (logistic) and define how “decoupling indices” are computed on the appropriate scale (e.g., deviance residuals or probability-scale residuals). Consider testing nonlinear age effects (splines/quadratic) given aging biology.

2. Global_MD vs ROI_MD: the manuscript does not discuss whether ROI residual effects are region-specific versus reflecting global microstructural variation; multicollinearity among ROIs (and with Global_MD) is likely (Sec. 2.3; Sec. 3.3.1–3.3.2).

Recommendation: Clarify intended use of $Global_M D$ in Sec. 2.4.3 (summary vs covariate). Add a brief correlation/variance-partition check (e.g., ROI residuals vs $Global_M D_{Residual}$) and consider models that adjust ROI residuals for global residual (or use global-first residualization) when the goal is region-specific inference.

3. Figures/presentation: figure numbering/cross-references are inconsistent, captions are sometimes duplicated, and several plots are difficult to read (small fonts, crowded labels) or potentially misleading (pie charts; correlation heatmap for non-viable behavior) (Sec. 3.1–3.3.2). Several figures do not clearly label synthetic vs real components or whether p-values are corrected.

Recommendation: Renumber figures sequentially and ensure each in-text reference points to the correct figure. Increase font sizes and use colorblind-safe palettes. Replace pie charts with bar charts. Remove/relocate any plots based on non-viable behavioral extraction (or explicitly label them as demonstrating failure). In every relevant caption, state N , whether data are synthetic, the model/covariates used, and whether multiplicity correction is applied.

4. Epigenetic age clock is central but under-contextualized (derivation, expected error, rationale for skin tissue) (Sec. 2.1; Sec. 4.2).

Recommendation: Add a brief description with citation(s): training/validation context, typical prediction error, and justification for using *DNAmAgeBat...skin* in this cohort; clarify whether any calibration or QC was applied to this dataset.

5. DTI/MD biological interpretation is somewhat oversimplified as “inverse integrity” (Sec. 2.3; Sec. 2.4.2; Sec. 3.2.2).

Recommendation: Revise wording to reflect that *MD* is sensitive but non-specific (can reflect multiple processes), and make the paper’s sign convention explicit for interpretability of “resilience/vulnerability” indices.

6. Limited discussion of sample-size/power and uncertainty: with $N \approx 30\text{--}33$ and many correlated predictors, estimates and R^2 can be unstable (Sec. 3.3–3.4).

Recommendation: Add uncertainty quantification (bootstrap CIs for β and R^2) and a short power/stability discussion in Sec. 3.4 or Sec. 4.4; emphasize estimation over dichotomous significance.

7. Reproducibility artifacts: ID harmonization includes ad hoc mappings (e.g., special ID renames; sex symbol conversions) without validation description (Sec. 2.1.1).

Recommendation: Briefly justify why special cases occurred and how correctness was verified (cross-checking logs/metadata). If possible, provide the harmonization map as supplementary material.

8. Broader contribution would be stronger with explicit code/data availability, especially if positioned as a pipeline/methods paper (Sec. 2–3).

Recommendation: Provide a public repository (or archived supplement) with: ROI extraction scripts, regression/plotting code, a data dictionary/schema, and (if raw data cannot be shared) a synthetic example dataset matching the required structure to run the full pipeline end-to-end.

Very minor issues

1. Keywords/metadata include unrelated terms (e.g., astronomy), suggesting a template carryover (Abstract/keywords).

Recommendation: Replace with domain-appropriate keywords (cognitive aging, DTI, mean diffusivity, epigenetic age, Egyptian fruit bat, cognitive reserve/resilience, residual modeling).

2. Formatting/typography inconsistencies (line breaks mid-word; inconsistent section heading formats; inconsistent variable naming with/without underscores; inconsistent unit formatting) appear throughout (Sec. 1–4; figure captions).

Recommendation: Proofread the compiled manuscript and standardize: section numbering, variable naming conventions (code vs prose), and unit/symbol formatting (e.g., consistent “ mm^2/s ”, consistent “ R^2 ”).

3. Some captions/axes use coded labels (ROI_#, underscores) without reader-friendly formatting, and some plots show excessive numeric precision.

Recommendation: Use human-readable axis labels (e.g., “ROI 6 (RegionName) MD residual”), round tick labels/percentages appropriately, and export figures as vector/high-DPI for legibility.

Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

Maths relevance: light

The paper’s core mathematics is a residual-based linear regression framework: fit normative linear models for each brain/cognitive metric using age/sex/colony, define decoupling indices as residuals (observed minus predicted), then regress cognitive residuals on *MD* residuals across ROIs. The formulas present are simple and mostly consistent, but two central internal-consistency problems affect the mathematical definability of the analysis: (i) conflicting cohort definitions (*N* and demographic counts/ranges), and (ii) inconsistent statements about whether synthetic behavioral data are metrics or residuals.

Checked items

1. ✓ **Decoupling residual definition (intro)** (Introduction, p.2 (residual definition $RM = M_{observed} - M_{predicted}$))
 - **Claim:** Defines a decoupling index/residual as observed value minus predicted (age-expected) value.
 - **Checks:** symbol/definition consistency, sign convention sanity check
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** A predictive (normative) model exists that outputs $M_{predicted}$ for each subject based on covariates.
 - **Notes:** The residual definition is standard and reused later with the same sign (observed – predicted).
2. ✓ **Perseveration index formula (STM/LTM)** (Sec. 2.2.2, p.3 ($STM_{perseveration_index}$ and $LTM_{perseveration_index}$))
 - **Claim:** Index = Entries to prior-phase location / (Total entries to all other incorrect locations + 1).
 - **Checks:** algebra/formula structure, normalization/constraints, edge-case handling
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Counts of entries are nonnegative integers., The “prior-phase location” is excluded from “all other incorrect locations.”

- **Notes:** Index is dimensionless; +1 prevents division by zero. Index is unbounded above, which is consistent with a ratio interpretation.
3. ✓ **Global MD definition from atlas mask** (Sec. 2.3.1, p.3)
- **Claim:** $Global_{MD}$ is the mean MD across all non-zero atlas voxels (union of ROIs).
 - **Checks:** definition consistency, dimensional/unit sanity check
 - **Verdict:** PASS; confidence: medium; impact: minor
 - **Assumptions/inputs:** Atlas labels define ROIs; non-zero voxels represent the union mask., MD map and atlas are in the same voxel space.
 - **Notes:** Averaging MD preserves MD units. No conflicting definition elsewhere.
4. ✓ **Normative model specification** (Sec. 2.4.1, p.4)
- **Claim:** Each metric is modeled linearly as a function of epigenetic age, sex, and origin colony.
 - **Checks:** notation/definition clarity, internal consistency with later steps
 - **Verdict:** PASS; confidence: medium; impact: moderate
 - **Assumptions/inputs:** Linear regression includes an intercept (not stated but typical)., Sex and origin colony are encoded as categorical predictors.
 - **Notes:** The model statement is consistent with the later residual computation and decoupling regressions, though the exact coding is not specified.
5. ✓ **Residual computation (methods)** (Sec. 2.4.2, p.4)
- **Claim:** Residual = Observed_Value – Predicted_Value for each metric for each bat.
 - **Checks:** symbol/definition consistency
 - **Verdict:** PASS; confidence: high; impact: critical
 - **Assumptions/inputs:** Predicted_Value comes from the normative model in Sec. 2.4.1.
 - **Notes:** Matches the introduction’s definition and is the foundation for all subsequent residual-vs-residual analyses.
6. ✓ **Sign interpretation for Time_to_First_Correct residual** (Sec. 2.4.2, p.4 (and reiterated Sec. 3.3.1, p.7))
- **Claim:** Negative residual for $Time_{tOFirstCorrect}$ indicates faster learning than expected.
 - **Checks:** limiting/sanity case
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Lower $Time_{tOFirstCorrect}$ means better/faster learning.

- **Notes:** If observed time < predicted time, residual is negative; that corresponds to faster-than-expected learning.
7. ✓ **Sign interpretation for MD residual** (Sec. 2.4.2, p.4 (and reiterated Sec. 3.3.1, p.7))
- **Claim:** Positive *MD* residual implies poorer integrity than expected; negative implies better-than-expected integrity.
 - **Checks:** definition consistency, limiting/sanity case
 - **Verdict:** PASS; confidence: medium; impact: minor
 - **Assumptions/inputs:** Higher *MD* corresponds to lower microstructural integrity within the paper’s interpretation.
 - **Notes:** Internally consistent given the stated “inverse measure” interpretation; the paper could benefit from a more explicit integrity↔ *MD* mapping statement.
8. ✓ **Decoupling regression (residual vs residual)** (Sec. 2.4.3, p.4)
- **Claim:** For each behavioral residual, regress it on each ROI *MD* residual via a simple linear regression.
 - **Checks:** derivation logic, definition consistency
 - **Verdict:** PASS; confidence: medium; impact: moderate
 - **Assumptions/inputs:** Residuals are computed from normative models that use the same covariates.
 - **Notes:** This is conceptually consistent: associating deviations-from-norm in brain and behavior after adjusting for the same covariates.
9. ✗ **Synthetic behavioral data: residuals vs metrics inconsistency** (Results Sec. 3.2.1, p.5 vs Results Sec. 3.3.1, p.7 (also Methods Sec. 2.4.1–2.4.2, p.4))
- **Claim:** Paper alternates between generating synthetic behavioral residuals directly and generating synthetic behavioral metrics then fitting normative regressions to obtain residuals.
 - **Checks:** definition consistency, derivation/pipeline consistency
 - **Verdict:** FAIL; confidence: high; impact: critical
 - **Assumptions/inputs:** If synthetic residuals are generated directly, they should be labeled residuals and the normative regression step is not applied (or must be described as producing identical residuals by construction)., If synthetic metrics are generated, residual generation proceeds via the normative model.
 - **Notes:** Sec. 3.2.1 states “generating synthetic behavioral residuals,” but Sec. 3.3.1 states regression models were fitted for “6 synthetic behavioral metrics” and residuals then computed. These cannot both be true without additional explanation (e.g., metrics generated then residualized, or residuals generated with covariate structure).

10. ✘ **Cohort size and demographic consistency (analysis set)** (Methods Sec. 2.1, p.2 vs Results Sec. 3.1, p.5 (also Abstract p.1))
- **Claim:** The final analytical cohort size and composition are consistent across sections.
 - **Checks:** symbol/definition consistency, data-dimension consistency (N)
 - **Verdict:** FAIL; confidence: high; impact: critical
 - **Assumptions/inputs:** A single final analysis cohort underlies the regression models unless explicitly partitioned.
 - **Notes:** Methods report a final analytical cohort of 30 bats (18M/12F; 17 Aseret/13 Herzeliya; age range includes 15.07). Results report a final analytical cohort of 33 bats (22M/11F; 18 Aseret/15 Herzeliya; age range max 13.84). The regression/residual definitions depend on N and covariate distributions, so the mathematical object being analyzed is unclear.
11. ✘ **Epigenetic age range inconsistency** (Abstract p.1 and Methods Sec. 2.1, p.2 vs Results Sec. 3.1, p.5)
- **Claim:** Epigenetic age ranges used in the paper are consistent.
 - **Checks:** definition consistency
 - **Verdict:** FAIL; confidence: high; impact: moderate
 - **Assumptions/inputs:** Epigenetic age values are defined on the same scale across the study.
 - **Notes:** Abstract/Results give range **6.62–13.84**, while Methods give final cohort range **6.62–15.07**. Without clarification (e.g., different cohorts), the normative regression domain is ambiguous.
12. ✔ **Units statement for MD** (Sec. 3.2.2, p.6)
- **Claim:** Global MD is reported in mm^2/s ; regional MD s are conceptually in the same unit.
 - **Checks:** dimensional/unit consistency
 - **Verdict:** PASS; confidence: medium; impact: minor
 - **Assumptions/inputs:** MD maps store values in mm^2/s or equivalent physical units.
 - **Notes:** No unit conflicts were found in text; MD remains in diffusion units throughout.

Limitations

- Audit is restricted to the provided PDF text/images; no supplementary materials, code, or appendices were available to disambiguate cohort definitions or the exact synthetic-data generation procedure.
- Most mathematical content is described narratively rather than with numbered equations; locations are therefore referenced by section and page.

- Because intermediate regression equations (e.g., explicit $\hat{\beta}$ formulas, coding of categorical variables) are not shown, verification is limited to definitional and logical consistency rather than step-by-step derivation.

Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

14 checks passed as specified, including internal sum/percent consistency for $N = 30$ (Methods) and $N = 33$ (Results), plus sanity/bounds checks. Two cross-section consistency checks flagged substantive discrepancies: final cohort size (30 vs 33) and epigenetic age maximum (15.07 vs 13.84). One implied-total calculation ($25 \times 6 = 150$ regressions) is computed but cannot be verified against an explicitly reported total.

Checked items

- ✓ **C1** (Methods §2.1 (page 2): 'final analytical cohort ... consisted of 30 bats (18 males, 12 females)')
 - **Claim:** Sex counts sum to the stated final analytical cohort size (30).
 - **Checks:** parts_vs_total
 - **Verdict:** PASS
 - **Notes:** Sum(parts)= 30.0 vs total= 30.0.
- ✓ **C2** (Methods §2.1 (page 2): 'Origin colonies were balanced, with 17 bats from Aseret and 13 from Herzeliya.')
 - **Claim:** Origin colony counts sum to the stated final analytical cohort size (30).
 - **Checks:** parts_vs_total
 - **Verdict:** PASS
 - **Notes:** Sum(parts)= 30.0 vs total= 30.0.
- ✓ **C3** (Methods §2.1 (page 2): 'mean epigenetic age of 9.75 ± 1.81 years (range: 6.62–15.07 years)')
 - **Claim:** Mean \pm SD should lie within the stated min/max range (a basic sanity check).
 - **Checks:** range_sanity
 - **Verdict:** PASS
 - **Notes:** Checked mean within [min,max] (sanity check).
- ✓ **C4** (Results §3.1 (page 5): 'final analytical cohort of 33 bats ... comprised 22 males (66.7%) and 11 females (33.3%)')
 - **Claim:** Sex counts sum to 33 and reported percentages match counts/total.

- **Checks:** parts_vs_total_and_percent
 - **Verdict:** PASS
 - **Notes:** Checked counts sum to total; percents match counts/total within tolerance; percent sum near 100.
5. ✓ **C5** (Results §3.1 (page 5): 'Aseret (n=18, 54.5\%) and Herzeliya (n=15, 45.5\%)')
- **Claim:** Origin counts sum to **33** and reported percentages match counts/total.
 - **Checks:** parts_vs_total_and_percent
 - **Verdict:** PASS
 - **Notes:** Checked counts sum to total; percents match counts/total within tolerance; percent sum near 100.
6. ✓ **C6** (Results §3.1 (page 5): 'epigenetic ages ranging from **6.62** to **13.84** years (Mean = 9.47 ± 1.58 years)')
- **Claim:** Mean lies within stated range (sanity check).
 - **Checks:** range_sanity
 - **Verdict:** PASS
 - **Notes:** Checked mean within [min,max] (sanity check).
7. ✓ **C7** (Results §3.2.2 (page 6): 'Mean Diffusivity (MD) ... mean of **0.000734** mm²/s (SD = **0.000035** mm²/s)')
- **Claim:** Coefficient of variation / SD-to-mean ratio is sensible and unit-consistent; also check SD < mean.
 - **Checks:** simple_ratio_sanity
 - **Verdict:** PASS
 - **Notes:** Recomputed CV and checked SD < mean.
8. ✓ **C8** (Methods §2.4.3 (page 4): '25 regression models for each behavioral metric' and Results §3.3.1 (page 7): '25 brain metrics (1 Global MD, 24 Regional MDs)')
- **Claim:** Count consistency: 1 global + 24 ROI = **25**; thus per behavioral metric there are **25** regressions.
 - **Checks:** count_identity
 - **Verdict:** PASS
 - **Notes:** Checked global_count + roi_count equals total.
9. ✓ **C9** (Methods §2.2.2 (page 3): 'six distinct cognitive metrics' and Results §3.3.1 (page 7): '... and the **6** synthetic behavioral metrics')
- **Claim:** Count consistency: six behavioral metrics is used consistently across sections.
 - **Checks:** repeated_constant_match
 - **Verdict:** PASS

- **Notes:** Checked the same constant reported in two sections matches.
10. ✓ **C10** (Results §3.3.1 (page 7): 'VIF scores ... (Age: 1.07, Sex: 1.04, Origin: 1.03)')
- **Claim:** All VIFs are ≥ 1 (a definitional sanity check for typical VIF implementations).
 - **Checks:** inequality_sanity
 - **Verdict:** PASS
 - **Notes:** Checked $\min(\text{VIF}) \geq 1$.
11. ✓ **C11** (Results §3.3.2 (page 8): ' $\beta = 8609.51$, $p = 0.0052$, $R^2 = 0.226$ ')
- **Claim:** R^2 is within $[0, 1]$ and p-value within $[0, 1]$.
 - **Checks:** bounds_check
 - **Verdict:** PASS
 - **Notes:** Checked p-value and R^2 bounds in $[0, 1]$.
12. ✓ **C12** (Results §3.3.2 (page 8): ' $\beta = 5306.37$, $p = 0.0137$, $R^2 = 0.180$ ')
- **Claim:** R^2 is within $[0, 1]$ and p-value within $[0, 1]$.
 - **Checks:** bounds_check
 - **Verdict:** PASS
 - **Notes:** Checked p-value and R^2 bounds in $[0, 1]$.
13. ✓ **C13** (Results §3.3.2 (page 8): ' $\beta = -5549.51$, $p = 0.0186$, $R^2 = 0.166$ ')
- **Claim:** R^2 is within $[0, 1]$ and p-value within $[0, 1]$.
 - **Checks:** bounds_check
 - **Verdict:** PASS
 - **Notes:** Checked p-value and R^2 bounds in $[0, 1]$.
14. ✓ **C14** (Methods §2.1 (page 2) vs Results §3.1 (page 5): cohort size discrepancy)
- **Claim:** Methods states final analytical cohort is 30 bats; Results states final analytical cohort is 33 bats. Flag inconsistency.
 - **Checks:** cross_section_constant_mismatch
 - **Verdict:** PASS
 - **Notes:** PASS indicates an inconsistency was found (values differ).
15. ✓ **C15** (Abstract (page 1) vs Methods/Results: epigenetic age range inconsistency)
- **Claim:** Abstract gives epigenetic age range 6.62–13.84 years; Methods gives range 6.62–15.07 years for the 30-bat final cohort; Results gives 6.62–13.84 years for $N = 33$. Flag mismatched maxima across sections.
 - **Checks:** cross_section_range_mismatch
 - **Verdict:** PASS
 - **Notes:** PASS indicates a mismatch in reported maxima across sections.

16. \triangle **C16** (Results §3.3.2 (page 8): '25 ... brain metric residuals ... and six synthetic ... residuals' (implied 25×6 matrix))
- **Claim:** Implied number of regressions: 25 brain residuals \times 6 behavioral residuals = 150 total regressions.
 - **Checks:** implied_total_count
 - **Verdict:** UNCERTAIN
 - **Notes:** Computed implied total (brain_metrics * behavioral_metrics); no explicit reported total provided to verify.

Limitations

- Only the provided PDF text was used; no access to underlying CSV/XLSX/NIfTI data, so most computed quantities (means, SDs, regression outputs) cannot be recomputed.
- Figures are present as images; numeric values not explicitly printed in text (e.g., individual MD values, histogram bins, heatmap cells) were not extracted from pixels.
- Several statements reference 'Figure X' numbering inconsistencies (e.g., 'Figure 4. Figure 10'); resolving figure-index mapping is editorial rather than a fast numeric recomputation.