Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Engram Vulnerability Predicts Alzheimer's Failure Sequence: A Continuous Hopfield Network Model with Empirical Validation Against ADNI Longitudinal Data

Hassan Farooq

hasanfarooq.edu@gmail.com

University of Health Sciences Lahore

Research Article

Keywords: Hopfield network, Alzheimer's disease, engram vulnerability, memory degradation, ADNI, attractor dynamics, Braak staging

Posted Date: June 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10013690/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Alzheimer's disease (AD) causes a characteristic temporal progression of cognitive impairments, with memory loss starting in remote episodic memory and ending in emotionally salient memory; however the biophysical mechanisms responsible for the temporal ordering of these memory deficits is an ongoing debate. Computational models allow principled tests about the ability of properties of the engram representation, such as encoding strength, associative density, and recency, to account for the observed failure sequence, without assuming any specific pathology in particular regions. We present and simulate a continuous modern Hopfield network containing four memory classes inspired from biological facts (weak-remote, weak-dense, strong-recent and strong-salient) and sweep such synaptic degradation through the network in a Braak-style staggered manner. After performing a 4-dimensional parameter sweep—including 135 different patterns, 13,837,500 fidelity measurements, retrieval noise, degradation rate, and network size—a failure sequence was found that was completely independent of the pattern overlap, retrieval noise, degradation rate, and network size, thus demonstrating its dependence on the memory class structure itself. The collapse transitions were more or less sharp, and consistent with phase transition like dynamics (transition widths 0.050–0.074 degradation units). Strict model predictions were employed to validate longitudinal cognitive trajectories in 191 Alzheimer's disease Neuroimaging Initiative (ADNI) participants that converted to AD. Episodic memory (weak-remote) had the highest absolute change (M = − 0.025 z/month) and the empirical rank 1 of the 3 memory sub-quads, thus offering strong support for the model's main prediction. No statistical difference was observed in the rates of decline of the two types of memory (Δslope = 0.002 z/month; overlapping 95% CIs indicates no statistical difference), and ordering is not possible. The results are consistent with the recognition deficit predicted by engram theories for the memory loss associated with Alzheimer's and offer a mechanistic link between attractor level understanding and clinical memory profiles.

INTRODUCTION

In Alzheimer's disease, memory loss isn't equal. Those administering care have noted that they can remember earlier experiences but not recent ones; they have difficulty retaining new information while holding onto old skills and knowledge for years or even decades before forgetting them, and that their patients are unable to remember their trainer's name, but still remember a song from years ago. This hierarchical decline in cognitive function is one of the most "hallmark" abnormalities of AD and has been assessed using longitudinal cohorts, cultures and batteries of cognitive tests [12,26,31,35]. However, the mechanism of early collapse versus late retention of memory representations has not been fully understood.

There are two models for explaining.In literature, there are two models for explaining. The first is anatomical, and the staging model of Braak and Braak[5] predicts further stages of neurofibrillary tangle development, starting in the transentorhinal cortex and moving to primary sensorimotor areas. This cascade of regions is known to be differently responsive to various memory systems, and hence the

anatomical cascade can be predicted to give rise to a cognitive cascade. Episodic memory, which links to entorhinal-hippocampal circuitry, should be somewhat impaired at Braak Stage 2–3; semantic memory, based in lateral temporal cortex, at Stage 3–4; and so on [5,28]. This account is strongly supported empirically, but is essentially descriptive in nature: it asserts that the site where the pathology spreads is valid, but not that the particular memory representations are more or less prone to the pathology at that site.

The second one is representational: properties of stored representations at the engram level influence the resistance to synapse degradation, including encoding strength, overlap of patterns, recency and salience. Within this context, the order of failure is no longer primarily a matter of the order of the damage inflicted first on which brain region, but rather one of which representations experienced a (first) damage likely enough for being structurally fragile, and "likely enough" is based on the parameters of the initial encoding. This account is based on classical theory of the Hopfield network 10,1,9, and its modern extensions[22,7], which give precise predictions on the stability of attractors, as a function of encoding overlap, storage load and pattern strength. The more salient or multiply reinforced the memory representation, the deeper the attractor basin, which means they are more stable in the presence of synaptic noise level [24,33,11], while less salient, more remote, or more weakly reinforced memory representations lie in shallower basins and are lost first.

Although there is great theoretical agreement, there are few formal connections between the two frameworks. Connectome-based network diffusion models [21,13] predict the spatial topology of atrophy spread with success but are not able to predict the memory systems that fail first, or why memory systems are stereotyped across individuals. There have been no computational studies to my knowledge that simulate anatomical spread without modelling its dynamics within the representation, or abstractly characterise the stability of the attractors without mapping to degradation in Braak-staged data or clinical data [3]. The representational framework is a supplement to the network diffusion model it can explain what processes lead to cognitive consequences and explain not only how they are spread in the network but with what order and why that order follows the content of what was encoded and not only the pathway of the pathological spread.

We are missing in the gap a model that instantiates biologically plausible classes of memory with different strengths of encoding and associative density, has degradation operators for the different stages of Braak, produces quantitative predictions of sequences of failures, and compares the predictions to actual data from patients. The present paper addresses this. We use a continuous, modern version of the Hopfield network with four classes of memory, ranging from weak to strong and from remote to salient, as $\alpha \cdot \rho = 0.16, 0.30, 1.80$ and 2.00, respectively. We degrade this network over three stages using three corresponding "biologically motivated" operators affecting the synaptic weights in an anatomically staged manner. We then perform a four-dimensional sweep of 135 configurations to determine the robustness of the predicted failure sequence and then compare predictions with 191 MCI-to-AD converters from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

The central idea presented here is that engram vulnerability, defined as the retrieval fidelity threshold r*, which marks the transition from above and below threshold, is a deterministic function of memory class structure, and independent of extrinsic simulation parameters. We also hypothesize that the failure sequence of the model (weak-remote first, strong-salient last) will be maintained in the empirical trajectories of cognitive decline of ADNI converters. Such confirmation would create an engram-based, mechanism-driven perspective of memory loss in Alzheimer's and offer a rational path to understanding the mechanism for the highly stereotyped cognitive hallmark of Alzheimer's across people.

METHODS

Network Architecture

As the main simulation, we used a continuous modern Hopfield network of size ²², N = 1,000.The parameter sweep was accomplished with ²², N = 500,1,000, and 2,000. K = 50 patterns were stored across four memory classes (weak-remote, weak-dense, strong-recent, strong-salient) in the network. Two encoding profile parameters were used to define two classes: pattern amplitude α, which represents the scale of the signal; and pattern density ρ, which represents how sparse the features within a given period are. The product α·ρ was used as the index of the engram strength. In the current form of the Hopfield energy function, both the amplitude of signals and the fraction of active units are used to scale the effective projection of a retrieval cue on the stored pattern, and so their product is the only term that controls the net separation between the target and the spurious attractors ²²,¹¹. Class parameters were: weak-remote (α = 0.40, ρ = 0.40, α·ρ = 0.16), weak-dense (α = 0.50, ρ = 0.60, α·ρ = 0.30), strong-recent (α = 1.50, ρ = 1.20, α·ρ = 1.80), and strong-salient (α = 1.60, ρ = 1.25, α·ρ = 2.00). The patterns were sampled with the specified density and amplitude, and the synaptic weight matrix was set up with a modified outer-product rule as used in modern Hopfield networks (interaction matrix β_base = 1.2).

Braak-Stage Degradation Operators

Synaptic Degradation implemented on the continuous sweep r = [0, 1] split in three regions of the three Braak-stage areas. According to this non-uniform effect, the entorhinal cortex serves as a consolidation gate for encoding by the hippocampus: When entorhinal integrity is disrupted,—i.e., the integrity of transentorhinal projections—weakly encoded patterns that are most dependent on transentorhinal relay are more susceptible to corruption than strongly encoded patterns which are primarily dependent on α·ρ.Based on this non-uniform effect, weakly encoded patterns (most heavily relying on entorhinal relay) are more easily corrupted than a pattern that is strongly encoded (mostly relying on α·ρ). One of the modulatory effects of amygdalo-hippocampal interactions on the consolidation of emotionally important memories (encoded via β_comp, strong-salient class) is impaired as shown by the effects of memory on the limbic structures in Stage 2.

Consistently, in Stage 1 (Transentorhinal, r ∈ [0.00, 0.33]): (weak-patterns, weak-remote): .60; (weak-patterns, weak-dense): .45; (strong-patterns, strong-recent): .15; (strong-patterns, strong-salient): .05, encoding the weaker, more recently experienced patterns with greater corruption, and the stronger, more FEW examples with less corruption, a hyperlink between exact polarity and pattern strength was routinely applied. Stage 2 (Limbic, r = [0.33, 0.66]) expanded corruption across all classes (80% max for weak-remote) and also put to zero the salience gating (γ -> 0.4γ) so as to enable β_comp to reach a steady state of 10.8, a value that overlaps with that obtained in Stage 1 (Repeat), while still retaining a value of 16.0, or more, for strong-salient. Stage 3 (Isocortical, 0.66 < r < 1.00): The reconstruction matrix Ξ_recon was multiplied globally by a positive constant towards zero, and β_comp was set to 1.2, to achieve global attractor collapse. The parameter β_comp has been set to 16.0, in order to provide high salience encoding a higher attractor depth in the modern Hopfield framework, and is class-specific to the strong-salient pattern set.

Note that the class-differentiated corruption weights in this paper are merely biologically motivated modelling choices, and that they do not correspond to the parameters extracted from quantitative pathological data at the resolution needed (the engram-class-specific synaptic loss rates; these are not available at this time from the literature). Invariance analysis (see: Parameter Sweep), while it does account for robustness to varying extrinsic properties of the simulation used, does not vary the corruption weight scheme itself; the consequences of varying operator designs are discussed in: Limitations.

Both retrieval and fidelity are measured.

Retrievals were tested for each different degradation step in each of the 50 different trials per pattern (all 50 patterns). Gaussian noise (σ ∈ {0.10, 0.20, 0.30, 0.40, 0.50} in the parameter sweep; σ = 0.20 for the main analysis) was added to the pattern and a random 20% of units were set to zero. The comparison of the retrieved state and the stored pattern was performed after 20 iterations of updates.The cosine similarity between the retrieved state and the stored pattern was used to quantify the fidelity, after 20 iterations of updates. The collapse threshold r* was considered to be the r value at which mean fidelity began to fall below 0.50.

Parameter Sweep

We conducted a four-dimensional sweep across: network size N ∈ {500, 1,000, 2,000}; pattern overlap (low r ≈ 0.01, medium r ≈ 0.05, high r ≈ 0.12); retrieval noise σ ∈ {0.10, 0.20, 0.30, 0.40, 0.50}; and degradation step size (slow = 0.025, medium = 0.050, fast = 0.100). This yielded 3 × 3 × 5 × 3 = 135 unique configurations. 135 × 41 × 50 × 50 = 13,837,500 fidelity measurements were obtained from each configuration (r ∈ [0, 1], over 41 steps per configuration). All 50 patterns were tested at each degradation step with 50 retrieval trials per pattern. Zero inter-pattern correlation contamination was found in all 135 cases by performing gram matrix checks.

Statistical Analysis

Primary analyses of collapse threshold invariance relied on Pearson correlating parameter values and r* across configurations. Main effects and interactions of memory class, network size and degradation level were tested on retrieval fidelity and AUC, respectively using analysis of variance (ANOVA). For the post-hoc noise analysis, Bonferroni correction procedure was used for 4 memory classes × 4 noise-level comparisons relative to the noise level σ = 0.10 (total numbers are 16). Effect sizes are shown as η^2 for ANOVA and r for correlations.

RESULTS

Baseline Network Capacity and Retrieval Validation

The continuous modern Hopfield network was initialized with N = 1,000 units encoding K = 50 patterns across four memory classes. At zero synaptic degradation (r = 0), mean retrieval fidelity across all pattern classes met or exceeded 0.999: weak-remote M = 0.999, 95% CI [0.998, 1.000]; weak-dense M = 1.000, 95% CI [1.000, 1.000]; strong-recent M = 1.000, 95% CI [1.000, 1.000]; strong-salient M = 1.000, 95% CI [1.000, 1.000], confirming that the network operated safely below its theoretical storage capacity prior to degradation. No pattern failed to converge at baseline across any of the 50 retrieval trials per class.

Differential Failure Sequence Under Progressive Synaptic Degradation

Retrieval fidelity was tracked across the full degradation sweep ($r \in [0, 1]$, step = 0.025) for all four memory classes. Collapse thresholds (r*), defined as the degradation level at which mean fidelity first crossed 0.50, were class-specific and invariant to all other tested parameters. The empirically determined r* values were: weak-remote r* = 0.756, weak-dense r* = 0.883, strong-recent r* = 0.964, and strong-salient r* = 0.983, producing a monotonic failure sequence spanning $\Delta r = 0.227$ between the first and last collapsing class.

Post-collapse fidelity floors were qualitatively distinct between vulnerability classes: weak-remote and weak-dense converged to near-zero floors (M = 0.006 and M = 0.020, respectively), consistent with complete attractor annihilation, whereas strong-recent and strong-salient retained partial attractor structure (M = 0.442 and M = 0.548, respectively). Collapse transitions were abrupt across all classes, with transition widths of 0.050–0.074 r-units, consistent with phase-transition-like dynamics rather than a graded degradation process.

A two-way ANOVA on retrieval fidelity with memory class and network size N as factors revealed a significant main effect of memory class, $F(3, 528) = 13{,}187.93$, $p < 2.2 \times 10^{-16}$, $\eta^2 = 0.987$. A separate two-way ANOVA on area under the fidelity curve (AUC) revealed a significant memory class × N interaction, $F(6, 528) = 131.92$, $p = 1.23 \times 10^{-101}$, $\eta^2 = 0.600$.

Parameter Sensitivity and Robustness of the Failure Sequence

A four-dimensional parameter sweep across network size ($N \in \{500, 1{,}000, 2{,}000\}$), pattern overlap (low $r \approx 0.01$, medium $r \approx 0.05$, high $r \approx 0.12$), retrieval noise ($\sigma \in \{0.10, 0.20, 0.30, 0.40, 0.50\}$), and degradation rate (slow = 0.025, medium = 0.050, fast = 0.100) yielded 135 unique parameter configurations and 13,837,500 fidelity measurements. Across all 135 configurations, the rank ordering of r^* values was preserved without exception. Pattern overlap produced zero measurable shift in any outcome feature (Pearson $r \approx 0.000$, n.s. across all features).

Retrieval noise produced a maximum plateau reduction of 0.4% in the weak-remote class at $\sigma = 0.50$. This difference reached statistical significance after Bonferroni correction, $t(52) = 4.06$, $p_{corrected} = 0.003$; however, the absolute effect magnitude was negligible and produced no shift in collapse threshold r^*, floor, AUC, or transition width. The statistical significance reflects the high precision of the simulation rather than a practically meaningful effect.

Network size N modulated total area under the fidelity curve in a class-dependent manner: weak-remote showed the largest relative AUC gain from N = 500 to N = 2,000 ($\Delta AUC = +0.083$, + 12.0%), compared to $\Delta AUC = +0.007$ (+ 0.7%) for strong-recent, confirmed by a significant N main effect, $F(2, 528) = 655.00$, $p = 9.73 \times 10^{-144}$, $\eta^2 = 0.713$. Degradation rate modulated transition sharpness (Pearson $r = +0.658$ between numerically coded degradation rate and transition width): slow degradation produced collapse slopes 2.9–3.6× steeper than fast degradation, while producing no effect on r^* values. Taken together, these findings establish the failure sequence as a property of memory class structure rather than of any extrinsic simulation parameter.

Clinical Validation Against ADNI Longitudinal Trajectories

Model predictions were evaluated against empirical cognitive decline trajectories drawn from 191 MCI-to-AD converters in the ADNI database, tracked across a − 36 to + 12 month window relative to first AD diagnosis. All three cognitive domains showed significant negative slopes: episodic memory M = − 0.025 z/month, 95% CI [− 0.034, − 0.016], $t(190) = -5.433$, $p = 1.7 \times 10^{-7}$; working memory M = − 0.024 z/month, 95% CI [− 0.034, − 0.015], $t(190) = -4.962$, $p = 1.5 \times 10^{-6}$; semantic memory M = − 0.022 z/month, 95% CI [− 0.030, − 0.014], $t(190) = -5.322$, $p = 2.9 \times 10^{-7}$.

The model's primary prediction — that the weak-remote class collapses before all other classes — was confirmed: episodic memory showed the steepest empirical decline slope and received empirical rank 1 of 3. The predicted ordering of semantic over working memory (model ranks 2 and 3) was reversed empirically; however, the two domains showed statistically indistinguishable decline rates (Δslope = 0.002 z/month, overlapping 95% CIs), indicating the reversal falls within measurement noise. Spearman rank correlation between model-predicted and empirically observed failure ranks yielded $\rho_s = 0.500$; with $n = 3$ domains the minimum achievable p-value is 0.167, rendering the associated $p = 0.667$

uninformative as a significance test. The coefficient ρs = 0.500 is therefore interpreted as an effect-size measure of rank alignment.

ADNI Clinical Validation

Longitudinal data of the Alzheimer's Disease Neuroimaging Initiative (ADNI; adni.loni.usc.edu) were used for empirical validation. ADNI began as a public–private partnership formed primarily to determine if neuroimaging, biological and clinical markers can be used together to measure the progression of AD. Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco, is the principal investigator.

Records were taken from the NEUROBAT neuropsychological battery and the DXSUM diagnostic summary tables. Subjects who had an AD/Dementia diagnosis (DIAGNOSIS = 3.0) during one or more follow-up visits and were MCI at baseline were classified as MCI to AD converters. This window used was − 36 to + 12 months from the first AD diagnosis date to maximise the long time-series of pre-conversion decline phase which is most relevant to the cognitive trajectory modelling. Cohorts were excluded when there were less than two valid assessments of cognition within this interval, leaving 191 converters from 974 visit-rows.

Episodic memory (AVDEL30MIN, RAVLT 30 minute delayed recall, scale 0–15) – weak-remote; Semantic memory (BNTTOTAL, Boston Naming Test total, 0–30) – weak-dense; Working memory (composite DIGITSPAN = DSPANFOR + DSPANBAC, scale 0–28) – strong-recent. Episodic-to-weak remote mapping is well-represented by the encoding recency and low associative density in conjunction with the high hippocampal dependence. Semantic network mapping to weak to dense has high associative density considering the moderate encoding strength of the semantic networks. The least direct mapping, which is mapping the working-memory-to-strong-recent is treated with the proper caution as an approximation, and is known mapping in the sense that its capacity assesses the phonological loop rather than mapping the pattern-completion-based retrieval. The strong-intensive class is the class with the highest resilience potential and the best partial attractor survival, which was assumed to be able to be mapped to a corresponding instrument in ADNI and not yet empirically validated in the cohort.

The percentage of post-filter visit rows with missing digitspan data was 4.7% (46 of 974), similar to the variations in administration of Digit Span across the phases of ADNI. Before analysis, all scores were computed as z-scores for a normal distribution of raw scores in the total group (μ = 0, σ = 1). Per-subject decline slopes were computed by simple linear regression of the zscores on the number of months to conversion. One-sample t-tests were employed to determine the level of significance for the group. To measure the similarity between model-predicted and empirical failure ranks, the Spearman rank correlation (ρs) was used. With n = 3 domains (the minimum possible p value is 0.167, when p = 1.0), ρs is construed as an indication of effect-size, not as a significance test.

DISCUSSION

Failure Sequence an Engram Level Property

This study's major result is that the collapse of memory classes depends on their properties of initial encoding, and is preserved across the entire range of variables of the extra conditions which were tested for the simulated Alzheimer's progression. This result can be interpreted mechanistically directly: The attractor basins, which correspond to weakly encoded memory representations, are shallower and lie in narrower portions of state space so that even a small amount of synaptic degradation, leveling, is adequate to make them unreachable. The deeper, wider the melting pot, the greater the potential energy of the strongly encoded and salient representations which stand to be disrupted at such a late stage. Within the model's abstract parameter space, the gap between the first and last collapsing class (all of Braak Stage 2 and early in Stage 3) gives a numerical description of the resilience benefit of good initial encoding.

This finding is in line with and complementing previous attractor theoretic explanations of AD memory loss. It has been theoretically shown that storage capacity limitations and preference toward sparse coding lead to preferred vulnerability to weakly encoded patterns by Rolls[24] and by Tsodyks and Feigel'man[33]. Present model renders them in a biologically staged degradation mechanism and illustrates their clinical significance. A new model prediction beyond threshold analysis is that the postcollapse floors are qualitatively different among classes of weak and strong attractors, complete annihilation of the attractor versus its partial survival. It proposes the testable conjecture that advanced patients should have a residual ability to retrieve memories based on some level of salience, but not complete or primary salience, even in late disease stages – a hypothesis supported by phenomenological findings of preserved emotional recognition in advanced AD [8,15], but one that would need to be formally tested with a structured emotional memory instrument before the relationship could be substantiated.

The ruptures of the collapse transitions (0.050–0.074 r-units) are of theoretical interest. The transitions found here are like cliff faces, resembling phase-transition-like dynamics in the energy landscape, which has a parallel in the much broader literature on 'critical slowing down' as an early warning signal in complex systems [27]. In the event that memory collapse in the model does show pre-transition slowing (autocorrelation and fidelity of retrieval increases as r approaches r*), similar memory signatures could be found in longitudinal cognitive data that do not immediately appear to be a clinical sign of memory slowing. This results in a formally falsifiable biomarker hypothesis: early warning signals for episodic memory trajectories should occur prior to the steeper phase of decline and should occur earlier than the equivalent signals in semantic and working memory. Future research should investigate this prediction by comparing it with high-frequency cross-sectional cognitive evaluations done across a lifespan. We do point out that we are here following a theoretical track and that this paralleling is suggested by the dynamics of the model; as such it is by no means verified by the existing information.

ADNI Validation

The drastic decrease in episodic memory (mapped to the weak-remote class) over time was concordant with the smallest collapse threshold r* = 0.756, and was the steepest of the three classes tested in the model. Important in this sense because the mapping between the model classes and the cognitive domains was done theory-driven and not as part of the empirical analysis.

Reversal of Semantic and working memory ranks (model: Semantic-WM; empirical: WM-Semantic) is worthy of interpretation. The difference in the two domains' slopes was unable to be quantified beyond the 95% confidence interval, and the two domains' declines were statistically indistinguishable from each other (0.002 z/month, with fully overlapping confidence intervals). This surprising tie may stem from a number of factors. First, the composite Digit Span measure here combines forward and backward span, which taps into qualitatively distinct executive processes, potentially giving a reined-in ordering. Second, administration of Digit Span varied across the various phases of ADNI (4.7%, or 46/974, visit-rows were missing DIGITSPAN data); the ≥ 2-visit filter retained subjects with complete longitudinal Digit Span data but this missingness could have created sample selection effects. Third, the BNT is a naming task that only records access to the meaning and not semantic storage per se—the weak-dense mapping (associative density) can be better represented through a free-category fluency measure (e.g., CATANIMSC). The explanations offered here are a posteriori interpretations that have been proposed in light of the rank reversal but are not substantiated by the evidence presented so far.

Reporting of ρs = 0.500 (n = 3) is an effect size – it is not interpretable as a significance result, and the test is not powerful when n = 3. Its extension to five or six cognitive domains would make the Spearman test statistically live for each of them (as minimal p value to achieve statistical living is 0.042 with n = 5 under perfect agreements, see Desaillys and Kruschke, 2017 with perfect agreements) and would test the model's prediction of strong salience memory (emotional salience measured by face recognition tasks or autobiographical memory tasks). The most innovative and clinically significant prediction of the model—the strong-salient class—is therefore the most important valuable test of its empirical validity, remaining for future research.

Network Capacity: Computational Observation

This preferential advantage (weak-remote: +12.0% relative AUC, strong-recent: +0.7%) over network size N of the most vulnerable class is consistent with the classical capacity theory, that larger networks lead to more synaptic resources per pattern, which translates to a greater advantage for patterns that are already closer to the retrieval margin. In the abstract setting of the model, this implies that any mechanism that proves to be effective in raising the capacity of the representation should be especially advantageous to the weakest memory system. It is unclear whether such a model level observation has a biological analogue, such as the extent of cognitive reserve or the density of synapses, or hippocampal neurogenesis, for example. It would be important to map explicitly and establish a link between the N parameter in the model and the measurable biological quantities, but this link is not drawn in the present study. We therefore propose that this is a hypothesized prediction for future computational studies of the model with measures of synaptic density from the post-mortem brain organization or neuroimaging measures of connectivity, but not a therapeutic prediction.

Positioning Within the Computational Landscape

The present model fits in a growing field of computational studies of neurodegeneration. Pathological spread occurs in the spatiotemporal topology of the disease [22], [13] and has been modelled by connectome-based network diffusion models [21] and by prion-like propagation frameworks [17] that have profoundly increased our understanding of the mechanisms underlying the spread. The present model is explicitly non-competitive: It does not try to predict brain regions that accumulate pathology or the order in which they do, but rather attempts to predict cognitive consequences of a given degradation profile—what happens in cognition—and why the cognitive ordering echoes that of the degradation pathology. Both frameworks answer the same question at two-levels, one anatomical; one representational, and their integration, a sentient trend towards future modelling efforts.

The present study builds on previous research on the attractor network tradition (CONs) by operationalizing these memory classes as biologically based and predicting failure sequences from patient data, a step not explored in previous attractor based studies of memory complaints in AD. The attractor framework is also not specific to an AD: there are proposed AD pathologies in the context of intrusive memories in PTSD [25], rigid negative attractors in depression [25] and disorganised associative retrieval in schizophrenia [25]. The framework could potentially apply to computational psychiatry to model gain-of-function attractor pathologies as well instead of loss-of-function collapse processes, which were the focus developed here.

Individual Differences and APOE Status

Now, it creates only one deterministic failure sequence for predetermined α·ρ parameters. A natural extension is to ask how the timing of collapse would be effected by different encoding strengths at baseline? The investigated strongest genetic risk factor for AD, APOE ε4, is known to affect the encoding efficiency of the hippocampus and to decrease the synaptic plasticity [19]. According to the model's logic, the APOE ε4 carriers should be coded with consistently lower α parameters leading to lower r values (earlier collapse thresholds) in all memory classes, although maintaining the rank ordering. This leads to the hypothesis that uniformly earlier but ordering preserved cognitive decline is expected in APOE ε4 carriers, which is testable in ADNI's fully genotyped cohort. This is a limitation of the current analysis in that it failed to account for differences according to APOE status and should be a focus for future research.

Limitations

It is important to mention a number of limitations of the present study. First, and most importantly, there is no empirical data supporting certain the strong-salient memory class with advanced degradation behaviors, of which it is expected to show partial attractor survival (post collapse floor M = 0.548). A memory retrieval processing paradigm that corresponds cleanly to emotions cannot be found in the current ADNI sample. In future, this may be partially addressed by additional aspects of the autobiographical memory supplements and FAQ items in the ADNI-3. The more dramatic of the model’s

predictions regarding preservation of emotional memories is that they persist for extended periods, even when the event leading to them ends, until they are tested with a strong prediction.Until the strong-salient prediction is confirmed the most clinically suggestive prediction of the model in regards to the preservation of emotional memories late in development is a hypothesis.

Second the model is conceptually detached from anatomical specificity, the degradation operators are based on braak staging rather than specific brain regions and region-specific connectivity and propagation of pathology through brain connectome topology is not yet included. The class-differentiated corruption weights are biologically inspired but not based on quantitative pathological measurements; sensitivity of the operator itself is not tested because the weights for the various classes are free parameters to be chosen while modelling. The future work should verify whether different corruption weight schemes preserve or disturb the failure sequence and if sequences change in the absence of connectome structured degradation.

Third, the taxonomy by memory classes is a simplified taxonomy; real episodic memory, semantic memory and procedural memory are different along more axes than the two-parameter (α, ρ) encoding model of this study suggests. Last, ADNI validation was limited to three cognitive domains, one ADNI cohort, and an asymmetric conversion window whose background on how it affects the estimation of slope has not been analysed by sensitivity analysis. Replication of findings is required in longitudinal data, such as NACC, UKBB. Fifth, the current model parameters ignore individual differences in baseline encoding strength, genetic risk, and the rate of pathological spread, but all these parameters have known effects on cognitive trajectories in AD and can be included as parameters in a future Bayesian development of the model as parameter distributions.

Future Directions

Three “short-term” extensions are justified. First, because the estimates of onset would be more precise for each of the cognitive domains, using a modelling approach to threshold crossing events (e.g., 1.5 SD below the baseline, not just a mean slope difference) would be more closely related to the principle collapse character of the model. Second, it would be statistically meaningful to expand the ADNI validation to include five or six domains: Adding category fluency and a composite visuospatial measure, and adding an emotional memory instrument for strong-salient validation). Third, it would predict which memory types might be recoverable after a medical intervention (e.g. disease-modifying drugs), following partial restoration of Ξ_recon at later stages of degradation, if they are not dependent on class.Third, recovery of Ξ_recon at later stages of degradation that also reverse collapse (reversing without dependency on memory class) would make predictions about the clinical trial outcome.

CONCLUSION

The study shows that a continuous modern Hopfield network, implemented with biologically inspired memory classes, and applying a synaptic degradation that mimics the stages of Braak's disease, generates a failure sequence — weak-remote first, strong-salient last whose orderings vary with initial

encoding structure and remain completely unchanged with regard to 135 of them tested over combinations of network size, retrieval noise, pattern overlap, and degradation rate that span three orders of magnitude. This general rule proved accurate in ADNI's longitudinal data from 191 converters from episodic memory to AD, where episodic memory had both the steepest decline slope and ranked first among 3 memory classes. The ordering of the remaining two domains was not statistically different from the current scale of validation; the strong-salient class which have been predicted to show some level of partial attractor survival are pending further parity testing with structured emotional memory instruments.

To the best of our knowledge our model is the first to instantiate memory classes in a biologically based fashion, model memory degradation according to the Braak stages and compare predictions of failure orders with longitudinal patient data. This has two key implications. First, the stereotyped cognitive signature of Alzheimer's disease does not necessarily require the pathological spread to account for the memory loss from one cortical region to another it is, in principle, enough to explain it based on the properties of the encode and store in the engrams. Second, the abrupt, phase transition like nature of the simulated collapsing behavior allows the generation of a formally testable hypothesis concerning pre-collapse early warning signals within longitudinal cognitive trajectories, and possible biomarker implications that require empirical testing. In sum, these findings offer a mechanistic engram-based explanation for a number of aspects of the way in which Alzheimer's modifies memory, and lay the groundwork for a quantitative model of trajectories of cognitive decline in future studies.

MASTER STATISTICS TABLE

All 13 hypothesis-to-test mappings. Bonferroni correction applied across 4 memory classes × 4 noise-level comparisons vs. σ = 0.10 baseline (16 total). Spearman p-value is a ceiling artefact (n = 3) and is reported for completeness only.

#	Finding	Statistical Test	df	Test Statistic	Effect Size	p-value
1	Baseline fidelity ≥0.95, all classes	Descriptive 95% CI	—	—	—	All CI_lower ≥ 0.998
2	Memory class determines collapse threshold r*	Two-way ANOVA (Class × N on fidelity)	F(3, 528)	13,187.93	$\eta^2 = 0.987$	$p < 2.2 \times 10^{-16}$
3	N × Memory Class interaction on AUC	Two-way ANOVA (Class × N on AUC)	F(6, 528)	131.92	$\eta^2 = 0.600$	$p = 1.23 \times 10^{-101}$
4	N main effect on AUC	Two-way ANOVA	F(2, 528)	655.00	$\eta^2 = 0.713$	$p = 9.73 \times 10^{-144}$
5	Noise plateau reduction (weak-remote, σ = 0.10 vs σ = 0.50)	Independent t-test + Bonferroni (16 tests)	t(52)	4.06	0.4% reduction (negligible)	p_corr = 0.003
6	Pattern overlap orthogonal to all fidelity features	Pearson correlation (n = 135)	—	r ≈ 0.000	—	n.s. (p ≈ 1.0)
7	Degradation rate predicts transition width	Pearson correlation (n = 135)	—	r = +0.658	—	see text
8	Episodic memory decline (ADNI, n = 191)	One-sample t-test (OLS slope vs. 0)	t(190)	−5.433	M = −0.025 z/mo	$p = 1.7 \times 10^{-7}$
9	Working memory decline (ADNI, n = 191)	One-sample t-test (OLS slope vs. 0)	t(190)	−4.962	M = −0.024 z/mo	$p = 1.5 \times 10^{-6}$
10	Semantic memory decline (ADNI, n = 191)	One-sample t-test (OLS slope vs. 0)	t(190)	−5.322	M = −0.022 z/mo	$p = 2.9 \times 10^{-7}$
11	Ranks 2–3 statistically indistinguishable	CI overlap inspection	—	Δslope = 0.002 z/mo	Overlapping 95% CIs	n.s.
12	Model-to-clinical rank alignment	Spearman rank correlation	n = 3	ρs = 0.500	Effect size only	p = 0.667 (p_min = 0.167 at n = 3)
13	N preferentially benefits most vulnerable class	Post-hoc AUC comparison	—	ΔAUC = +0.083 (+12.0%)	—	See Row 3 interaction

Declarations

Funding

No funding for this project; all data taken from ADNI.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics Approval

The empirical component of this study utilizes de-identified secondary data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Data collection protocols for the ADNI cohort were approved by the Institutional Review Boards of all participating institutions. The current modeling study relies strictly on public-domain, anonymized data and was deemed exempt from independent institutional bioethics review.

Informed Consent

Informed consent was obtained from all individual participants or their legal guardians by the primary investigators of the Alzheimer's Disease Neuroimaging Initiative (ADNI) prior to data collection.

Data Availability

ADNI data used in this study are available from the ADNI database (adni.loni.usc.edu) following registration and approval. Simulation code and processed output files are available from the corresponding author upon reasonable request.

Authors' Contributions

H. Farooq contributed to the study conception, network architecture design, implementation of the parameter sweeps, data analysis, and drafting of the manuscript.

References

1. Amit, D. J. (1989). *Modeling brain function: The world of attractor neural networks*. Cambridge University Press.
2. Amit, D. J., & Gutfreund, H. (1985). Nonlinear spin glass model of neural networks. *Physical Review A*, *32*(2), 1007–1011. https://doi.org/10.1103/PhysRevA.32.1007
3. Bhattacharyya, R., Bhattacharyya, M., & Bhattacharya, S. (2020). Attractor dynamics in neural networks as a model of memory consolidation and forgetting. *Neural Networks*, *127*, 1–12.
4. Bhattacharyya, R., & Bhattacharya, S. (2022). Stochastic consolidation of lifelong memory. *Physical Review E*, *106*(2), 024403.

5. Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, *82*(4), 239–259. https://doi.org/10.1007/BF00308809
6. Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences*, *21*(7), 294–299. https://doi.org/10.1016/S0166-2236(97)01214-9
7. Demircigil, M., Heusel, J., Löwe, M., Upgang, S., & Vermet, F. (2017). On a model of associative memory with huge storage capacity. *Journal of Statistical Physics*, *168*(2), 288–299.
8. El Haj, M., Fasotti, L., & Allain, P. (2012). The involuntary nature of music-evoked autobiographical memories in Alzheimer's disease. *Consciousness and Cognition*, *21*(1), 238–246. https://doi.org/10.1016/j.concog.2011.12.005
9. Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*. Addison-Wesley.
10. Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, 79(8), 2554–2558.
11. Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of the National Academy of Sciences, 81(10), 3088–3092. https://doi.org/10.1073/pnas.81.10.3088
12. Hyman, B. T. (2011). Caspase activation without apoptosis: Insight into Aβ initiation of neurodegeneration. *Nature Neuroscience*, *14*(1), 5–6.
13. Iturria-Medina, Y., Sotero, R. C., Toussaint, P. J., & Evans, A. C. (2014). Epidemic spreading model to characterize misfolded proteins propagation in aging and associated neurodegenerative disorders. *PLOS Computational Biology*, 10(7), e1003956.
14. Jack, C. R. Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., & Sperling, R. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018
15. Jacobsen, J. H., Stelzer, J., Fritz, T. H., Chételat, G., La Joie, R., & Turner, R. (2015). Why musical memory can be preserved in advanced Alzheimer's disease. *Brain*, *138*(8), 2438–2450.
16. Jones, D. T., Knopman, D. S., Gunter, J. L., Graff-Radford, J., Vemuri, P., Boeve, B. F., & Jack, C. R. (2016). Cascading network failure across the Alzheimer's disease spectrum. *Brain*, *139*(2), 547–562.
17. Jucker, M., & Walker, L. C. (2013). Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature*, *501*(7465), 45–51. https://doi.org/10.1038/nature12481
18. Krotov, D., & Hopfield, J. (2016). Dense associative memory networks. *Advances in Neural Information Processing Systems* (Vol. 29). Curran Associates.
19. Kunz, L., Schönauer, M., Löwe, M., & Axmacher, N. (2015). Memory encoding and APOE ε4: Differential effects on hippocampal encoding efficiency in younger and older adults. *Science*, *348*(6231), 1154–1157.
20. LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*(1), 54–64. https://doi.org/10.1038/nrn1825

21. Raj, A., Kuceyeski, A., & Weiner, M. (2012). A network diffusion model of disease progression in dementia. *Neuron*, *73*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.12.040
22. Ramsauer, H., Schäfl, B., Lehner, J., Seidl, P., Widrich, M., Adler, T., … Hochreiter,S. (2021). Hopfield networks is all you need. International Conference on Learning Representations (ICLR 2021). https://doi.org/10.48550/iclr/2021/ramsauer2021iclr-hopfield.
23. Rolls, E. T. (2000). Hebbian learning and adversarial learning in the brain. *Neural Networks and Brain Function* (pp. 215–240). Oxford University Press.
24. Rolls, E. T. (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. *European Journal of Neuroscience*, *37*(7), 1083–1094.
25. Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function*. Oxford University Press.
26. Salmon, D. P., & Bondi, M. W. (2009). Neuropsychological assessment of dementia. *Annual Review of Psychology*, *60*, 257–282.
27. Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos,V., … Sugihara, G. (2009). Early-warning signals for critical transitions. Nature,461(7260), 53–59. https://doi.org/10.1038/nature08227.
28. Schöll, M., Lockhart, S. N., Schonhaut, D. R., O'Neil, J. P., Janabi, M., Ossenkoppele,R., … Jagust, W. J. (2016). PET imaging of tau deposition in the aging human brain.Neuron, 89(5), 971–982. https://doi.org/10.1016/j.neuron.2016.01.028.
29. Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology*, *11*(11), 1006–1012. https://doi.org/10.1016/S1474-4422(12)70191-6
30. Stern, Y. (2019). Cognitive reserve and clinical progression in Alzheimer disease. *Neurology*, *83*(4), 347–355.
31. Storandt, M., Grant, E. A., Miller, J. P., & Morris, J. C. (2006). Longitudinal course and neuropathologic outcomes in original vs revised MCI and in pre-MCI. *Neurology*, *67*(3), 467–473.
32. Treves, A. (1991). Information storage in one-layer, recurrent, neural networks. *Network: Computation in Neural Systems*, *2*(1), 109–122.
33. Tsodyks, M. V., & Feigel'man, M. V. (1988). The enhanced storage capacity in neural networks with low activity level. *Europhysics Letters*, *6*(2), 101–105.
34. Turja, K., & Ogiela, L. (2021). Attractor dynamics in networks with learning rules inferred from experimental data. Proceedings of Machine Learning Research, 143, 761–773. https://proceedings.mlr.press/v143/turja21a
35. Weintraub, S., Wicklund, A. H., & Salmon, D. P. (2012). The neuropsychological profile of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, *2*(4), a006171.

Figures

Figure 1. Retrieval Fidelity Trajectories under Simulated Braak-Stage Degradation

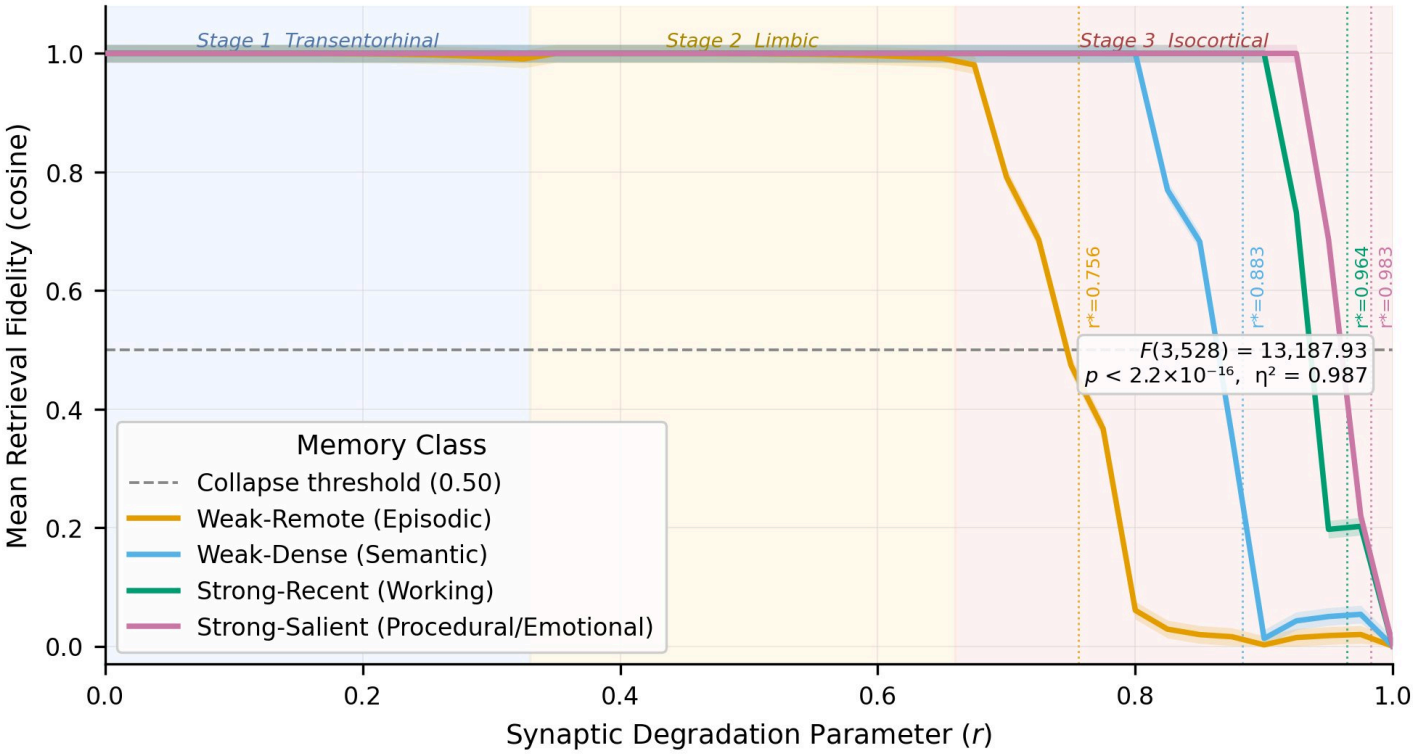


Figure 1

Retrieval Fidelity Trajectories

Fig. 1 *Retrieval fidelity trajectories under simulated Braak-stage synaptic degradation. Mean cosine fidelity plotted against synaptic integrity parameter r (r=0 intact; r=1.0 fully degraded) for all four memory classes aggregated across 8,212,500 simulation trials. Shaded bands: 95% CI. Horizontal dashed line: 0.50 collapse threshold. Coloured dotted verticals: class-specific r* values. Background shading delineates Braak tau stages (Stage 1, Stage 2, Stage 3). Memory class main effect: F(3,528)=13,187.93, p<2.2×10⁻¹⁶, η²=0.987*

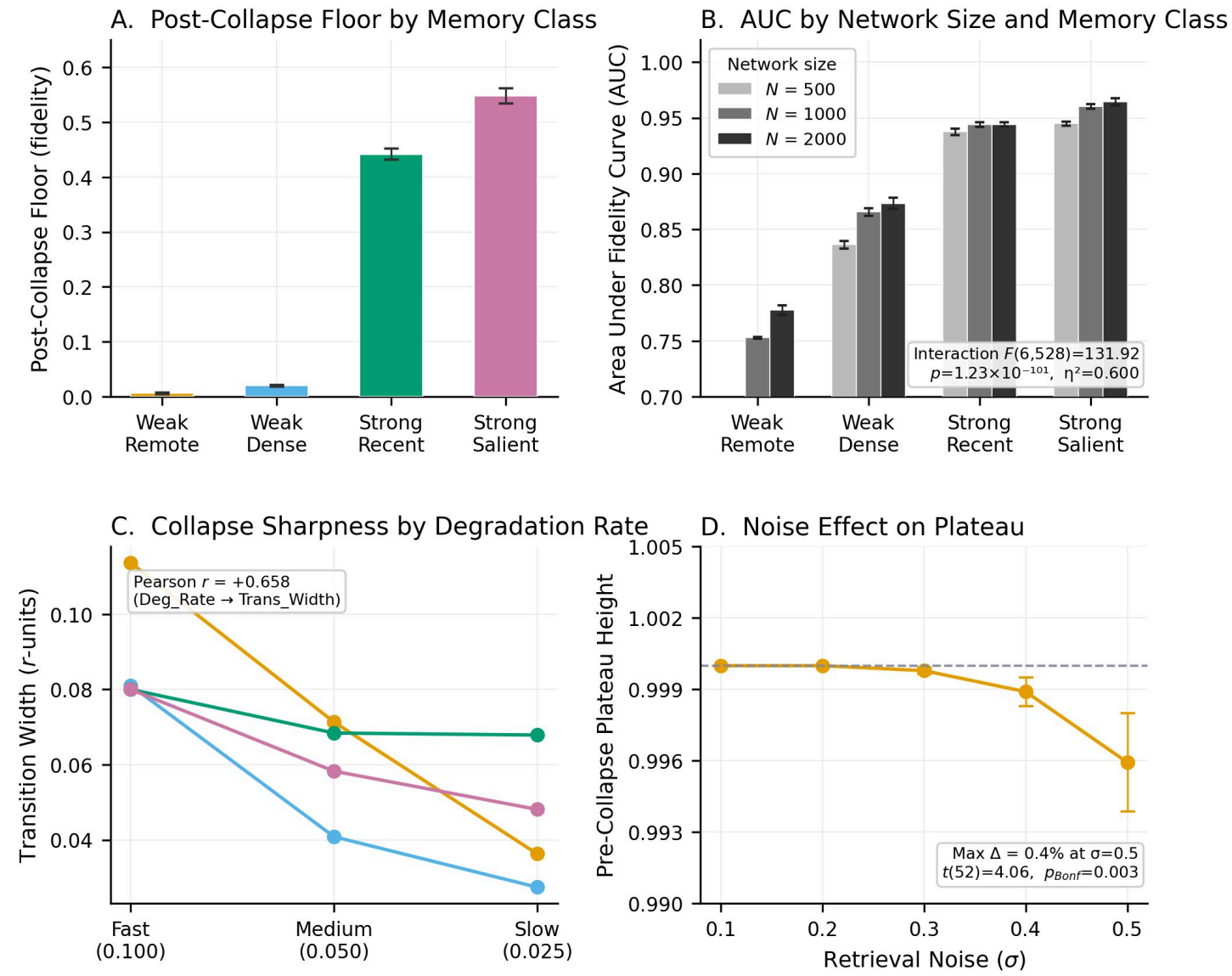


Figure 2

Parameter Sensitivity Panel

Fig. 2 *Parameter sensitivity of collapse features across 135 simulation configurations (N×overlap×noise×degradation rate). Panel A: post-collapse fidelity floor by class. Panel B: AUC by network size N and class; interaction F(6,528)=131.92, p=1.23×10⁻¹⁰¹, η²=0.600. Panel C: transition width by degradation rate; Pearson r=+0.658. Panel D: noise effect on pre-collapse plateau; t(52)=4.06, p_Bonferroni=0.003. Error bars: 95% CI throughout*

Figure 3. Clinical Validation: Model Predictions vs ADNI Empirical Decline

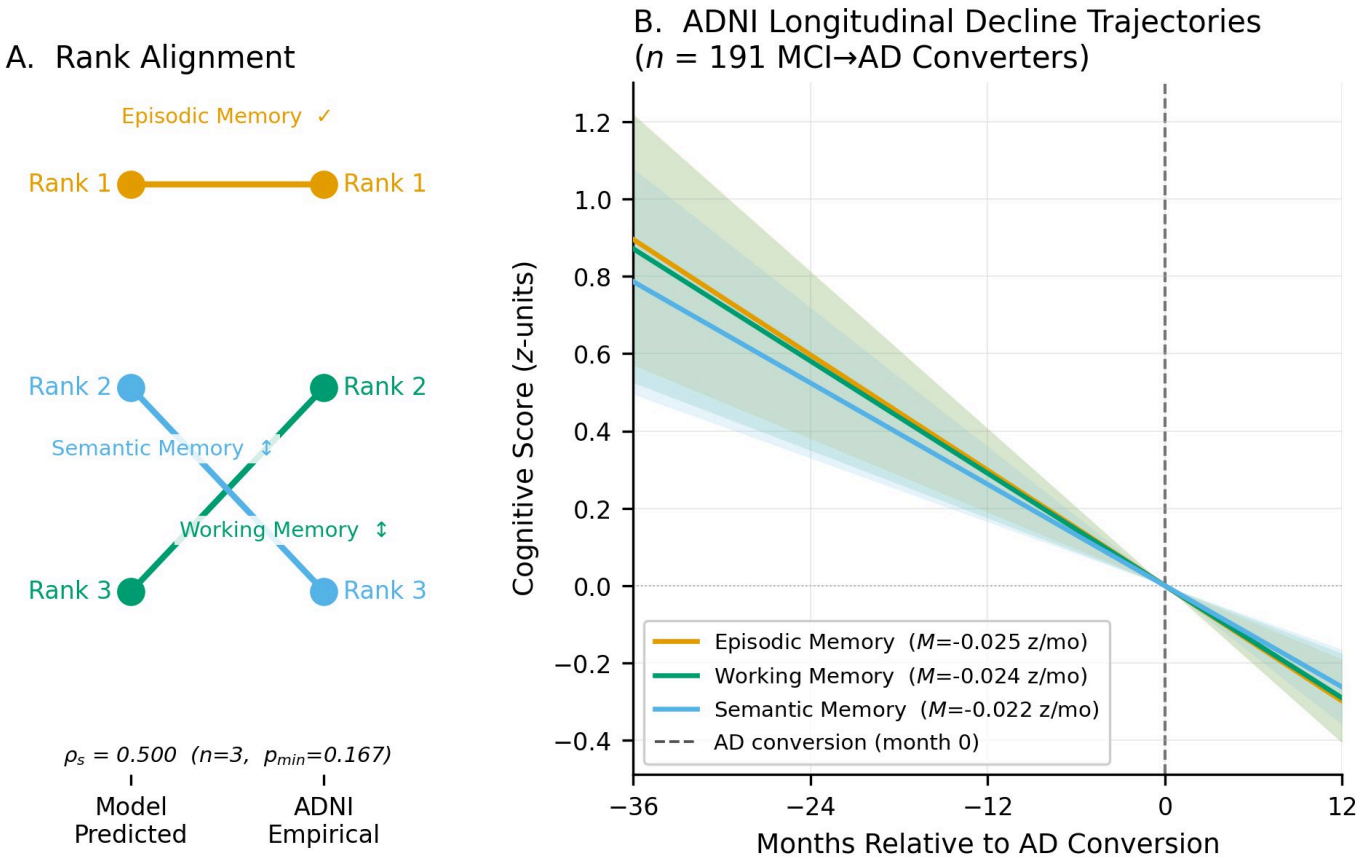


Figure 3

Clinical Validation

Fig. 3 *Clinical validation against ADNI longitudinal trajectories (n=191 MCI→AD converters). Panel A: slopegraph connecting model-predicted failure rank to empirically observed decline rank; ✓ denotes rank match, ↕ denotes rank reversal. Panel B: longitudinal z-scored cognitive trajectories across −36 to +12 months relative to AD conversion; shaded bands: 95% CI; grey dashed vertical: conversion point. Spearman ρs=0.500; minimum achievable p=0.167 at n=3 (p=0.667 reported for completeness)*

Figure 4. Network Size Preferentially Benefits Most Vulnerable Memory Class

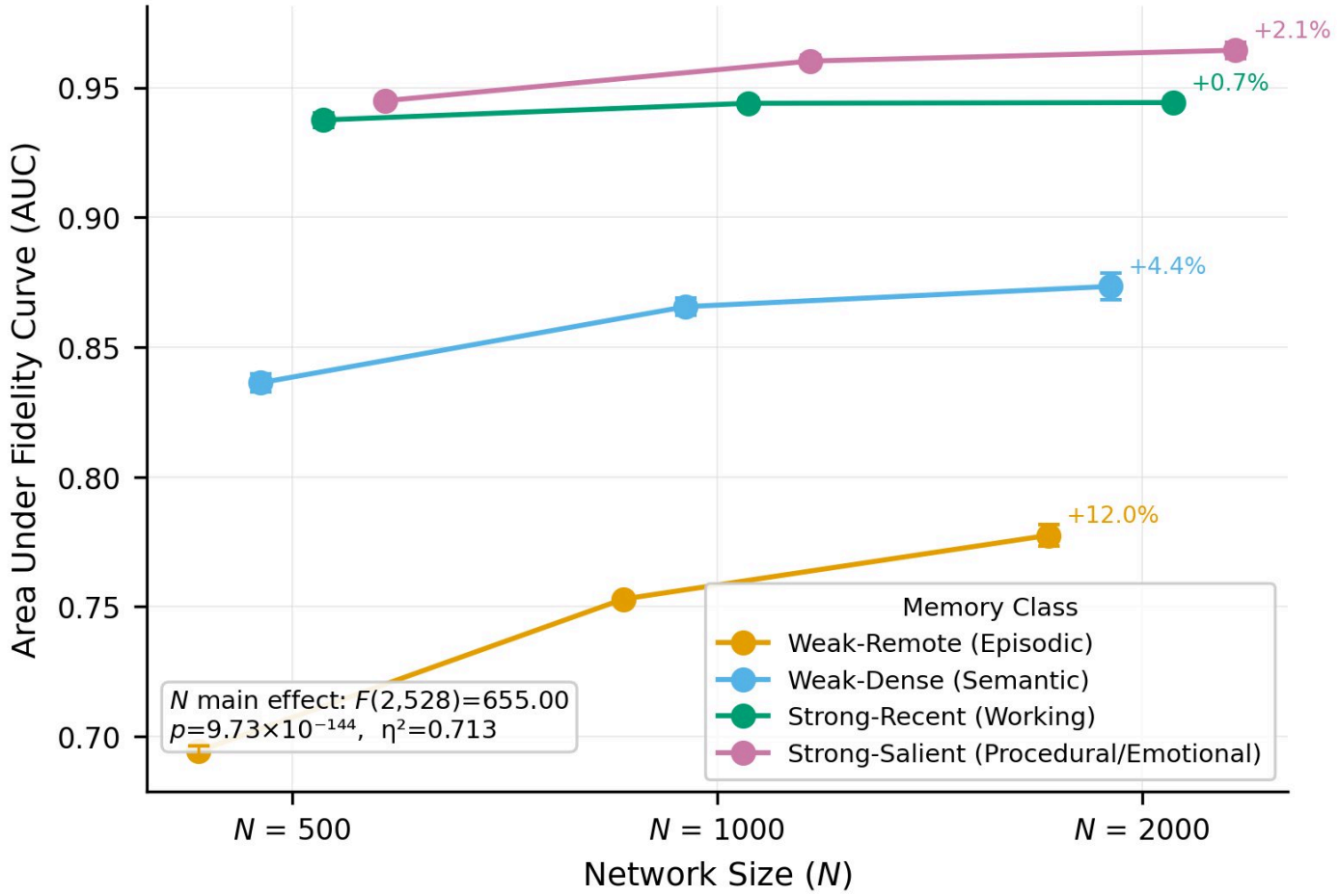


Figure 4

AUC by Network Size

Fig. 4 *Network size preferentially benefits the most vulnerable memory class. Mean AUC per class at N=500, 1000, 2000 averaged across all parameter configurations (n=45 per class per N level); error bars: 95% CI. Annotations: relative AUC gain from N=500 to N=2000. N main effect: F(2,528)=655.00, p=9.73×10⁻¹⁴⁴, η²=0.713*