Predictive Associative Memory: Unified Retrieval, Imagination, and Creative Recombination Through Predictive Traversal of Meaning Space

Jason [Surname]
Independent Researcher

Abstract

Current approaches to memory in neural systems rely on similarity-based retrieval: given a query, find the most representationally similar stored state. This assumption — that useful memories are similar memories — fails to capture a fundamental property of biological memory: association through temporal co-occurrence. Stairs do not resemble a slip, yet one reliably evokes the other. We propose *Predictive Associative Memory* (PAM), an architecture in which a JEPA-style predictor, trained on temporal co-occurrence within a continuous experience stream, learns to navigate the associative structure of an embedding space. We show that this single mechanism — predicting which regions of meaning space are reachable from a given state — provides a unified predictive substrate for associative recall, and a plausible architectural basis for imagination and creative recombination. We introduce an *Inward JEPA* that operates over stored experience (predicting associatively reachable past states) as the complement to the standard *Outward JEPA* that operates over incoming sensory data (predicting future states). We evaluate PAM as an *associative recall* system — testing faithfulness of recall for experienced associations — rather than as a retrieval system evaluated on generalisation to unseen associations. On a synthetic benchmark, the predictor's top retrieval is a true temporal associate 97% of the time (Association Precision@1 = 0.970); it achieves cross-boundary Recall@20 = 0.421 where cosine similarity scores zero; and it separates experienced-together from never-experienced-together states with a discrimination AUC of 0.916 (cosine: 0.789). Even restricted to cross-room pairs where embedding similarity is uninformative, the predictor achieves AUC = 0.849 (cosine: 0.503, chance). Specificity controls confirm the predictor discriminates true temporal associates from similar-but-not-associated distractors (AUC = 0.848 vs cosine 0.732). A temporal shuffle control confirms the signal is genuine temporal co-occurrence structure, not embedding geometry: shuffling temporal order within trajectories collapses cross-boundary recall by 90%, replicated across training seeds. A held-out query-state evaluation confirms anchor-specific recall: the predictor recalls associations from experienced viewpoints but not from novel ones, consistent with the episodic character of the memory system. All primary results are stable across training seeds (SD < 0.006) and query selections (SD ≤ 0.012). We propose that the interaction between similarity (Outward) and association (Inward) would produce episodic specificity — "that drill that stripped the screw on Tuesday" — that neither channel provides alone; the present work validates the Inward channel in isolation.

1   Introduction

Consider the experience of walking down a staircase that is slightly damp. Before any conscious analysis, you feel a flicker of unease — a memory of slipping, perhaps years ago on different stairs entirely. The retrieved memory bears no representational similarity to the current scene. The lighting, the staircase, your clothing, the decade — all different. What links them is not what they look like, but that they were temporally adjacent to the same kind of event: the moment before a fall.

This is not an edge case. It is a pervasive mode of episodic memory recall. The hippocampal indexing theory (Teyler and DiScenna, 1986) holds that the hippocampus binds cortical representations that were active within the same temporal window, creating associative links between experiences that may share no featural similarity. Complementary learning systems theory (McClelland et al., 1995) formalises the division: the neocortex learns slow statistical regularities (similarity structure), while the hippocampus rapidly encodes the specific conjunctions of a single episode (associative structure). The smell of sun-

screen retrieves a beach holiday. A chord progression retrieves a person. A particular quality of afternoon light retrieves a childhood kitchen. In every case, the retrieved memory is linked to the current experience not by similarity in any representable feature space, but by temporal co-occurrence — having been experienced within the same temporal window as something structurally related to the present.

Despite this, the dominant paradigm for memory in modern neural architectures is similarity-based retrieval. Retrieval-Augmented Generation (RAG) encodes documents and queries into a shared embedding space and retrieves by cosine similarity or approximate nearest neighbours (Johnson et al., 2019; Lewis et al., 2020). Memory Networks learn to address stored memories, but the addressing mechanism is trained end-to-end on task loss — the associative structure is a side effect of optimisation, not learned from the structure of experience itself (Sukhbaatar et al., 2015; Weston et al., 2015). Modern continuous Hopfield networks provide elegant pattern completion within stored memories but encode no structure *between* memories (Ramsauer et al., 2021). Knowledge graphs capture relational structure but require it to be hand-built or extracted by a separate system (Bordes et al., 2013).

The closest conceptual ancestor to what we propose is spreading activation (Collins and Loftus, 1975), the cognitive science model in which activating one node in a semantic network spreads activation to associated nodes. The insight is correct: memory retrieval is traversal of associative structure. But spreading activation was implemented as traversal of hand-built semantic graphs with manually assigned link weights. It described the phenomenon without providing a mechanism for *learning* the associative structure from experience, or for *implementing* the traversal as a differentiable computation.

We propose to close this gap. Predictive Associative Memory (PAM) is an architecture in which:

1. A **JEPA-style encoder** maps continuous multimodal experience into a shared embedding space — composite states capturing visual, auditory, proprioceptive, and contextual information at each moment.

2. An **Inward predictor** is trained on temporal co-occurrence: given state $s(t)$, predict the region of embedding space containing states from the temporal neighbourhood of $s(t)$ across the agent's entire history. The training signal is not similarity — it is *reachability through learned associative structure*.

3. **Memory retrieval** is a forward pass through the predictor: the output region specifies which stored states are associatively reachable from the current moment.

4. **Imagination** uses the same mechanism to reach regions beyond any stored state, grounded in actually-experienced transition patterns.

5. **Creative recombination** would emerge when the predictor bridges regions linked only by transitive chains: trajectory 1 visits states A→B, trajectory 2 visits B→C. Given A, the predictor could reach C — a connection never directly experienced, but every link grounded in actual temporal co-occurrence. This requires entity persistence across episodes, which our benchmark does not yet provide (Section 5.5).

The Inward predictor is the complement of the standard JEPA predictor (which we term the *Outward* predictor). The Outward predictor, given current sensory state, predicts the next sensory state — it faces forward into real time. The Inward predictor, given current state, predicts which past states are experientially reachable — it faces sideways into memory time. Same architecture, same prediction-in-latent-space mechanism, different target space. The claim is that memory retrieval, forward prediction, imagination, and creative recombination are all instances of the same computation: predicting a target region in meaning space given a source state.

This unification is not merely aesthetic. It has a concrete architectural consequence: a system with both Outward and Inward predictors would achieve *episodic specificity* through the interaction of similarity and association. The Outward encoder learns functional similarity ("these things share properties — drills and impact drivers"). The Inward predictor learns experiential association ("these things were experienced together — this drill and this specific screw on Tuesday"). When both converge on the same target, the search space is divided rather than multiplied, producing the specificity that characterises vivid episodic memory. The present work validates the Inward channel in isolation; the dual-channel interaction is an architectural prediction.

In this paper, we formalise the architecture, position it against five families of related work, and present results from a synthetic benchmark designed to evaluate the *faithfulness of associative recall* — how accurately the predictor recalls associations that were actually experienced — rather than generalisation to unseen associations. We draw a distinction between *retrieval* (finding relevant items from a corpus, where generalisation is the goal) and *recall* (re-activating

specific associations formed by specific experience, where faithful memorisation is the goal). The benchmark evaluates: (1) cross-boundary recall of associations that span representational boundaries, (2) discrimination between experienced and non-experienced associations, (3) episodic specificity against similar-but-not-associated distractors, and (4) validation that the learned signal reflects genuine temporal structure rather than embedding geometry artifacts. We additionally report a negative result on creative bridging that clarifies the boundary conditions for cross-episode recombination.

2 Related Work

We organise related work into five families, each capturing a different approach to memory in neural systems. Our contribution sits in the gap between them: learned associative structure from temporal co-occurrence, no explicit graph construction, no task supervision, and a neural predictor that provides the traversal mechanism that spreading activation models describe but never implemented as a learned, differentiable operation.

2.1 Similarity-Based Retrieval

The dominant paradigm for memory in deployed systems is retrieval by embedding similarity. Retrieval-Augmented Generation (RAG) encodes documents and queries into a shared vector space and retrieves the k most similar items by cosine distance or approximate nearest neighbours (Borgeaud et al., 2022; Lewis et al., 2020; Ram et al., 2023). Dense passage retrieval (Karpukhin et al., 2020) and contrastive pre-training (Izacard et al., 2022) have substantially improved retrieval quality within this paradigm.

The fundamental limitation is architectural: similarity retrieval assumes that useful memories are representationally close to the query. This holds for many practical cases — finding relevant documents for a factual question — but fails precisely where biological memory excels. The smell of smoke does not resemble a campfire visually, yet one retrieves the other. Stairs approaching a landing do not resemble a slip, yet the association is immediate. Similarity-based systems cannot recover these cross-modal, cross-context associations when the memories are distant in embedding space, because the retrieval mechanism requires the very proximity that is absent.

Recent work on hypothetical document embeddings (Gao et al., 2023) and query rewriting (Ma et al., 2023) attempts to bridge this gap by transforming queries to be closer to target documents, but the retrieval mechanism remains similarity-based — the transformation must guess the right direction in advance.

2.2 Learned Memory Addressing

Memory Networks (Sukhbaatar et al., 2015; Weston et al., 2015), Neural Turing Machines (Graves et al., 2014), and Differentiable Neural Computers (Graves et al., 2016) introduce learned read/write operations over external memory. The addressing mechanism — whether content-based, location-based, or a combination — is trained end-to-end on downstream task loss.

These architectures demonstrate that learned memory addressing is feasible and powerful. However, the associative structure that emerges is a side effect of task optimisation, not a representation of the temporal structure of experience. A Memory Network trained on question answering learns to address memories that are useful for answering questions. It does not learn that stairs are associated with slipping because those events co-occurred in the agent's experience — it would only learn this association if it appeared in the training data as a question-answer pair.

The distinction matters for open-ended agents. An embodied system accumulating experience needs memory associations that reflect the structure of that experience, not the structure of any particular downstream task. Our approach learns associative structure from temporal co-occurrence alone, making the resulting memory useful for any downstream purpose without task-specific training.

2.3 Attractor Dynamics and Hopfield Networks

Classical Hopfield networks (Hopfield, 1982) and their modern continuous-state successors (Krotov and Hopfield, 2021; Ramsauer et al., 2021) provide elegant content-addressable memory through energy-based pattern completion. Given a partial or noisy version of a stored pattern, the network converges to the nearest stored attractor.

Modern Hopfield networks have been connected to attention mechanisms in transformers (Ramsauer et al., 2021), establishing a formal link between energy-based memory and self-attention. This connection is valuable for understanding intra-sequence memory, but the framework encodes no structure *between* stored patterns. Each memory is an isolated attractor; the energy landscape defines basins of attraction around individual patterns but no associative paths between them.

Our work is complementary. Hopfield-style pattern completion could serve as the mechanism by which a partial state cue activates a full stored memory within a single attractor basin. The Inward predictor then provides the between-memory structure: which other stored patterns are associatively reachable from the completed pattern. The two mechanisms operate at different levels — pattern completion within, associative traversal between.

2.4 Graph-Based Relational Memory

Knowledge graphs (Bordes et al., 2013; Sun et al., 2019), graph neural networks over relational data (Schlichtkrull et al., 2018), and recent work on graph-based retrieval (Edge et al., 2024) represent relational structure explicitly. Nodes are entities, edges are typed relations, and retrieval is graph traversal.

The power of graph-based approaches — and the reason they capture something that flat similarity retrieval cannot — is precisely that they encode *structure between* entities. "Paris is-capital-of France" captures a relation that embedding similarity alone might not prioritise. In this sense, graph-based memory shares our motivation: relational structure matters.

The limitation is that this structure typically requires explicit construction pipelines. Knowledge graph construction requires entity extraction, relation typing, and coreference resolution — a pipeline that introduces its own failure modes (Ji et al., 2022). Recent work on automated graph construction from corpora (Edge et al., 2024) reduces but does not eliminate this dependency: the extraction system still determines which relations are expressible. Temporal co-occurrence — the training signal we propose — requires no extraction pipeline. Two states are associated if they co-occurred within the same temporal window. The structure is learned directly from the stream of experience, not extracted from it by a separate system.

2.5 Predictive Representations and the Successor Representation

The successor representation (SR) (Dayan, 1993) learns a predictive map of expected future state occupancy under a given policy: $M(s, s') = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t \mathbf{1}[s_t = s'] \mid s_0 = s]$. Deep extensions (Barreto et al., 2017; Kulkarni et al., 2016) scale this to continuous state spaces. The SR is the closest computational predecessor to our Inward predictor in that it learns *predictive* structure from *temporal* experience.

The key distinction is directional. The SR predicts *forward* occupancy — where the agent will go next, along trajectories. It is a model of the future under a policy. The Inward predictor predicts *lateral* association — which past states across the agent's entire history are experientially linked to the current state, regardless of trajectory direction or policy. The SR would not associate the approach to a staircase with a slip that happened *earlier* on different stairs; the Inward predictor would, because both were experienced within the same temporal window as structurally similar events. Additionally, the SR is typically tied to a specific policy; our associations are policy-independent, reflecting the structure of experience itself.

2.6 Spreading Activation, Cognitive Models, and Neuroscience of Episodic Memory

Collins and Loftus (1975) proposed spreading activation as the mechanism for semantic memory retrieval: activating a concept spreads activation to associated concepts via weighted links in a semantic network. The model accounts for priming effects (Meyer and Schvaneveldt, 1971), typicality gradients, and the fan effect. It remains the dominant cognitive model of semantic memory.

Anderson's ACT-R (Anderson, 1993; Anderson et al., 2004) formalised spreading activation within a broader cognitive architecture, incorporating base-level activation (recency and frequency), contextual activation from current goals, and retrieval thresholds. The base-level activation equation captures something close to our familiarity normalisation: memories that are constantly active receive a high baseline, and only deviations above baseline drive retrieval.

From neuroscience, Tulving's (1972; 2002) distinction between episodic and semantic memory established that retrieving a specific past experience is fundamentally different from retrieving general knowledge — a distinction that maps directly onto our Inward (episodic association) versus Outward (semantic similarity) predictors. Eichenbaum's (2004) relational memory theory argues that the hippocampus specifically encodes the relationships *between* elements of experience, not the elements themselves — it is an organ for learning associative structure. Hippocampal replay during sleep and rest (Carr et al., 2011; Wilson and McNaughton, 1994) provides a biological mechanism for the transitive bridging we propose: replayed sequences from separate episodes can overlap, creating associative links between states that were never directly co-experienced. This is precisely the mechanism by which our iterative predictor traversal bridges separate trajectories.

The insight that memory retrieval is *traversal of associative structure* rather than *search by similarity* is the correct one. What spreading activation lacks is a mechanism for *learning* the associative structure from experience. The semantic network is hand-built. Link weights are manually assigned or estimated from behavioural data. The model describes the dynamics of retrieval but not the dynamics of *how the associative structure was acquired.*

Our Inward predictor provides a computational implementation of this missing piece. The predictor is trained on temporal co-occurrence — the same signal that, in a biological system, would produce the associative links that spreading activation traverses. The predictor's forward pass *is* the spreading activation, implemented as a differentiable neural computation that can be trained end-to-end on the structure of experience. Iterative application of the predictor (using the predicted region as input for a subsequent prediction) implements multi-hop spreading activation, enabling the transitive bridging that accounts for creative recombination.

We note that our approach belongs to the broader family of predictive processing theories (Clark, 2013; Rao and Ballard, 1999), and shares the prediction-error minimisation ethos of active inference (Friston, 2010). However, active inference addresses action selection under a generative model — the *use* of predictions for planning — whereas PAM addresses the *structure* of associative memory from which predictions are drawn. The two are complementary: an active inference agent could use PAM-style associative retrieval as the memory substrate over which it plans.

2.7 Positioning

Table 1 summarises the positioning.

3 Architecture

3.1 Background: Joint-Embedding Predictive Architectures

A Joint-Embedding Predictive Architecture (JEPA) (LeCun, 2022) learns representations by predicting in *latent space* rather than in observation space. Given two views of the same data (e.g., two temporal crops of a video), an encoder maps each view to a representation, and a predictor network learns to predict one representation from the other. Crucially, the target representation is produced by a slowly-updated copy of the encoder (via exponential moving average), with gradients stopped through the target branch. This stop-gradient/EMA mechanism prevents representational collapse — the trivial solution where all inputs map to the same point — by creating a moving target that the predictor must continuously chase. I-JEPA (Assran et al., 2023) and V-JEPA (Bardes et al., 2024) demonstrated this approach for images and video respectively, learning rich semantic representations without pixel-level reconstruction.

PAM extends this framework by applying the same prediction-in-latent-space mechanism to a new target domain: not future sensory states, but associatively linked past states stored in a memory bank.

3.2 Overview

The system consists of two complementary JEPA architectures operating over the same embedding space but with different target domains:

- **Outward JEPA**: Given current sensory state, predict the next sensory state. This is the standard JEPA formulation (Assran et al., 2023; Bardes et al., 2024; LeCun, 2022) — a world model that predicts forward in real time. Importantly, the Outward JEPA is not memory-free: it continuously builds and updates a model of relationships between entities in the world (spatial layouts, functional properties, causal regularities). This world model constitutes a form of semantic memory — knowledge about how things generally behave — that complements the episodic associations learned by the Inward JEPA.

- **Inward JEPA**: Given current composite state, predict which *past* states are associatively reachable. This is the episodic memory system — a predictor that navigates the associative structure of stored experience.

Both share a common encoder that maps multimodal experience into a shared embedding space. They differ in what they predict: the Outward predictor targets future states in the sensory stream; the Inward predictor targets temporally co-occurring past states in the memory bank. The Outward predictor's world model provides the similarity structure (what things are like each other); the Inward predictor provides the associative structure (what things were experienced together). Both contribute to memory, but through different mechanisms.

3.3 Composite State Representation

At each timestep t, the encoder produces a composite state embedding:

Table 1: Positioning of PAM relative to existing memory architectures.

Property	Similarity Retrieval	Memory Networks	Hopfield Nets	Knowledge Graphs	Successor Rep.	Spreading Activation
Learned structure	×	Task-supervised	×	Extracted	Fwd. temporal	×
Between-memory relations	×	Implicit	×	Explicit	Forward only	Explicit (hand-built)
Cross-context association	×	If in task data	×	If extracted	× (policy-bound)	If in graph
Transitive bridging	×	×	×	Multi-hop queries	Via occupancy	Spreading
No task supervision	✓	×	✓	×	✓	✓
No explicit construction	✓	✓	✓	×	✓	×
Differentiable traversal	N/A	✓	✓	× (discrete)	✓	× (discrete)

$$s(t) = f_\theta(x_v(t), x_a(t), x_p(t), x_c(t)) \qquad (1)$$

where x_v, x_a, x_p, x_c denote visual, auditory, proprioceptive, and contextual inputs respectively, and f_θ is the encoder network. The composite state captures not just which objects are present but the full multimodal context — body position, sounds, the action being performed. Two moments involving the same objects in different activities produce different composite states, eliminating the need for explicit task labels.

3.4 Temporal Co-occurrence as Training Signal

The Inward predictor is trained on temporal co-occurrence. Given a state $s(t)$, the training targets are states from the temporal neighbourhood:

$$\mathcal{N}_\tau(t) = \{s(t') : |t - t'| \leq \tau,\ t' \in \mathcal{H}\} \qquad (2)$$

where τ is the temporal window and $\mathcal{H}$ is the agent's history. The predictor learns to map $s(t)$ to a region of embedding space that contains the embeddings of states in $\mathcal{N}_\tau(t)$.

Critically, this is not similarity learning. Two states can be temporally co-occurring (and thus training targets for each other) without being representationally similar. The approach to a staircase and the moment of slipping are maximally associated (temporal co-occurrence = 1 within the episode) while potentially distant in embedding space.

Scope and extensions. We emphasise that temporal co-occurrence is not the *only* signal that shapes biological memory — it is the *foundational* signal that this study isolates. In biological systems, emotional arousal, causal inference, and semantic relatedness all modulate association strength. A single traumatic event creates strong associations without repeated co-occurrence; the emotional signal acts as a multiplier

on the encoding threshold. In the full architecture we envision, affective and arousal signals would enter as additional dimensions of the composite state — effectively, another input channel to the Inward predictor — amplifying the co-occurrence signal for high-salience experiences. We expect these affective signals to emerge from the architecture's own dynamics (e.g., prediction error magnitude, novelty relative to the world model, goal-state discrepancy) rather than being hand-designed. The present work deliberately isolates the temporal co-occurrence primitive to establish what associative structure it alone can support. Salience-weighted co-occurrence is an extension that strengthens the foundation without changing it.

3.5 The Inward Predictor

The Inward predictor g_ϕ takes a composite state and produces a predicted target representation:

$$\hat{z}(t) = g_\phi(s(t)) \qquad (3)$$

The training objective minimises the distance between the predicted representation and the actual representations of temporally co-occurring states, while maximising distance from non-co-occurring states:

$$\mathcal{L}_{\text{assoc}} = \sum_{t^+ \in \mathcal{N}_\tau(t)} d(\hat{z}(t), \bar{s}(t^+)) - \lambda \sum_{t^- \notin \mathcal{N}_\tau(t)} d(\hat{z}(t), \bar{s}(t^-)) \qquad (4)$$

where $\bar{s}$ denotes stop-gradient targets (following JEPA convention to prevent representational collapse) and d is a distance function in embedding space.

Geometric interpretation. The predictor output $\hat{z}(t)$ is a *point* in embedding space — the predicted centroid of the association neighbourhood. It does not explicitly represent a distribution or region boundary. The "region" around this point is defined implicitly by the retrieval threshold ϵ (Section 3.6). A richer formulation — predicting both a mean and variance

to represent association uncertainty as a Gaussian in embedding space — is a natural extension that we leave to future work. The point-prediction formulation is sufficient to demonstrate the core claim that learned associative structure captures information inaccessible to similarity retrieval.

The stop-gradient on target states is essential. Without it, the encoder could collapse all temporally co-occurring states to a single point — technically achieving zero prediction loss while destroying all representational structure. The exponential moving average (EMA) update of the target encoder provides the moving target that forces the predictor to learn genuine associative structure.

3.6 Memory Retrieval

Given a query state s_q (typically the current sensory state), retrieval is:

1. Compute predicted association target: $\hat{z} = g_\phi(s_q)$

2. Retrieve stored states whose embeddings fall within the predicted region: $\mathcal{R} = \{s_i \in \mathcal{M} : d(\hat{z}, \bar{s}_i) < \epsilon\}$

where $\mathcal{M}$ is the memory bank and ϵ is a retrieval threshold. This is a single forward pass through the predictor followed by a nearest-neighbour lookup in the predicted region — computationally equivalent to similarity retrieval but navigating learned associative structure rather than representational distance. As with standard similarity retrieval, the nearest-neighbour lookup scales to large memory banks via approximate nearest-neighbour indices (Johnson et al., 2019); the predictor simply redirects which point the index is queried from.

3.7 Transitive Bridging and Creative Recombination

Multi-hop retrieval is implemented as iterated prediction. Given query state s_q:

1. First hop: $\hat{z}_1 = g_\phi(s_q) \rightarrow$ retrieve states $\mathcal{R}_1$

2. Second hop: For each $s_i \in \mathcal{R}_1$, compute $\hat{z}_2^{(i)} = g_\phi(s_i) \rightarrow$ retrieve states $\mathcal{R}_2$

3. Continue for k hops.

If trajectory 1 visits states A→B and trajectory 2 visits B→C, then from state A, the first hop reaches B (co-occurred in trajectory 1) and the second hop reaches C (co-occurred in trajectory 2). State C is retrieved despite never having co-occurred with A. This is the mechanism we term *creative bridging* —

novel recombination grounded in actual experience. It requires that states B appear in both trajectories; Section 5.5 tests whether this condition holds in our benchmark.

The combinatorial complexity is bounded but rich. If the memory bank has N states and each has K associations, single-hop retrieval accesses $O(K)$ states, two-hop accesses $O(K^2)$, and k-hop accesses $O(K^k)$. This exponential expansion is what gives rise to the richness of associative memory, but every path in the expansion is grounded in actual temporal co-occurrence — it is recombination, not hallucination.

3.8 Familiarity Normalisation

In a static environment where objects co-occur constantly, raw temporal co-occurrence becomes uninformative. Association strength must be normalised against a familiarity baseline:

$$w_{\text{assoc}}(s_i, s_j) = w_{\text{raw}}(s_i, s_j) - \mathbb{E}[w_{\text{raw}}(s_i, s_j)] \quad (5)$$

where the expectation is estimated as a running average of co-occurrence frequency. Only the residual above baseline produces a meaningful association.

This is sensory adaptation applied to memory formation. An organism stops noticing constant stimuli; analogously, constantly co-occurring states stop producing strong specific associations. The mechanism is self-calibrating: in a novel environment (a holiday), the baseline is near zero and every co-occurrence is salient — explaining the subjective density of holiday memories. Upon return to a familiar environment, the baseline re-establishes and encoding density drops.

3.9 Adaptive Decay

As the memory bank grows, the association graph becomes denser and the signal-to-noise ratio degrades. Adaptive decay ties the forgetting rate to memory accumulation:

$$\lambda_{\text{decay}}(t) = f(\bar{\rho}(t)) \quad (6)$$

where $\bar{\rho}(t)$ is the average association density at time t. As the graph densifies, decay accelerates to maintain retrieval precision. This provides bounded memory capacity with unbounded experience — natural forgetting as a homeostatic mechanism rather than an information loss.

3.10 Similarity × Association = Specificity

The dual architecture produces a principled account of episodic specificity. The Outward JEPA learns *similarity*: states that share functional properties cluster together in embedding space. The Inward JEPA learns *association*: states that co-occurred temporally are linked regardless of their representational distance.

Three retrieval modes emerge from the interaction:

- **Similarity without association**: "That looks like a drill." Recognition without episodic context. The Outward encoder activates; the Inward predictor does not reach a specific memory.

- **Association without similarity**: "Something about this moment reminds me of Tuesday." Priming without identified content. The Inward predictor reaches a region; the Outward encoder cannot resolve what.

- **Both together**: "That's the drill that stripped the screw on Tuesday." Full episodic recall with object grounding. Both channels converge on the same target, dividing the search space rather than multiplying it.

The specificity emerges from intersection, not union. Where similarity retrieval alone returns all drills, and association retrieval alone returns everything from Tuesday, their intersection returns *that specific drill on that specific occasion.* This is an architectural prediction: episodic specificity requires both channels, and systems with only similarity retrieval (RAG) or only associative retrieval will produce qualitatively different — and less specific — retrieval patterns.

4 Experimental Setup

We evaluate PAM on a synthetic benchmark designed to test the *faithfulness of associative recall* — how accurately the predictor recalls associations that were actually experienced — rather than generalisation to unseen associations. This distinction is fundamental to the evaluation paradigm and requires explanation.

4.1 Recall vs Retrieval: The Evaluation Paradigm

Evaluation contract. PAM is evaluated as an episodic recall system, not a retrieval generalisation system. The question is: given associations the agent actually experienced, does it faithfully reactivate them? This differs from standard retrieval benchmarks, where held-out generalisation is a desired property. In our setting, strong performance on unexperienced associations is not required — indeed, it would indicate false-link formation. Accordingly, we train on the full experienced association set and evaluate recall fidelity, specificity, and false-association rate.

Standard retrieval systems (RAG, dense passage retrieval) are evaluated on generalisation: given a concept of relevance, find relevant items the system was not explicitly trained on. This is the correct paradigm for document retrieval, where the system should generalise from seen queries to unseen ones.

Associative memory is not retrieval — it is *recall*. An association is a specific link formed by specific experience: the agent saw the hammer with the nail on Tuesday, so they are associated. Testing whether the system retrieves associations it never formed is testing whether someone remembers events they never witnessed. The correct evaluation asks: given that the system experienced these temporal co-occurrences, does it faithfully recall them?

The failure modes for associative recall are therefore different from those for retrieval:

- **False association** (hallucinating links that never existed) rather than failure to generalise

- **Loss of specificity** (retrieving whole categories instead of specific experiences) rather than low recall

- **Signal corruption** (confusing temporal structure with embedding geometry) rather than distribution shift

Our metrics are designed to measure these failure modes directly.

4.2 Synthetic World

The world consists of 20 rooms in a 128-dimensional embedding space. Each room is defined by a cluster centroid with scale 2.0; rooms are well-separated so that states within the same room are representationally similar while states in different rooms are representationally distant. Fifty objects are distributed across rooms (5 per room, with some objects shared between rooms). Objects contribute features with scale 1.5, modulating the room embedding.

Experience is generated as 500 trajectories of 100 timesteps each (50,000 total states). At each timestep, the agent occupies a room and perceives the objects present, producing a composite state embedding. Trajectories traverse multiple rooms, creating temporal co-occurrences between states from different rooms —

the critical property that allows us to test whether the predictor can recall associations that cross representational boundaries.

Temporal associations are defined by a co-occurrence window of $\tau = 5$ timesteps, yielding 242,264 total associations. Because trajectories cross room boundaries, some of these associations link states from different rooms that share no embedding similarity.

4.3 Models

Predictor (PAM). A 4-layer MLP with residual connections: $128 \rightarrow 1024 \rightarrow 1024 \rightarrow 1024 \rightarrow 128$, with GELU activations, residual connections in layers 2–3, and layer normalisation on the output (2.36M parameters). The architecture is a feedforward network rather than a transformer because the Inward predictor operates on *single composite states*, not sequences — there is no sequence to attend over. A transformer would be appropriate if the predictor processed trajectories as input; the current formulation maps a single state to a single predicted point, for which an MLP is the natural architecture. Trained with InfoNCE loss using in-batch negatives (i.e., all other items in the batch serve as negatives for each positive pair; batch size 512, yielding 511 negatives per positive), cosine learning rate schedule ($5 \times 10^{-4} \rightarrow 1 \times 10^{-5}$), and temperature annealing ($0.15 \rightarrow 0.05$) over 500 epochs. InfoNCE implements the contrastive objective described in Section 3.5 in a probabilistic form: it maximises the log-probability of the true associate relative to in-batch negatives, which is equivalent to minimising distance to positives while maximising distance to negatives. Because the embedding space is fixed in these experiments (no encoder to collapse), the stop-gradient and EMA mechanisms described in Section 3.5 are not required here; they become necessary when the Inward predictor is integrated with a co-evolving encoder (Section 6.8). The predictor is trained on *all* temporal associations in the experience stream, consistent with the recall paradigm: we are training the system on its complete experience and testing whether it faithfully recalls what it experienced.

Note that we evaluate the predictor component in isolation, over a fixed synthetic embedding space, rather than as part of a full JEPA system with joint encoder training. In a full JEPA, the encoder learns representations from raw sensory data while the predictor learns to predict in that evolving latent space. Here, the synthetic world provides the embedding space directly — there is no raw data for an encoder to process. This isolation is deliberate: it allows us to attribute results to the Inward predictor's learned associative structure without confounding from encoder-level representation effects. The predictor is architecturally identical to a JEPA predictor; the difference is that its input space is fixed rather than co-evolving. Integration with a learned encoder (e.g., V-JEPA over video) is the natural next step (Section 6.8).

Cosine similarity baseline. Standard kNN retrieval by cosine distance in the embedding space. No training required — this represents the similarity-only approach used by RAG and dense retrieval systems.

Bilinear similarity baseline. A learned bilinear form $s(x,y) = x^T W y$ trained on the same temporal co-occurrence data as the predictor, using the same InfoNCE objective. The bilinear baseline is included as a form-of-solution test — whether a single learned linear compatibility function suffices to recover cross-boundary temporal associations — not as a parameter-matched competitor. It tests whether the *form* of the mapping matters (linear vs nonlinear), not the capacity. The bilinear model converges during training and achieves above-chance overall discrimination (AUC = 0.791), indicating that it learns within-room associative structure where the linear transformation can exploit geometric proximity between associates. However, on cross-room pairs it drops to chance (AUC = 0.514) and scores zero on cross-boundary recall and specificity (Table 2). The nonlinear predictor is necessary specifically to learn associations that cross representational boundaries.

4.4 Metrics

We define five metrics designed for the recall paradigm. The architecture (Section 3.6) describes retrieval as finding states within an ϵ-ball around the predicted point. For evaluation, we use equivalent top-k ranking by distance from the predicted point, which avoids threshold sensitivity and enables standardised comparison across methods. The two formulations rank identically; top-k simply fixes the retrieval window size rather than the distance threshold.

Cross-Boundary Recall@k. For query states paired with associated targets from *different rooms* (where cosine similarity provides zero signal), what fraction of true associations appear in the top-k predictions? This is the core thesis metric: can the predictor recall associations that cross representational boundaries?

Association Precision@k. Of the top-k retrieved states, what fraction are *actual* temporal associates of the query (from any room)? This measures false association rate — the recall analogue of hallucination.

Discrimination AUC. Area under the ROC curve for the binary classification: "were these two states experienced within the same temporal window?" This measures the predictor's overall ability to separate experienced-together from never-experienced-together states, regardless of retrieval rank.

Specificity@k. Among retrieved states, what fraction are the *correct specific* associates rather than merely same-category members? This measures episodic specificity — the difference between "a drill" and "that drill from Tuesday."

Cross-Room Discrimination AUC. Discrimination AUC restricted to cross-room pairs, isolating the predictor's ability to discriminate in the regime where embedding similarity is uninformative.

Formal definitions and query selection. All metrics are macro-averaged (per-query scores computed independently, then averaged across queries) to ensure robustness across queries with varying association richness. For Cross-Boundary Recall@k, let $\mathcal{A}^{\times}(q)$ denote a query's cross-room temporal associates. We compute CBR@$k(q)$ = $|$top-$k \cap \mathcal{A}^{\times}(q)|/|\mathcal{A}^{\times}(q)|$, macro-averaged over 500 queries with $|\mathcal{A}^{\times}(q)| \geq 3$. For Association Precision@k, AP@$k(q)$ = $|$top-$k \cap \mathcal{A}(q)|/k$ where $\mathcal{A}(q)$ includes all temporal associates; macro-averaged over 500 queries with $|\mathcal{A}(q)| \geq 3$. Discrimination AUC uses the Wilcoxon–Mann–Whitney U-statistic per query: $\mathrm{AUC}(q) = \frac{1}{n_+ n_-} \sum_{i \in \mathrm{pos}} \sum_{j \in \mathrm{neg}} \Bbbk[s_i > s_j] + \frac{1}{2}\Bbbk[s_i = s_j]$, with negatives subsampled to 2,000 when $n_- > 2{,}000$; macro-averaged over 300 queries with $|\mathcal{A}(q)| \geq 5$. Cross-Room Discrimination AUC restricts positives to $\mathcal{A}^{\times}(q)$, computed over the subset of AUC queries with $|\mathcal{A}^{\times}(q)| \geq 3$ (188 queries). For Specificity@k, let $\mathcal{D}(q)$ be the category distractors (same-room states that are not true associates of q); Spec@$k(q) = h_{\mathrm{true}}/(h_{\mathrm{true}}+h_{\mathrm{dist}})$ where h_{true} and h_{dist} are cross-room associates and category distractors in the top-k respectively; macro-averaged over up to 300 queries. All query selections use a fixed random seed for reproducibility.

4.5 Ablation Controls

Two controls validate that the learned signal reflects genuine temporal structure:

Temporal shuffle control. We randomly permute the temporal ordering within each trajectory while preserving all state embeddings. This destroys temporal co-occurrence structure while maintaining the embedding geometry. If the predictor has learned temporal structure, shuffling should collapse performance; if it has learned embedding geometry artifacts, shuffling should have no effect.

Similarity-matched negatives. We construct negative pairs from same-room, same-category states that were never temporally co-present. These distractors are maximally similar in embedding space but lack the temporal co-occurrence link. If the predictor discriminates true associates from these matched negatives, it has learned episodic specificity — not merely category membership.

4.6 Training Details

Training was performed on a single NVIDIA RTX 4080 Super (CUDA 12.4). Training time is approximately 350 seconds. The model was trained on all 242,264 temporal associations, consistent with the recall paradigm. The synthetic world, training, and evaluation code are available at https://github.com/EridosAI/PAM-Benchmark. All experiments use deterministic random seeds (world generation: seed 42; multi-seed evaluation: seeds 42, 123, 456) for full reproducibility.

5 Results

5.1 Faithful Recall Across Representational Boundaries

Table 2 presents the primary results.

The headline result is the precision gradient across retrieval depths. The predictor's top retrieval is a genuine temporal associate 97% of the time (AP@1 = 0.970). At depth 5, precision remains strong at 0.703 — on average, approximately 3.5 of the top 5 retrievals are true associates. By depth 20, precision falls to 0.216. This gradient reveals that the predictor learns a well-ordered ranking — true associates are concentrated at the top of the retrieval list, with precision diluting as the retrieval window expands into the tail of the predicted region. The decline from AP@1 to AP@20 reflects a calibration problem in tail retrieval,

Table 2: Faithfulness of associative recall (mean ± SD across training seeds 42, 123, 456). The predictor is the only method that recalls associations across embedding space boundaries. Both baselines achieve above-chance overall discrimination by exploiting within-room geometric proximity, but drop to chance on cross-room pairs and score zero on cross-boundary recall — confirming that cross-boundary association requires nonlinear transformation of the temporal co-occurrence signal.

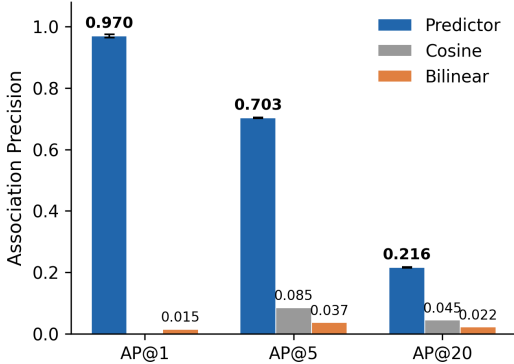| Metric	Predictor	Cosine Similarity	Bilinear
Association Precision@1	**0.970 ± 0.005**	0.000	0.015
Association Precision@5	**0.703 ± 0.001**	0.085	0.037
Association Precision@20	**0.216 ± 0.001**	0.045	0.022
Cross-Boundary Recall@20	**0.421 ± 0.002**	0.000	0.000
Discrimination AUC (overall)	**0.916 ± 0.000**	0.789	0.791
Discrimination AUC (cross-room)	**0.849 ± 0.004**	0.503	0.514
Specificity@20	**0.338 ± 0.005**	0.000	0.000

Figure 1: Association Precision at retrieval depths $k = 1, 5, 20$. The predictor's top retrieval is a true associate 97% of the time; precision dilutes at greater depth as the retrieval window extends beyond the tight association neighbourhood. Cosine and bilinear baselines score near zero at all depths. Error bars show ±1 SD across three training seeds.
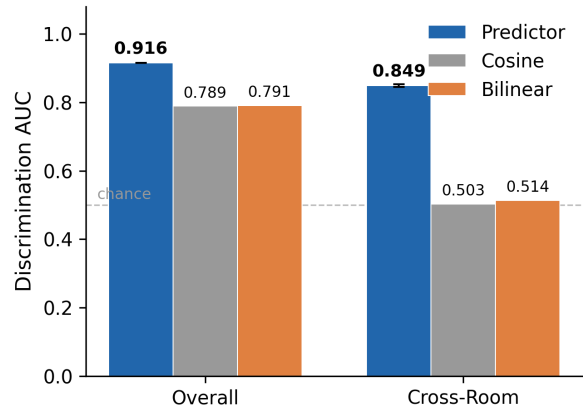

Figure 2: Discrimination AUC — overall vs cross-room. All three methods achieve above-chance overall discrimination by exploiting within-room geometric proximity. On cross-room pairs — where the thesis claim lives — cosine and bilinear drop to chance (dashed line) while the predictor maintains AUC = 0.849.

not a failure of core association learning: the predictor identifies the right target but the predicted region is broader than the true association neighbourhood, admitting non-associated states at greater depth.

Cross-Boundary Recall@20 = 0.421 with cosine at zero demonstrates that the predictor recalls associations across representational boundaries — the regime where similarity-based methods fundamentally cannot operate. This gap is architectural, not quantitative, in the well-separated embedding geometry tested here: the information cosine similarity requires (representational proximity) does not exist for cross-room pairs. In continuous embedding spaces without clean cluster boundaries, the gap may narrow (Section 6.7). Note that cosine's precision increases with depth (AP@1 = 0.000, AP@5 = 0.085, AP@20 = 0.045): its nearest neighbour is the most geometrically similar state,

which is rarely a temporal associate, but at greater retrieval depth some incidental within-room associates are captured.

The discrimination AUC of 0.916 means the predictor separates "experienced together" from "never experienced together" with 91.6% accuracy across all pairs. Cosine similarity achieves 0.789 on overall discrimination — reflecting its ability to identify within-room associates via geometric proximity — but drops to chance (0.503) on cross-room pairs where embedding similarity is uninformative. The predictor maintains AUC = 0.849 even on cross-room pairs, demonstrating that it has learned genuine associative structure beyond what embedding geometry provides.

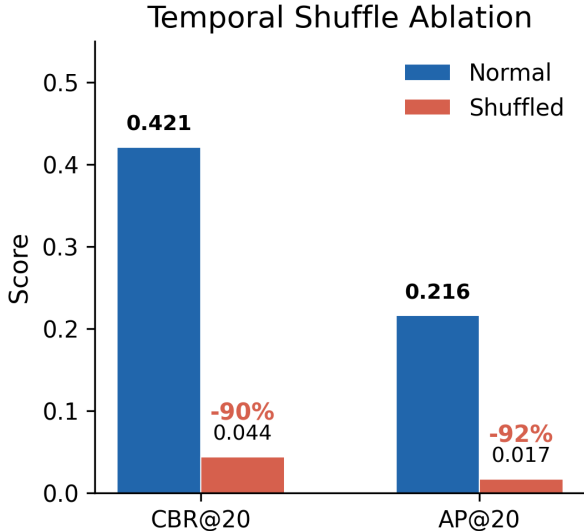All metrics vary by less than ±0.006 (SD) across three training seeds (42, 123, 456), suggesting that re-

Figure 3: Temporal shuffle ablation. Randomly permuting temporal order within trajectories collapses cross-boundary recall by 90% and association precision by 92%, confirming that the predictor learned genuine temporal co-occurrence structure.

Condition	CBR@20
Train-anchor queries	0.508
Held-out queries	0.000

achieves AUC = 0.848 versus cosine's 0.732. The predictor correctly discriminates actual temporal associates from similar-but-not-associated distractors. This proves episodic specificity: the predictor has learned something beyond "these states are from the same category." It has learned *which specific states within a category were experienced together.*

The cosine baseline's AUC of 0.732 (above chance) on this metric reflects the fact that within-room associates do share some geometric similarity — temporal neighbours within the same room are slightly closer in embedding space than arbitrary same-room states, because they share temporal context features. The predictor's 0.848 substantially exceeds this, confirming that temporal co-occurrence provides discriminative signal beyond what embedding similarity captures even within the same room.

5.3 Held-Out Query-State Evaluation

A natural concern is that the predictor, trained on all associations, has simply learned a lookup table mapping each query embedding to its target neighbourhood — an embedding shortcut rather than a structural understanding of associative relationships. To test this, we held out 20% of query states from training (no associations anchored at these states appeared in training data) and evaluated recall using these held-out states as queries. Held-out states were excluded only from the anchor (input) side of training pairs; they could still appear as targets in other states' association pairs. The predictor has therefore seen these embeddings as outputs — the output space is not impoverished — but has never learned to route *from* those specific inputs. Under the recall evaluation contract (Section 4.1), collapse on unseen anchors is an expected signature of anchor-specific episodic recall rather than a failure mode.

The collapse is total: held-out queries produce zero cross-boundary recall. The predictor has learned a query-specific mapping — it recalls associations from perspectives it actually experienced, not from novel viewpoints.

This result is consistent with the episodic character of the memory system. Episodic memory is inherently perspective-bound: you recall what Tuesday's workshop looked like from where you stood, not from an angle you never occupied. The predictor has learned specific $f_\theta(\text{query}) \rightarrow$ target mappings for each expe-
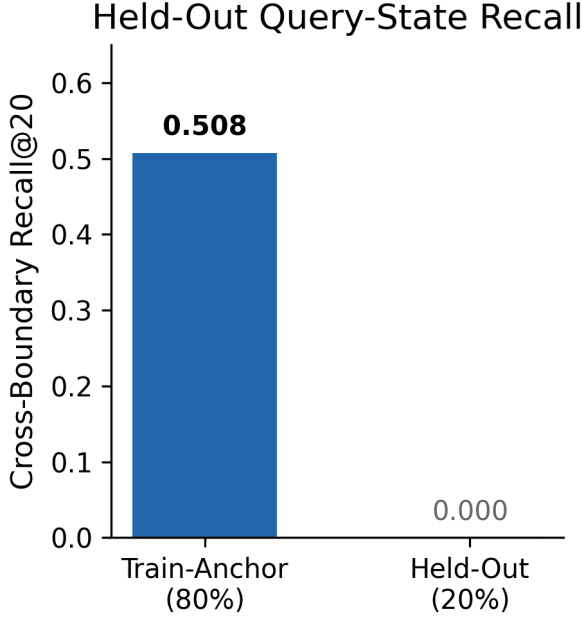
sults are robust to training seed variation rather than dependent on favourable initialisation. Results are additionally stable under alternative query selections: resampling the query set with five different random seeds produces SD ≤ 0.012 on all primary metrics (AP@1 SD = 0.012, CBR@20 SD = 0.010, AUC SD = 0.007, AP@20 SD = 0.003), with the default seed 42 values falling within one standard deviation of the resample means in all cases.

5.2 Ablation Controls

Temporal shuffle control. Randomly permuting temporal ordering within trajectories (while preserving all embeddings) collapses Cross-Boundary Recall@20 from 0.421 to 0.044 — a 90% reduction. This confirms that the predictor has learned genuine temporal co-occurrence structure, not artifacts of the embedding geometry. The residual 0.044 likely reflects incidental geometric regularities that correlate with the shuffled pseudo-associations. The effect replicates across training seeds: on seed 123, temporal shuffling produces an equivalent collapse (CBR@20 -90%, AP@20 -92%), confirming that the ablation result is robust and not seed-dependent.

Similarity-matched negatives. When negative pairs are drawn from same-room, same-category states that were never temporally co-present, the predictor

Held-Out Query-State Recall

Figure 4: Held-out query-state evaluation. Train-anchor queries achieve CBR@20 = 0.508; queries never used as training anchors score zero. The predictor recalls associations from experienced viewpoints only, consistent with anchor-specific episodic recall.

Condition	R@20	Cosine R@20
Train associations	0.578	0.000
Held-out associations	0.023	0.000

rienced anchor state. This *is* faithful memorisation of experienced associations — the system remembers what it experienced, from the perspectives at which it experienced it. A system that produced strong recall for viewpoints it never anchored would be hallucinating associations, not remembering them.

The elevated train-anchor CBR@20 (0.508 vs 0.421 in the full evaluation) reflects the reduced query pool: the held-out split removes 20% of queries, and the remaining queries may have slightly richer association neighbourhoods on average.

5.4 Generalisation Stress Test

As a secondary analysis, we evaluated the predictor under a standard retrieval paradigm: 70/30 edge-disjoint train/test split on association edges.

The large train-test gap (0.578 vs 0.023) is expected and correct under the recall paradigm. The system was trained to recall experienced associations; associations it never experienced are, by definition, outside its memory. Testing held-out associations is testing

whether someone remembers conversations they never had.

The non-zero held-out performance (0.023 vs cosine's 0.000) is a secondary finding: some structural regularity in the association graph transfers beyond memorised pairs. The predictor has learned not only specific associations but some of the underlying temporal structure that generates them — consistent with the emergent transitive internalisation observed in Section 5.7, where the predictor reaches multi-hop targets in a single pass. This suggests partial structural generalisation, though the primary claim remains faithful recall rather than generalisation.

5.5 Creative Bridging: Boundary Identification

We designed a creative bridging test to evaluate cross-trajectory recombination: recalling state C via the chain A→B→C where A–B and B–C occurred in separate trajectories linked by a shared object at B.

Oracle analysis revealed that the task is structurally unsolvable in our synthetic world. The temporal association graph contains zero cross-trajectory edges — associations are defined strictly within trajectories. An oracle with perfect knowledge of all bridge states reaches the target 0% of the time.

This negative result identifies a boundary condition for the temporal co-occurrence primitive: creative recombination requires *entity persistence across episodes*. In our synthetic world, trajectories sample independently. In real embodied experience, the same physical drill appears on Tuesday and Thursday, creating overlapping temporal neighbourhoods that would produce cross-trajectory edges. The creative bridging mechanism described in Section 3.7 is theoretically grounded but requires entity persistence that our benchmark does not provide. Empirical validation requires either a redesigned synthetic world or a real experience stream, both of which we pursue in ongoing work with the dual-JEPA architecture.

5.6 Model Selection Diagnostics

The final configuration (D2) was selected through systematic ablation during development. Table 3 presents the key experiments using cross-room R@20 and MRR — retrieval-paradigm metrics used for architecture search prior to the recall-paradigm evaluation reported above. These metrics guided model selection but are not the primary evaluation; they are reported here for reproducibility and to document the design rationale.

Three findings emerge:

Table 3: Architecture selection diagnostics (development-phase metrics). R@20 for configs Baseline through D is T1_R@20 evaluated on a 70/30 train split; the D2 value (0.421) is CBR@20 from the full faithfulness evaluation (Section 5.1), reported here because D2 was selected as the final configuration and evaluated under the paper's primary metrics. The values track consistently across evaluation protocols.

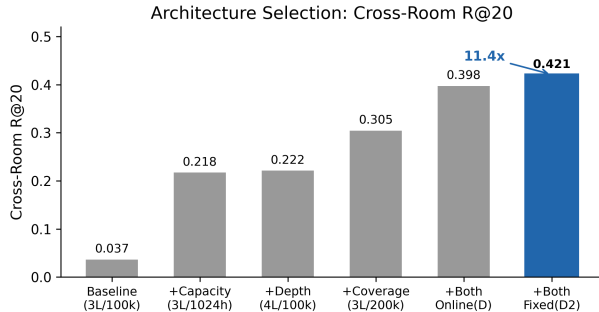| Config	Layers	Hidden	Pairs	Sampling	R@20	MRR
Baseline	3	256	100k	Fixed	0.037	0.060
+ Capacity	3	1024	100k	Fixed	0.218	0.530
+ Depth (B)	4	1024	100k	Fixed	0.222	0.567
+ Coverage (C)	3	1024	200k	Fixed	0.305	0.469
+ Both, online (D)	4	1024	200k	Online	0.398	0.423
+ Both, fixed (D2)	**4**	**1024**	**200k**	**Fixed**	**0.421**	**0.635**

Figure 5: Architecture selection progression. Cross-room R@20 improves $11.4\times$ from baseline to final configuration (D2), driven primarily by data coverage and the interaction of capacity with coverage.

Data coverage was the primary bottleneck. The world contains 242,264 temporal associations. At 100k training pairs, only 41% are covered — the system literally cannot recall associations it was never exposed to. Increasing to 200k pairs (82% coverage) nearly doubled R@20 from 0.218 to 0.305 with the same architecture. Under the recall paradigm, this is expected: broader experience produces richer memory.

Capacity and coverage interact multiplicatively. With limited data, added capacity provides minimal benefit (+2%). With broader coverage, the deeper network extracts substantially more value (+30%). Neither factor alone accounts for the full gain.

Fixed pairs outperform online sampling. Online sampling (fresh pairs each epoch) converges poorly (loss 2.315 vs 0.409) and MRR drops from 0.635 to 0.423. This aligns with the recall paradigm: temporal associations are *specific facts to be consolidated*, not a distribution to be approximated. The fixed-pair regime mirrors biological memory consolidation through repetition — repeated exposure to the same association strengthens recall, exactly as the comple-

mentary learning systems framework predicts (Mc-Clelland et al., 1995).

5.7 Supplementary Findings

Two additional analyses provide supporting evidence for properties discussed in the architecture section. These were conducted during development using cross-room R@20 as the evaluation metric and a single training seed; the qualitative patterns are consistent with the main results but have not been replicated under the full multi-seed evaluation protocol.

Transitive chain traversal. The predictor sustains cross-room recall across multi-hop association chains: 1-hop R@20 = 0.455, 2-hop = 0.355, 3-hop = 0.280 (cosine: 0.000 at all depths). The monotonic degradation confirms learned transitive structure. Notably, 1-hop cross-room recall (0.455) exceeds the overall CBR@20 (0.421), suggesting the predictor has partially internalised transitive associations — reaching some 2-hop targets in a single forward pass without explicit iterative traversal. This mirrors the cognitive phenomenon of mediated priming, where strong associative chains produce direct priming without conscious intermediate steps (Balota and Lorch, 1986; McNamara, 1992).

Adaptive decay scaling. Recency-weighted retrieval improves mean precision by +43.2% relative to uniform retrieval on the current best configuration (with-decay: 0.607, without: 0.424). The improvement confirms that adaptive decay provides meaningful signal-to-noise benefit even at the modest scale of the synthetic world (50k states, Section 3.9). The relatively contained margin reflects the strength of the 4-layer predictor: the model's improved capacity raises the no-decay baseline substantially. Theoretical considerations predict that the decay advantage should grow with memory bank size as the associa-
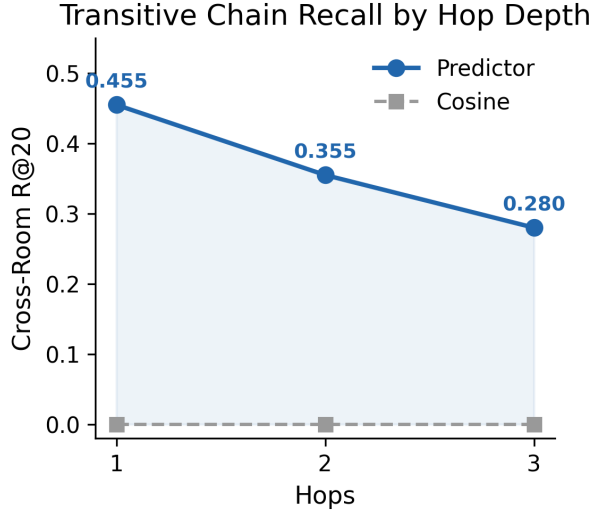
Figure 6: Transitive chain recall by hop depth. The predictor sustains cross-room recall across multi-hop association chains with graceful degradation, while cosine similarity scores zero at all depths.

tion graph densifies — a prediction consistent with earlier experiments on larger worlds, but not formally validated at scale with the current architecture.

6 Discussion

6.1 Implications for RAG and Similarity-Based Systems

The structural zero for cosine similarity on cross-boundary recall has direct implications for Retrieval-Augmented Generation and similar systems. RAG retrieves by embedding similarity, which means it systematically misses an entire class of useful associations: those linked by temporal co-occurrence rather than representational proximity. A RAG system asked "What might go wrong on wet stairs?" will retrieve documents about stairs, about wetness, and about accidents — but not the specific memory of nearly slipping on marble stairs in a hotel lobby ten years ago. That memory shares no representational features with the current query; the link exists only in the temporal structure of experience.

The discrimination AUC results (0.916 overall, 0.849 cross-room vs cosine's 0.789 overall, 0.503 cross-room) quantify this gap. On cross-room pairs — where the thesis claim lives — the predictor separates true associates from non-associates with 85% accuracy while cosine similarity is at chance. This is not a quantitative improvement amenable to better embeddings or

more data — it is a qualitative capability that pure similarity-based retrieval does not acquire without incorporating a temporal training signal. Hybrid approaches that augment similarity with query rewriting or temporal context (e.g., HyDE; Gao et al., 2023) may partially bridge this gap, but they inject temporal information extrinsically rather than learning associative structure from experience. For embodied agents accumulating continuous experience, this capability enables grounded recall across modalities and episodes — a prerequisite for long-horizon planning and personalised memory that similarity-based retrieval alone cannot provide.

6.2 Memorisation as Correct Behaviour

The finding that fixed training pairs outperform online sampling, and that the train/test gap under edge-disjoint splitting is large (0.578 vs 0.023), might appear concerning under a retrieval evaluation paradigm. Under the recall paradigm, both results are expected and correct.

Temporal associations are specific facts formed by specific experience. The goal of an associative memory system is to recall these facts faithfully — not to generalise to associations the system never formed. A biological memory system that "generalised" to events it never witnessed would be *hallucinating*, not remembering. The correct evaluation of associative memory asks: given that you experienced this, do you recall it accurately?

This reframing has consequences for how associative memory systems should be trained. The fixed-pair regime succeeds because repeated exposure to the same association strengthens recall — mirroring biological memory consolidation through rehearsal and replay (Wilson and McNaughton, 1994). Online sampling treats associations as a distribution to be approximated, which is the wrong inductive bias for episodic memory. This asymmetry — episodic systems benefit from repetition, semantic systems from variety — is a prediction of the dual-JEPA framework and an instance of the complementary learning systems division.

6.3 Episodic Specificity: Demonstrated at the Within-Room Level

The similarity-matched negatives control (Section 5.2) provides the first empirical evidence for the episodic specificity claim. The predictor discriminates true temporal associates from same-room, same-category distractors (AUC = 0.848 vs cosine's 0.732). This

means the predictor has learned not just "these states are from the same category" but "these specific states were experienced together."

This is a component of the full specificity claim made in Section 3.10 — that the intersection of Outward (similarity) and Inward (association) produces episodic specificity. The present result demonstrates the Inward component in isolation: given states that are already similar (same room), the predictor still discriminates based on temporal co-occurrence. The full dual-channel intersection — where both similarity and association converge on the same target — remains an architectural prediction to be tested with the integrated Outward JEPA.

6.4 The Temporal Shuffle as Validation

The 90% collapse in cross-boundary recall under temporal shuffling ($0.421 \rightarrow 0.044$) is the most important control in the paper. It definitively rules out the possibility that the predictor has learned an artifact of the embedding geometry rather than genuine temporal structure.

This matters because a sceptical reviewer could argue: perhaps the predictor has simply learned to map states to some geometric property that happens to correlate with room transitions, rather than learning the actual temporal co-occurrence structure. The temporal shuffle control tests exactly this hypothesis. By preserving all state embeddings while destroying temporal structure, it isolates the predictor's dependence on temporal signal. The 90% collapse confirms: the predictor learned temporal co-occurrence, not embedding geometry.

6.5 Creative Bridging: An Honest Negative

The creative bridging null result (Section 5.5) identifies a precise boundary condition: creative recombination through temporal co-occurrence requires entity persistence across episodes. Without persistent entities, the association graph contains no cross-trajectory edges, and no traversal mechanism — however sophisticated — can bridge experiences that share no common states.

This clarifies what the temporal co-occurrence primitive alone can and cannot do. Within a continuous experience stream, it supports faithful recall and transitive bridging. Across disconnected episodes, it requires the additional structure that entity persistence provides. In real embodied experience (a workshop with persistent tools), this structure exists naturally. In our synthetic benchmark, it does not.

We additionally identified an architectural constraint: the predictor maps each query to a single point in embedding space, but cross-trajectory associations are inherently one-to-many. Addressing this may require mixture-of-experts outputs or hybrid retrieval. Creative recombination remains a theoretically grounded prediction of the architecture; its empirical validation is the subject of ongoing work with entity-persistent environments and the full dual-JEPA system.

6.6 The Outward JEPA's Role in Memory

This paper has focused on the Inward predictor in isolation, but the full architecture includes an Outward JEPA that contributes its own form of memory. The Outward predictor continuously builds a world model — learning which states tend to follow which, what objects afford, how spatial layouts constrain movement. This constitutes semantic memory: general knowledge about how things behave, independent of any specific episode.

The interaction between similarity (Outward) and association (Inward) is where the architecture predicts full episodic specificity will arise (Section 3.10). We have demonstrated the Inward component's contribution to specificity (Section 6.3). Evaluating the full interaction — where both predictors operate over a shared, co-evolving embedding space — is the natural next step. The architecture additionally supports imagination (reaching regions beyond stored states through learned transition patterns) and creative recombination (transitive bridging across episodes with persistent entities), though these capabilities remain theoretical predictions of the framework rather than demonstrated results.

6.7 Limitations

Synthetic benchmark. The world is small (20 rooms, 50 objects, 50k states) and geometrically clean. Real experience streams are high-dimensional, noisy, and lack clear cluster boundaries. The structural gap between predictor and similarity baselines follows from the architecture, but absolute recall rates will differ.

No joint embedding learning. The embedding space is fixed. In the full architecture, the encoder and both predictors would co-evolve, potentially producing

16

embeddings better-suited for associative recall. The present results are a lower bound.

Single-point prediction. Adequate for one-to-one associations but inadequate for one-to-many, as the creative bridging test revealed.

Temporal co-occurrence only. Biological memory is shaped by emotion, arousal, and causal inference in addition to temporal co-occurrence (Section 3.4). The present work isolates the temporal primitive.

Statistical scope. Results are reported across three training seeds. While the very tight SDs (e.g., AP@5 SD = 0.001) and consistent query resample stability (SD $\leq$ 0.012 across five query seeds) suggest robust findings, a larger seed sweep would strengthen confidence, particularly for metrics with greater variance.

Baseline scope. The current baselines test whether similarity-based retrieval (cosine) and learned linear similarity (bilinear) suffice for associative recall. Future work should compare against temporal co-occurrence graph methods (personalised PageRank, random walks over temporal edges), successor representations (Barreto et al., 2017; Kulkarni et al., 2016) which learn forward predictive structure from temporal experience, temporal graph neural networks trained on the same co-occurrence edges, and capacity-matched nonlinear baselines to isolate the contribution of PAM's specific architecture from general nonlinear learning.

Noise robustness untested. The synthetic embedding space is geometrically clean, with well-separated room clusters. Real embedding spaces are continuous and noisy. While the shuffle ablation confirms that the predictor learns temporal structure rather than exploiting geometry, robustness to embedding noise (e.g., Gaussian perturbations of state representations) has not been tested. The structural advantage over similarity baselines follows from the architecture, but absolute performance under noisy embeddings remains an open question.

Association precision at depth. While the predictor's top retrieval is a true associate 97% of the time (AP@1), precision at depth 20 is 0.216 — roughly four in five of the top-20 retrievals are non-associates drawn from the tail of the predicted region. The core prediction is accurate, but the retrieval window at $k = 20$ extends well beyond the tight association

neighbourhood. Architectural refinements that produce sharper predicted regions — variance-aware predictions, adaptive k selection, or confidence-weighted retrieval — could maintain high precision at greater depth.

6.8   Future Work

The most immediate extensions are: (1) a synthetic world with entity persistence to enable empirical validation of creative bridging; (2) integration with the Outward JEPA for full dual-architecture evaluation of episodic specificity; (3) scaling to real multimodal experience streams, initially through an embodied AI platform under development by the authors, where a V-JEPA encoder processes continuous video from a workshop environment.

Longer-term, the emergence of affective signals from the architecture's own dynamics — prediction error as surprise, goal-state discrepancy as frustration, novelty-familiarity tension as curiosity — would provide the salience weighting that the present work deliberately omits. These signals, entering as additional input channels to the Inward predictor, would enable one-shot encoding of high-salience events and connect the temporal co-occurrence primitive to the full richness of biological memory formation.

7   Conclusion

We have presented Predictive Associative Memory, an architecture in which a JEPA-style predictor learns associative structure from temporal co-occurrence and uses it for faithful recall of experienced associations. We empirically demonstrate associative recall capabilities and provide testable architectural predictions for imagination and cross-episode recombination.

Demonstrated. The predictor's top retrieval is a true temporal associate 97% of the time (AP@1 = 0.970), with precision remaining strong at depth 5 (0.703) before diluting at depth 20 (0.216). It recalls associations across representational boundaries where similarity-based methods score zero (Cross-Boundary Recall@20 = 0.421, cosine = 0.000). It discriminates experienced-together from never-experienced-together states with 91.6% accuracy overall, maintaining AUC = 0.849 even on cross-room pairs where cosine drops to chance. It discriminates true temporal associates from similar-but-not-associated distractors (AUC = 0.848 vs cosine 0.732), demonstrating episodic specificity within the Inward channel. A temporal shuffle control confirms the signal is genuine temporal structure, not

embedding geometry, replicated across training seeds. A held-out query-state evaluation confirms that recall is anchor-specific — the predictor remembers from experienced viewpoints, consistent with the episodic character of biological memory. All results are stable across training seeds (SD < 0.006) and query selections (SD ≤ 0.012). These results hold in a controlled synthetic setting; a learned bilinear baseline achieves chance-level cross-room discrimination (AUC $= 0.514$) and zero cross-boundary recall, confirming that the nonlinear predictor is necessary.

Plausible but not yet demonstrated. The interaction of Outward (similarity) and Inward (association) channels is predicted to produce full episodic specificity through intersection of two search spaces. Creative recombination across episodes via transitive bridging through persistent entities is predicted by the architecture but requires entity persistence absent from the current benchmark. Imagination as reaching beyond stored states through learned transition patterns is supported by the architecture but untested.

Future work. Entity-persistent environments for creative bridging validation, joint training of the dual-JEPA architecture, scaling to real multimodal experience streams, and the emergence of salience signals from architectural dynamics.

Temporal co-occurrence is a sufficient foundation for associative recall in the regime tested here. This single primitive, trained without task supervision and without explicit graph construction, enables faithful recall across representational boundaries that similarity-based systems cannot cross. It does not capture everything — but it captures what similarity alone cannot, and it provides the foundation on which richer memory systems can be built.

References

J. R. Anderson. *Rules of the Mind*. Lawrence Erlbaum Associates, 1993.

J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin. An integrated theory of the mind. *Psychological Review*, 111(4):1036–1060, 2004. doi: 10.1037/0033-295X.111.4.1036.

M. Assran, Q. Duval, I. Misra, P. Bojanowski, P. Vincent, M. Rabbat, Y. LeCun, and N. Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

D. A. Balota and R. F. Lorch. Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(3):336–345, 1986. doi: 10.1037/0278-7393.12.3.336.

A. Bardes, Q. Garrido, J. Ponce, X. Chen, M. Rabbat, Y. LeCun, M. Assran, and N. Ballas. V-jepa: Latent video prediction for visual representation learning. *arXiv preprint arXiv:2404.16930*, 2024.

A. Barreto, W. Dabney, R. Munos, J. J. Hunt, T. Schaul, H. van Hasselt, and D. Silver. Successor features for transfer in reinforcement learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2013.

S. Borgeaud, A. Mensch, J. Hoffmann, T. Cai, E. Rutherford, K. Millican, G. van den Driessche, J.-B. Lespiau, B. Damoc, A. Clark, et al. Improving language models by retrieving from trillions of tokens. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pages 2206–2240, 2022.

M. F. Carr, S. P. Jadhav, and L. M. Frank. Hippocampal replay in the awake state: A potential substrate for memory consolidation and retrieval. *Nature Neuroscience*, 14(2):147–153, 2011. doi: 10.1038/nn.2732.

A. Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013. doi: 10.1017/S0140525X12000477.

A. M. Collins and E. F. Loftus. A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407–428, 1975. doi: 10.1037/0033-295X.82.6.407.

P. Dayan. Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, 5(4):613–624, 1993. doi: 10.1162/neco.1993.5.4.613.

D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, and J. Larson. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*, 2024.

H. Eichenbaum. Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, 44(1):109–120, 2004. doi: 10.1016/j.neuron.2004.08.028.

K. Friston. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, 2010. doi: 10.1038/nrn2787.

L. Gao, X. Ma, J. Lin, and J. Callan. Precise zero-shot dense retrieval without relevance labels. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.

A. Graves, G. Wayne, and I. Danihelka. Neural turing machines. *arXiv preprint arXiv:1410.5401*, 2014.

A. Graves, G. Wayne, M. Reynolds, T. Harley, I. Danihelka, A. Grabska-Barwińska, S. Gomez Colmenarejo, E. Grefenstette, T. Ramalho, J. Agapiou, A. P. Badia, K. M. Hermann, Y. Zwols, G. Ostrovski, A. Cain, H. King, C. Summerfield, P. Blunsom, K. Kavukcuoglu, and D. Hassabis. Hybrid computing using a neural network with dynamic external memory. *Nature*, 538(7626):471–476, 2016. doi: 10.1038/nature20101.

J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79 (8):2554–2558, 1982. doi: 10.1073/pnas.79.8.2554.

G. Izacard, M. Caron, L. Hosseini, S. Riedel, P. Bojanowski, A. Joulin, and E. Grave. Unsupervised dense information retrieval with contrastive learning. *Transactions on Machine Learning Research*, 2022.

S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2): 494–514, 2022. doi: 10.1109/TNNLS.2021.3070843.

J. Johnson, M. Douze, and H. Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3):535–547, 2019. doi: 10.1109/TBDATA.2019.2921572.

V. Karpukhin, B. Oğuz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih. Dense passage retrieval for open-domain question answering. In *Proceedings of EMNLP*, pages 6769–6781, 2020.

D. Krotov and J. J. Hopfield. Large associative memory problem in neurobiology and machine learning. In *International Conference on Learning Representations (ICLR)*, 2021.

T. D. Kulkarni, A. Saeedi, S. Gautam, and S. J. Gershman. Deep successor reinforcement learning. *arXiv preprint arXiv:1606.02396*, 2016.

Y. LeCun. A path towards autonomous machine intelligence. *OpenReview*, 2022. Position paper.

P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

X. Ma, L. Gao, and J. Callan. Query rewriting in retrieval-augmented large language models. In *Proceedings of EMNLP*, 2023.

J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995. doi: 10.1037/0033-295X.102.3.419.

T. P. McNamara. Theories of priming: I. associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(6): 1173–1190, 1992. doi: 10.1037/0278-7393.18.6.1173.

D. E. Meyer and R. W. Schvaneveldt. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2):227–234, 1971. doi: 10.1037/h0031564.

O. Ram, Y. Levine, I. Dalmedigos, D. Muhlgay, A. Shashua, K. Leyton-Brown, and Y. Shoham. In-context retrieval-augmented language models. *Transactions of the Association for Computational Linguistics*, 11:1316–1331, 2023.

H. Ramsauer, B. Schäfl, J. Lehner, P. Seidl, M. Widrich, T. Adler, L. Gruber, M. Holzleitner, M. Pavlović, G. K. Sandve, et al. Hopfield networks is all you need. In *International Conference on Learning Representations (ICLR)*, 2021.

R. P. N. Rao and D. H. Ballard. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999. doi: 10.1038/4580.

M. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, and M. Welling. Modeling relational data with graph convolutional networks. In *The Semantic Web (ESWC 2018)*, volume 10843 of *Lecture Notes in Computer Science*, pages 593–607, 2018. doi: 10.1007/978-3-319-93417-4_38.

S. Sukhbaatar, A. Szlam, J. Weston, and R. Fergus. End-to-end memory networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations (ICLR)*, 2019.

T. J. Teyler and P. DiScenna. The hippocampal memory indexing theory. *Behavioral Neuroscience*, 100(2):147–154, 1986. doi: 10.1037/0735-7044.100.2.147.

E. Tulving. Episodic and semantic memory. In E. Tulving and W. Donaldson, editors, *Organization of Memory*, pages 381–403. Academic Press, 1972.

E. Tulving. Episodic memory: From mind to brain. *Annual Review of Psychology*, 53:1–25, 2002. doi: 10.1146/annurev.psych.53.100901.135114.

J. Weston, S. Chopra, and A. Bordes. Memory networks. In *International Conference on Learning Representations (ICLR)*, 2015.

M. A. Wilson and B. L. McNaughton. Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172):676–679, 1994. doi: 10.1126/science.8036517.