Mem8: Ecphoric Reconstruction and the Dynamics of Resonant Memory

A Mathematical Monograph and Idealized System Specification

Flyxion
Independent Researcher

October 2026

Abstract

This monograph develops an idealized mathematical architecture for memory systems in which recollection is modeled as the context-dependent reconstruction of previously organized states rather than the retrieval of static representations. The framework combines complex-valued semantic encoding, phase-sensitive resonance, distributed oscillatory modes, recurrent associative dynamics, predictive inference, and history-dependent modification of persistent memory. Its motivating distinction is between the preservation of representational content and the preservation of conditions sufficient for its subsequent reinstatement.

Beginning with wave-based semantic memory and resonance-based retrieval, the framework generalizes static interference scores into an evolving system of coupled memory fields, contextual operators, temporal modes, and constrained reconstruction processes. Ecphoric recall is formulated as the interaction between a partial cue, a persistent trace, and the current dynamical state of the system. The resulting reconstruction is evaluated not solely by similarity to an encoded representation but also by contextual adequacy, historical fidelity, uncertainty, and its contribution to viable future continuation.

The specification develops conditions for stable encoding, selective resonance, associative completion, temporal binding, interference control, consolidation, reconsolidation, forgetting, and controlled revision. It distinguishes mathematically necessary properties from neuroscientific analogies and from experimentally testable architectural hypotheses. Formal results address identifiability, observability, stability, representational equivalence, reconstruction error, and the conditions under which phase-sensitive dynamics could offer advantages over conventional real-valued memory systems. Among them are a closed form for the resonance score of Listopad (2025) that exposes exactly what it adds to ordinary inner-product similarity, a realification theorem showing that no expressive advantage of complex over real vectors can be claimed, a necessary and sufficient condition for the existence of a decoder given an encoding, and an explicit stability threshold for the recurrent reconstruction map. A capacity theorem is stated as a baseline associative-retrieval result for the softmax readout, with its temperature requirement, and is not claimed to give an episodic capacity; a dynamic sufficiency theorem characterizes which memory transformations preserve the ability to answer a changing family of historically grounded questions; and the recurrent reconstruction iteration is related to Bayesian inference by showing that it computes posterior modes, which is a different operator from the posterior mean.

Mem8 is proposed as an idealized computational target rather than a claim of achieved biological equivalence. It supplies a common mathematical language for studying semantic memory, episodic reconstruction, dynamical association, and the preservation of organized experience across changing contexts.

Scope, Sources, and Status

The text is a specification and a theory. It is not a report of a completed empirical program. Every chapter in Part IX describes an experiment as a protocol with hypotheses, baselines, metrics, and acceptance criteria, and nothing in that part should be read as a result. The only numerical outputs reported anywhere in the monograph are the small deterministic verifications of Appendix M, the figures computed from the analytic results by the script supplied with the source, and the worked examples of Parts VI to VIII, whose every number is produced by the scripts `examples.py` and `lifecycle.py`. These confirm algebraic identities, probability bounds and exact computations on synthetic data, and say nothing about the architecture's competence at memory tasks.

Two sources organize the mathematical program. The first is the wave-based semantic memory proposal of Listopad (2025), which models patterns as complex waveforms and retrieves them by an interference score; it is treated here as one special case of a more general theory of ecphoric reconstruction, and its equations are reproduced with attribution and independently checked. The second is the study of Cabral, Fernandes, and Shemesh (2023), which used ultrafast functional magnetic resonance imaging in rats to identify intrinsic macroscale oscillatory modes with standing-wave properties. That study supplies empirical motivation for modeling distributed temporal coordination. It does not establish a mechanism of episodic recollection, and the monograph keeps that boundary visible wherever the two meet.

Two features of the first source shape the treatment that follows and are stated here so that later chapters can rely on them. The four named semantic operators of the wave-based proposal (negation, a controlled phase shift, and two amplitude operators) are named and discussed in its text but are not given formal definitions there, and results for the announced phase and amplitude perturbation experiments are not reported. The monograph therefore proposes candidate definitions of the operators where it needs them, labels them as such, and designs the reproduction protocol around what the source does specify. The second feature is that the source's own mapping from real embeddings assigns phase 0 or π according to the sign of each coordinate, a fact whose consequences for the comparison with real-valued retrieval are worked out in Chapter 9.

A third source is an earlier specification of the same name, by Chenoweth and Chenoweth [4], whose architecture of wave records, emotional modulation, temporal blankets, a protective controller and reactive layers is mapped onto the present framework in Appendix N. The mapping gives some of its mechanisms a precise form with small checks, records the internal inconsistencies of its printed figures, and lists its performance, safety and consciousness claims as reported by their source and not adopted. The relation between the two projects is for the author to state, and this monograph makes no claim of derivation.

A fourth group of sources consists of two companion manuscripts by the author on manifold-aligned prediction and semantic dynamics [15, 14]. Their tangent–normal decomposition is used in Chapter 36 as the curved generalization of an exactly checkable linear statement about the softmax readout, and their empirical claims about recent generative models are cited as claims of those manuscripts and are not independently asserted here.

Three labels mark material that is not an established theorem of the literature. A *Worked*

Example applies results already stated in the text to specific numbers or a finite instance, and every number in it was checked by a deterministic script. A *Counterexample* exhibits a situation in which a plausible inference fails, and each states which condition fails and which does not. An *Extension* is a proposed result that is proved where it is stated; it is built from classical ingredients, which are cited, and the claim that it is new is limited to the combination or the quantification, never to the ingredients. Where an extension restates a classical result, it says so. Theorem, proposition and chapter numbers of the earlier text are unchanged, and the new labels carry their own numbering within each chapter.

Contents

Abstract ii

Scope, Sources, and Status iii

I Foundations: What a Memory System Must Accomplish 1

1 Introduction: From Retrieval to Reconstruction 2

2 The Conceptual History of Ecphory 4

3 The Limits of Conventional Memory Architectures 6

4 The Idealized Mem8 Problem 8

II Wave-Based Semantic Representation 10

5 Complex-Valued Semantic States 11

6 Resonance and Interference as Retrieval Operations 13

7 Phase Transformations and Semantic Operators 16

8 ResonanceDB as a Baseline Architecture 18

9 The Representational Equivalence Problem 20

III Distributed Oscillatory Memory 22

10 Empirical Foundations of Macroscale Resonance 23

11 Spectral Decomposition of Distributed Activity 25

12 Coupled Oscillators and Phase Synchronization 27

13 Oscillatory Modes as a Computational Substrate 29

14 Temporal Binding and Phase-Coded Associations 31

IV Ecphoric Reconstruction 33

15 The Ecphoric Interaction Operator 34

16 Associative Completion and Attractor Geometry 35

17 Episodic Memory and Temporal Reinstatement **37**

18 Context-Sensitive Reconstruction **38**

19 Recollection, Familiarity, and Source Discrimination **39**

20 Memory Interference and False Recollection **40**

V Predictive Inference and Memory **42**

21 Memory as Inference over Latent Causes **43**

22 Variational Ecphory **46**

23 Precision Weighting and Resonance Selection **47**

24 Active Inference and Prospective Memory **48**

25 Amortized Deliberation and Learned Retrieval Policies **49**

VI Persistence, Consolidation, and Transformation **50**

26 Encoding as the Creation of Recoverable Constraints **51**

27 Consolidation and Multiscale Memory **52**

28 Reconsolidation and Memory Modification **55**

29 Forgetting, Interference, and Resource Constraints **58**

30 Memory Compression and Future Admissibility **60**

VII Formal Theory of Reconstructive Memory **63**

31 State Spaces and Transition Systems **64**

32 Observability and Reconstructive Identifiability **65**

33 Stability of Recurrent Reconstruction **72**

34 Resonance Selectivity and Capacity **77**

35 Information Geometry of Memory **81**

36 Projection, Abstraction, and Sufficient Memory States **83**

37 Memory as Preservation of Viable Continuation **88**

VIII The Mem8 Computational Architecture **93**

38 Representational Substrate **94**

39 The Oscillatory Coordination Engine **96**

40 The Ecphoric Retrieval Engine **98**

41 Persistent Memory and Historical Provenance **100**

42 Adaptive Memory Organization **103**

43 Resource Scheduling and Bounded Computation **106**

44 Distributed and Asynchronous Memory **108**

45 Reference Implementation **110**

IX The Experimental Program **115**

46 Reproduction of the Wave-Based Retrieval Paper **117**

47 Ecphoric Pattern Completion **118**

48 Temporal Sequence Reconstruction **119**

49 Contextual Interference and Source Attribution **120**

50 Consolidation and Reconsolidation **121**

51 Oscillatory Ablation Experiments **122**

52 Scaling and Computational Efficiency **124**

53 Robustness and Failure Analysis **125**

X Interpretation and Limits **126**

54 What the Architecture Explains **127**

55 Biological Plausibility and Computational Equivalence **129**

56 Recollection without Phenomenology **130**

57 Memory, Identity, and Historical Continuity **131**

58 The Ideal Architecture and Its Realizable Approximations **132**

59 Open Mathematical Problems **134**

60 Conclusion: Memory as Organized Reinstatement **136**

A Complex Hilbert Spaces and Realification **137**

B The Resonance Score: Derivation and Properties **138**

C Spectral Graph Theory **139**

D Coupled-Oscillator Stability **140**

E Recurrent Attractor Convergence **141**

F Variational Inference **142**

G Observability and Identifiability **143**

H Rate–Distortion Theory **144**

I Dynamical Sufficiency under Projection **145**

J Computational Complexity **146**

K Reproducibility Protocols **147**

L A Minimal Mem8 Model **148**

M Pilot Algebraic Verification **149**

N Reconciliation with the July 2025 MEM|8 Specification **153**
N.1 Mapping of components 153
N.2 Emotional modulation as a prior 153
N.3 Context-dependent decay 154
N.4 Noise-floor storage 154
N.5 Repetition and the Custodian 155
N.6 Reactive layers 155
N.7 Coherence normalization 155
N.8 Reported by the source and not adopted 155
N.9 Arithmetic of the printed figures 155

Part I

Foundations: What a Memory System Must Accomplish

Chapter 1

Introduction: From Retrieval to Reconstruction

The dominant engineering picture of memory treats it as content-addressable storage. A record is written under some key or embedding, a query of the same kind is later presented, and the system returns the stored record whose key best matches the query. This picture is adequate for a large class of tasks and has been refined for decades through hashing, indexing, vector quantization, and approximate nearest-neighbor search. It is inadequate as an account of remembering, because it silently assumes that what is stored at encoding is what is wanted at retrieval, that the retrieval cue is of the same kind as the stored key, and that the correct output is always an element of the store. None of these assumptions survives contact with the phenomena that the psychology of memory has catalogued since Bartlett's studies of serial reproduction [13], in which remembered material is shaped by the schema, context, and purpose present at recall, or with the experimental literature on suggestion and misinformation [9], in which the very same stored episode yields different recollections under different cues.

Five distinct capacities are routinely run together under the single word memory, and the first task of a formal treatment is to separate them. Semantic lookup returns a stored item or a stored attribute of an item given a cue that names it, and is correctly modeled by a map from keys to values. Associative completion returns a whole pattern given a fragment of it, and is correctly modeled by a dynamical system with attractors or by a posterior over patterns given partial evidence. Episodic reinstatement returns not only the content of a past event but the organization that bound that content together at the time, including order, duration, and the surrounding context. Reconstructive remembering returns a representation assembled at the time of recall from stored constraints and current inference, which may differ from every stored element and which may be correct or incorrect with respect to the history it purports to describe. Prospective simulation returns a representation of a state that has not occurred, assembled from the same constraints and used to guide action. A system may be excellent at the first and incapable of the others, and the converse is also possible; success at one capacity does not imply success at another.

Proposition 1.1 (Non-implication of capacities)**.** *There exist memory systems that perform exact semantic lookup and exact associative completion on every cue from a stated family, and that cannot reinstate the temporal order of any episode composed of two or more items.*

Proof. Let the item vocabulary be a set of V orthonormal vectors $e_1, \dots, e_V$ in a Hilbert space, and let an episode be an ordered pair $h = (a, b)$ of distinct items. Define the encoding $E(h) = e_a + e_b$, which is symmetric in its arguments. Lookup of either item and completion of the pair from one member, by projection onto the stored sum, are exact. For the two episodes $h_1 = (a, b)$ and $h_2 = (b, a)$ one has $E(h_1) = E(h_2)$. A target recollection function $T(h, c)$ that returns the first item of the episode takes the values a and b respectively for any cue c that does not itself reveal the order, so no decoder $\mathcal{E}$ can satisfy $\mathcal{E}(E(h), c) = T(h, c)$ for both histories. The failure is not one of retrieval strategy. It is a property of the encoding, and Chapter 32 states the exact condition that characterizes it. □

The proposition has a simple moral that organizes the rest of the monograph. A memory system is judged by the family of future recollections it can support, and the capacity to support one family does not extend to another. The question is therefore never whether a given architecture stores information in some abstract sense, but which recollection tasks the information it preserves is sufficient for.

The central problem can now be stated. Let a past organization be denoted x_{t_0}, a latent object that may be an item, a pattern, an episode, or a trajectory. At a later time t, the system holds persistent memory M_t, receives a cue c_t that is generally partial and may be of a different kind from anything stored, and is in a contextual state z_t that includes its current inferential and dynamical condition. The goal is to produce an estimate

$$\widehat{x}_{t_0} = \mathcal{E}(c_t, M_t, z_t) \tag{1.1}$$

that is constrained by the history that produced M_t. The ecphoric operator $\mathcal{E}$ is the central object of the theory, and it must be defined independently of the particular encoding scheme, the retrieval algorithm, and the dynamical substrate used to implement it. The wave-based resonance score of Listopad [1], the modern continuous associative memories of Ramsauer and colleagues [18], and the oscillatory reconstruction processes developed in later chapters are all candidate realizations of $\mathcal{E}$, and the theory is designed so that they can be compared on equal terms.

The monograph is organized in ten parts. The first fixes vocabulary and the system state. The second reconstructs the wave-based semantic representation with independent checks and proves what it can and cannot claim over real-valued vectors. The third examines distributed oscillatory dynamics and the empirical study that motivates them. The fourth through sixth develop reconstruction, inference, and the persistence and transformation of traces. The seventh states the formal theory, including the sufficiency criterion around which the work is built. The eighth specifies the computational architecture, the ninth the experimental program, and the tenth separates what has been demonstrated from what has only been motivated.

Chapter 2

The Conceptual History of Ecphory

The vocabulary of ecphory comes from Richard Semon, whose Mneme books distinguished the engram, the lasting change left in an organism by a stimulus, from the processes by which that change is acquired and later brought to expression [39, 40]. Semon called the acquisition engraphy and the awakening of a stored engram by a retrieval stimulus ecphory. Two features of his account remain structurally important. The first is that the engram is not itself the remembered experience: it is a persistent condition of the organism, and the experience occurs only when the condition is brought into interaction with an ecphoric stimulus. The second is that ecphory is partial and cue-driven. A fragment of the original stimulus complex is sufficient to evoke the whole, which Semon called the principle of homophony, and the same engram may be evoked by different fragments with different results.

Tulving restored the term to experimental psychology and made the interaction between cue and trace the central object of study [11, 12]. His encoding specificity principle holds that what is retrievable depends on the overlap between the information present in the cue at retrieval and the information encoded with the trace at storage, so that a cue that is objectively weaker may nonetheless succeed if it reinstates the encoding context. Two consequences follow. Retrievability is a relation between trace and cue, never a property of the trace alone, and the accessibility of a trace can change without any change to the trace. Tulving further distinguished the retrieval information that guides output from the recollective experience of re-living an episode, a distinction taken up in Chapter 56, where the monograph declines to infer the second from the first.

Tulving's later term for the joint dependence is synergistic ecphory: the experience that results is a product of the trace and the cue together, so that a cue carrying misleading information changes what is reconstructed, and no inspection of the trace alone predicts the outcome [12]. The classic demonstration is the cue-wording effect of Loftus and Palmer, in which changing a single verb in a question about a filmed collision shifted estimates of speed and the later endorsement of details that had not occurred [9]. In the notation introduced next, synergy is the statement that $\mathcal{E}(m, c, z)$ is not a function of m alone, and the cue-wording effect is the statement that two cues c, c' with the same intended content can send one trace to different reconstructions. Bartlett's account of remembering as an effortful reconstruction guided by schemas [13] contributes the complementary point that the reconstruction draws on the rememberer's prior organization as well as on trace and cue, which in the formal theory enters through the context argument z.

For the formal theory three maps are separated. Let $\mathcal{H}ist$ be a space of histories, understood as whatever organized past the system could in principle be asked about, and let $\mathcal{M}$ be a space of memory states.

Definition 2.1 (Engraphy, persistence, ecphory)**.** Trace formation, or engraphy, is a map $E : \mathcal{H}ist \to \mathcal{M}$ that assigns to each history the memory state it leaves, possibly stochastically. Trace persistence is a family of maps $P_s : \mathcal{M} \to \mathcal{M}$, $s \geq 0$, with $P_0 = \mathrm{id}$ and $P_{s+s'} = P_{s'} \circ P_s$, describing the autonomous evolution of the memory state during an interval without new

input. Ecphory is a map $\mathcal{E} : \mathcal{M} \times \mathcal{C} \times \mathcal{Z} \to \mathcal{Y}$ from memory state, cue, and context to a reconstruction.

The semigroup law for persistence is an idealization, since real persistence is perturbed by interference from intervening events, a feature added in Part VI. It is adopted here so that the three maps can be studied in isolation and then composed.

The separation leads to a notion that the rest of the monograph uses repeatedly. Two physically different memory states can be equivalent with respect to a specified family of future recollections, in the sense that no cue from the family can tell them apart through $\mathcal{E}$. This is the memory analogue of observational equivalence in automata theory.

Definition 2.2 (Recollective equivalence)**.** Fix a cue family $Q \subseteq \mathcal{C}$ and a context family $\mathcal{Z}_0 \subseteq \mathcal{Z}$. Two memory states $m, m' \in \mathcal{M}$ are $(Q, \mathcal{Z}_0)$-equivalent, written $m \sim_{Q,\mathcal{Z}_0} m'$, if $\mathcal{E}(m, c, z) = \mathcal{E}(m', c, z)$ for all $c \in Q$ and $z \in \mathcal{Z}_0$.

Equivalence is an equivalence relation, and the quotient $\mathcal{M}/\sim_{Q,\mathcal{Z}_0}$ is the space of memory states that matter for the stated family. A state with surplus structure, in the sense of distinctions no admissible cue can elicit, lies in the same class as one without it. This is the sense in which the monograph speaks of preserving conditions for reinstatement rather than preserving content: what must be preserved is the class, and everything finer is free to vary, to be compressed, or to be forgotten.

Chapter 3

The Limits of Conventional Memory Architectures

Several families of architecture have been proposed as models or engineering substitutes for memory, and each fixes a different compromise among the capacities distinguished in Chapter 1. A brief comparison clarifies what a new architecture must add and which of its claims will need independent support.

Symbolic stores and relational databases keep records under explicit keys and answer exact queries. They support lookup with perfect fidelity and offer no completion from partial or corrupted cues unless the query language is extended with similarity operators. Key-value memories, including the attention mechanism of transformers [2], generalize the exact match to a soft match, so that the output is a weighted combination of stored values with weights given by a similarity between the query and the keys. Vector embeddings with approximate nearest-neighbor indexes trade a controlled loss of recall for sublinear query time and support retrieval by semantic proximity, but they return elements of the store, and their similarity is typically a cosine, which discards magnitude and carries no notion of context beyond what the encoder placed in the vector.

Hopfield networks introduced the idea that memory can be a property of a dynamical system, with stored patterns as stable fixed points of an energy-descending update [25]. The classical network stores a number of random patterns proportional to the number of units, with a critical ratio near 0.14 patterns per unit [6]. Dense associative memories raised the capacity to a polynomial of higher degree in the dimension [7], an exponential-interaction variant achieves capacity exponential in the dimension [33], and the continuous formulation of Ramsauer and colleagues showed that a single update step of such a network is the attention operation, with exponentially large capacity and retrieval in one or few steps [18]. Sparse distributed memory offers a complementary construction, in which addresses and contents are high-dimensional binary vectors and storage is distributed over all locations within a Hamming radius of the address [37]. Holographic reduced representations compose and decompose structured items by circular convolution, which is multiplication in the frequency domain of complex-valued vectors [44], and are the closest precedent for phase-based binding. Recurrent networks and differentiable neural memories add learned dynamics and learned read and write operations at the cost of capacity guarantees.

A comparison on common terms requires dimensions along which any memory can be assessed. Seven are used throughout: storage capacity, which is the number of items or episodes that can be stored with recall error below a stated bound; cue dependence, which is the class of cues for which recall succeeds; reconstruction accuracy, which is distance to the target recollection, not merely the identity of the nearest stored item; interference, which is the degradation of recall of one item caused by the presence of others; modification, which is whether and how stored material can be updated with its history preserved; latency; and computational cost per query and per write. Table 3.1 gives a qualitative placement of the families on these dimensions, based on the literature cited above. The entries are intended to show where each family is structurally strong or weak and are not measurements.

Family	Capacity	Cue dependence	Output	Modification
Database / key-value	Limited by storage only	Exact or soft key match	Stored element	Explicit, with versioning if designed
Embedding with ANN index	Large, with recall loss from approximation	Similarity in encoder space	Stored element	Insert and delete, no provenance by default
Classical Hopfield	Order 0.14 per unit [6]	Partial pattern, same format	Fixed point, possibly spurious	Hebbian rewrite, destructive
Dense / continuous Hopfield	Polynomial to exponential in dimension [7, 33, 18]	Partial pattern, same format	Fixed point or metastable mixture	Rewrite of stored matrix
Sparse distributed memory	Large for high dimension [37]	Noisy address within a radius	Blended content	Additive writes
Phase-aware resonance [1]	Exact scan, no training	Resonance in complex space	Stored element	Insert, no stated provenance
Mem8 (this work)	Bound derived in Chapter 34	Partial, contextual, cross-format	Reconstruction with confidence and provenance	Controlled, provenance-preserving

Table 3.1: Qualitative comparison of memory families on the dimensions used in this monograph. Entries for earlier families summarize the cited literature. The final row states design targets, not achieved properties.

Two limits are common to most of these architectures. First, they conflate the stored object with the output: even when retrieval is dynamical, the stable output is intended to equal a stored pattern, so a memory is judged correct when it returns one of the things that was put in. Second, they treat context as a property of the cue, not as a persistent condition of the system that alters how every cue is read. Both limits are consequential for episodic and reconstructive tasks, where the correct output may be a composite of stored constraints, where the same cue must yield different outputs under different contexts, and where the output must carry an assessment of its own historical support.

Chapter 4

The Idealized Mem8 Problem

The theory is first posed for an unconstrained system, in which representation spaces are infinite-dimensional where convenient, computation is exact, and no budget limits retrieval. Constraints enter in Parts VIII and IX. The advantage of the unconstrained formulation is that it separates what is impossible for information-theoretic reasons from what is impossible for resource reasons, a separation that Chapter 32 makes precise.

The full system state is the tuple

$$X_t = (M_t, \Psi_t, \Theta_t, Z_t, H_t, B_t), \tag{4.1}$$

in which M_t denotes the persistent traces, Ψ_t the distributed representational fields, Θ_t the oscillatory phases, Z_t the contextual and inferential state, H_t the historical provenance, and B_t the computational resource constraints. The first specification requirement is that every component have a defined state space, a transition law, an observation map, and a persistence policy. Table 4.1 gives the idealized assignment, and the sections that follow develop each row.

The system evolves as a controlled stochastic process, written formally in Chapter 31. At each step it may receive an observation to be encoded, a cue to be answered, or neither, and it may take actions that change the context or the resources. The accessible mnemonic state, meaning whatever can be read out through some sequence of admissible cues, is in general a strict function of the physical state: the same accessible state can be realized by many physical states, which is the content of the recollective equivalence of Definition 2.2.

A requirement on the problem statement itself is that it separate three kinds of success. A reconstruction can be accurate with respect to the history that produced the trace, which is historical fidelity. It can be adequate to the present context, in the sense that it supplies what the current inference needs, which is contextual adequacy. It can be valuable for the future, in the sense that it improves the continuation of the system under its objectives, which is viability. The three are related but distinct, and a design that collapses them cannot represent the case in which a useful reconstruction is historically false, or the case in which a historically accurate reconstruction is useless in the present. The viability formulation of Chapter 37 keeps historical fidelity as a separately measurable constraint for exactly this reason.

Component	State space	Transition law	Observation	Persistence policy
M_t traces	Finite set of tuples in $\mathcal{H} \times \mathcal{T} \times \ldots$ (Ch. 38)	Write, update, merge, split, forget (Ch. 28, 42)	Through $\mathcal{E}$ only	Persistent, append-only for inscriptions
Ψ_t fields	Complex fields on a spatial index set	Field equation (Ch. 13)	Mode coefficients	Transient, reconstructed on demand
Θ_t phases	Torus $\mathbb{T}^N$	Coupled-oscillator dynamics (Ch. 12)	Order parameters, relative phases	Transient, reset by events
Z_t context	Context space $\mathcal{Z}$ with a group action on cues	Inference and task dynamics (Ch. 18)	Context reports	Slowly varying, logged
H_t provenance	Append-only ledger of typed records	Append only (Ch. 41)	Ledger queries	Permanent
B_t resources	Budgets on time, memory, energy	Scheduler (Ch. 43)	Remaining budget	Per task

Table 4.1: State components of Mem8 with the four required specifications. Chapter references indicate where each specification is developed.

Part II

Wave-Based Semantic Representation

Chapter 5

Complex-Valued Semantic States

The wave-based proposal of Listopad models each semantic pattern as a discrete complex-valued waveform

$$\psi(x) = A(x)\, e^{\mathrm{i}\phi(x)}, \qquad A(x) \geq 0, \quad \phi(x) \in [-\pi, \pi), \tag{5.1}$$

where x indexes the coordinates of the pattern, A is an amplitude that the source interprets as semantic intensity or salience, and ϕ is a phase that the source interprets as contextual or operator-level modulation [1]. The generalization used here replaces the coordinate index by an arbitrary index set and the pattern by an element of a complex Hilbert space $\mathcal{H}$, with inner product $\langle \psi, q \rangle = \sum_x \overline{\psi(x)}\, q(x)$ in the finite-dimensional case, linear in the second argument. Individual traces are vectors $m_i \in \mathcal{H}$, and the representation above is recovered by writing each coordinate in polar form.

The basic quantities are standard but must be fixed because the later arguments depend on exactly which of them a given score uses. The energy of a pattern is $E(\psi) = \|\psi\|^2 = \sum_x |\psi(x)|^2 = \sum_x A(x)^2$, which is independent of phase. The inner product splits into a real and an imaginary part, $\langle \psi, q \rangle = \mathrm{Re}\langle \psi, q \rangle + \mathrm{i}\, \mathrm{Im}\langle \psi, q \rangle$, and the real part carries the sign-sensitive agreement between the two patterns coordinate by coordinate: $\mathrm{Re}\langle \psi, q \rangle = \sum_x A_\psi(x) A_q(x) \cos(\phi_q(x) - \phi_\psi(x))$. Interference between two patterns, the squared norm of their sum, satisfies $\|\psi + q\|^2 = \|\psi\|^2 + \|q\|^2 + 2\,\mathrm{Re}\langle \psi, q \rangle$, which is the algebraic content of the statement that waves in phase reinforce and waves in antiphase cancel. Coherence is used here in a precise sense: the modulus of the normalized inner product, $\kappa(\psi, q) = |\langle \psi, q \rangle| / (\|\psi\| \|q\|) \in [0, 1]$, measures the degree of alignment up to a single global phase, while the normalized real part $c(\psi, q) = \mathrm{Re}\langle \psi, q \rangle / (\|\psi\| \|q\|) \in [-1, 1]$ measures alignment including the phase offset. A pattern and a copy of it multiplied by $e^{\mathrm{i}\alpha}$ have coherence 1 and phase-aware agreement $\cos \alpha$.

The distinction between phase as an encoding convention and phase as a dynamically meaningful relational variable can now be stated exactly. Multiplication of every pattern in a system by the same unit complex number $e^{\mathrm{i}\alpha}$ is a gauge transformation: it changes every vector and no observable that depends only on norms and on $|\langle \psi, q \rangle|$, and it changes no quantity that depends only on $\mathrm{Re}\langle \psi, q \rangle$ computed between two patterns transformed together. Phase is therefore a convention to the extent that only global phase is arbitrary. Phase is a relational variable to the extent that relative phases between components, or between a query and a stored pattern, change the outcome of a computation. The distinction matters because a representation in which phase carries information is only as good as the discipline that keeps phases comparable across encodings: if the encoder of a pattern and the encoder of a query assign phases by different conventions, the relative phase between them is an artifact and the score it modulates is noise.

Proposition 5.1 (Gauge invariance)**.** *Let $U_\alpha \psi = e^{\mathrm{i}\alpha} \psi$. Then $\langle U_\alpha \psi, U_\alpha q \rangle = \langle \psi, q \rangle$ for every α, while $\langle U_\alpha \psi, q \rangle = e^{-\mathrm{i}\alpha} \langle \psi, q \rangle$. In particular $|\langle \psi, q \rangle|$ is invariant under independent phase changes of the two arguments, and $\mathrm{Re}\langle \psi, q \rangle$ is invariant only under a common phase change.*

Proof. Conjugation of the first argument contributes $e^{-\mathrm{i}\alpha}$ and the second argument $e^{\mathrm{i}\alpha}$, which cancel when the phases are equal. Independent phases α, β yield $e^{\mathrm{i}(\beta-\alpha)}\langle \psi, q \rangle$, which has the same modulus and a rotated real part. □

Chapter 6

Resonance and Interference as Retrieval Operations

The retrieval score of the wave-based proposal is built from the interference of a stored pattern with a query, scaled by an energy-balance factor [1]. Writing $E_1 = \|\psi_1\|^2$ and $E_2 = \|\psi_2\|^2$, the source defines

$$S(\psi_1, \psi_2) = \frac{1}{2} \cdot \frac{\sum_x |\psi_1(x) + \psi_2(x)|^2}{E_1 + E_2} \cdot R, \qquad R = \frac{2\sqrt{E_1 E_2}}{E_1 + E_2}, \tag{6.1}$$

with $S = 0$ when $E_1 + E_2 = 0$. The factor R equals one when the energies agree and falls as they diverge. The source also gives the equivalent form obtained from $\sum_x |\psi_1 + \psi_2|^2 = E_1 + E_2 + 2\,\mathrm{Re}\langle\psi_1, \psi_2\rangle$, namely

$$S = \frac{\big(E_1 + E_2 + 2\,\mathrm{Re}\langle\psi_1, \psi_2\rangle\big)\sqrt{E_1 E_2}}{(E_1 + E_2)^2}, \tag{6.2}$$

and in the notation of norms, with ψ the stored pattern and q the query, the score is

$$S(\psi, q) = \frac{\|\psi + q\|^2 \, \|\psi\| \, \|q\|}{\big(\|\psi\|^2 + \|q\|^2\big)^2}. \tag{6.3}$$

The source states several properties: boundedness in $[0, 1]$, symmetry, self-match $S(\psi, \psi) = 1$ for nonzero ψ, invariance under a global phase, $S = 0$ for antiphase pairs, the reduction $S = (1 + \cos\theta)/2$ in the real equal-norm case, and the fact that $1 - S$ is not in general a metric. It also states that $S = 1$ if and only if the two patterns are equal and nonzero, without a proof. This chapter gives a closed form from which all of these follow at once, supplies the missing proof, and identifies exactly what the score contributes beyond ordinary inner-product similarity.

Theorem 6.1 (Closed form of the resonance score)**.** *Let $\psi, q \in \mathcal{H}$ be nonzero, $a = \|\psi\|$, $b = \|q\|$, and define the energy balance $\rho = 2ab/(a^2 + b^2) \in (0, 1]$ and the phase-aware cosine $c = \mathrm{Re}\langle\psi, q\rangle/(ab) \in [-1, 1]$. Then*

$$S(\psi, q) = \frac{\rho}{2}\,(1 + \rho\, c). \tag{6.4}$$

Proof. From the interference identity, $\|\psi + q\|^2 = a^2 + b^2 + 2abc = (a^2 + b^2)(1 + \rho c)$, using $2ab = \rho(a^2 + b^2)$. Substituting into (6.3) gives $S = (a^2 + b^2)(1 + \rho c)\, ab/(a^2 + b^2)^2 = \frac{ab}{a^2 + b^2}(1 + \rho c) = \frac{\rho}{2}(1 + \rho c)$. □

The identity was checked numerically on twenty thousand random pairs of complex vectors with unequal energies, with maximum absolute discrepancy below 10^{-15} (Appendix L); the check confirms the algebra and is not evidence for anything else.

Corollary 6.2 (Bounds, symmetries, and extremal cases)**.** *For nonzero ψ, q: (i) $0 \le S \le 1$; (ii) $S(\psi, q) = S(q, \psi)$; (iii) $S = 1$ if and only if $q = \psi$; (iv) $S = 0$ if and only if $q = -\psi$; (v) for every real orthogonal map O of the underlying real space $\mathbb{R}^{2n}$, $S(O\psi, Oq) = S(\psi, q)$, in particular for*

every unitary map and every common phase; (vi) $S(\psi, q) \leq \frac{\rho}{2}(1+\rho)$ regardless of the alignment of the two patterns.

Proof. Since $\rho \in (0, 1]$ and $c \in [-1, 1]$ we have $1 + \rho c \in [0, 2]$ and so $S \in [0, \rho] \subseteq [0, 1]$, which gives (i), and (vi) follows from $c \leq 1$. Symmetry holds because ρ and $\mathrm{Re}\langle\psi, q\rangle$ are symmetric. For (iii), $S = 1$ requires $\rho(1 + \rho c) = 2$, which forces $\rho = 1$ and $c = 1$ because both factors are at most 1 and 2 respectively; $\rho = 1$ means $a = b$, and $c = 1$ with equal norms means $\mathrm{Re}\langle\psi, q\rangle = \|\psi\|\|q\|$, which is the equality case of Cauchy–Schwarz for the real inner product and forces $q = \psi$. The converse is immediate. For (iv), $S = 0$ with $\rho > 0$ forces $1 + \rho c = 0$, hence $\rho = 1$ and $c = -1$, hence $a = b$ and $\mathrm{Re}\langle\psi, q\rangle = -ab$, hence $q = -\psi$. For (v), ρ depends only on the norms and c only on the real inner product, and both are preserved by any real orthogonal map. □

Item (v) is the structural observation that matters most. The score is invariant under the full orthogonal group $O(2n)$ of the real space underlying $\mathcal{H} = \mathbb{C}^n$, which strictly contains the unitary group $U(n)$, and a function with this much symmetry cannot depend on the complex structure of $\mathcal{H}$ in any way that a real inner-product space does not already encode. The complex structure enters retrieval only through the operators that act on the patterns before the score is computed, a point developed in Chapter 7.

For normalized patterns the score reduces to

$$S(\psi, q) = \frac{1 + \mathrm{Re}\langle\psi, q\rangle}{2}, \qquad \|\psi\| = \|q\| = 1, \tag{6.5}$$

since $\rho = 1$ and $c = \mathrm{Re}\langle\psi, q\rangle$. The reduction exhibits exactly which part of the retrieval mechanism is attributable to the complex representation and which part is ordinary inner-product similarity: on unit-norm patterns, S is an affine function of the real part of the complex inner product, and ranking by S is ranking by $\mathrm{Re}\langle\psi, q\rangle$. The one quantity that distinguishes the complex case from a purely real one is that $\mathrm{Re}\langle\psi, q\rangle$ is sensitive to relative phase, through $\cos(\phi_q - \phi_\psi)$ in each coordinate, and this sensitivity is the entire phase-awareness of the score on normalized data.

Proposition 6.3 (Ranking under equal energies)**.** *Let stored patterns $m_1, \ldots, m_P$ have a common norm a and let q be a query. Then $S(m_i, q)$ is a strictly increasing function of $\mathrm{Re}\langle m_i, q\rangle$, so retrieval by S ranks the traces exactly as retrieval by the real-part cosine does. If the patterns have unequal norms the rankings can differ, and the difference is produced solely by the balance factor ρ.*

Proof. With equal stored norm and fixed query norm, ρ is the same for all traces and $S = \frac{\rho}{2}(1 + \rho c_i)$ is strictly increasing in c_i, and $c_i = \mathrm{Re}\langle m_i, q\rangle/(ab)$ is proportional to $\mathrm{Re}\langle m_i, q\rangle$. □

The behavior under unequal energies deserves a worked case because it is the part of S that is not inner-product similarity. For a perfectly aligned query of scale λ, $q = \lambda\psi$ with $\lambda > 0$, one has $c = 1$ and $\rho = 2\lambda/(1 + \lambda^2)$, so $S = \frac{\rho}{2}(1 + \rho)$. A query of twice the energy scale, $\lambda = 2$, gives $\rho = 0.8$ and $S = 0.72$, and a query of half the scale gives the same value, because $\rho(\lambda) = \rho(1/\lambda)$. A pattern is therefore penalized for energy mismatch even when its direction is exactly right, and the penalty is symmetric in the logarithm of the scale. For antiphase scaled queries, $q = -\lambda\psi$, one has $c = -1$ and $S = \frac{\rho}{2}(1 - \rho)$, whose maximum over ρ is $1/8$ at $\rho = 1/2$ and whose value is zero at $\lambda = 1$ and in the limit $\lambda \to 0$. This

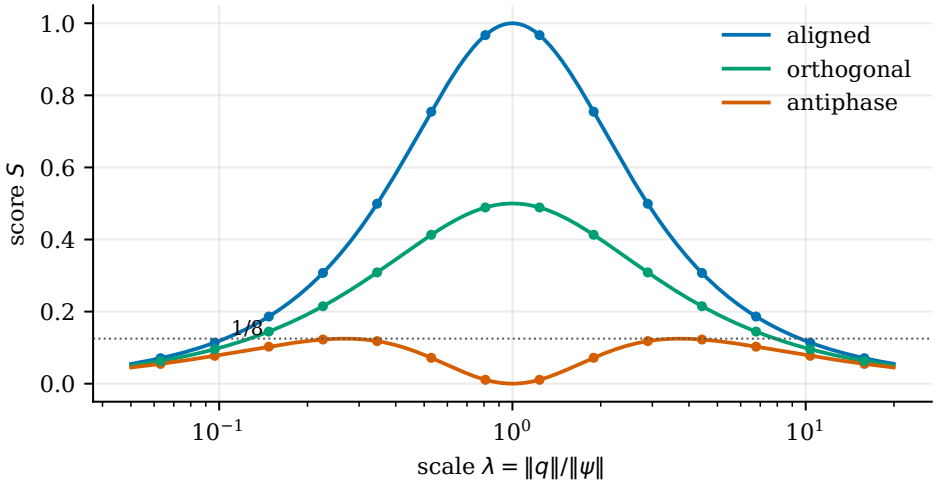


Figure 6.1: Resonance score as a function of the query scale λ for an aligned, orthogonal and antiphase query. Curves are the closed form (B.2); points are the score computed from its definition on random vectors in $\mathbb{C}^8$. The dotted line marks the maximum $1/8$ of the antiphase curve. Computed by the script `figs.py` with seed 8; an illustration of the stated analytic result and not an evaluation of the architecture.

design choice has consequences for any store in which amplitude carries information: it makes retrieval sensitive to the absolute energy of stored traces, which Chapter 29 uses to show that amplitude decay produces a recency bias without any added mechanism.

On the question of metric properties, the source's remark that $1 - S$ does not generally satisfy the triangle inequality can be sharpened. On unit-norm patterns, $1 - S = \frac{1}{2}(1 - \mathrm{Re}\langle\psi, q\rangle) = \frac{1}{4}\|\psi - q\|^2$, a quarter of the squared Euclidean distance, so $\sqrt{1 - S} = \frac{1}{2}\|\psi - q\|$ is a metric on the unit sphere while $1 - S$ itself is not: three coplanar unit vectors at angles 0, $\pi/3$, and $2\pi/3$ give $1 - S$ values $1/4$, $1/4$, and $3/4$ for the two short pairs and the long pair, and $1/4 + 1/4 < 3/4$.

Figure 6.1 shows the closed form of Theorem 6.1 against the definition. The aligned curve peaks at one only when the scales match, the antiphase curve never exceeds $1/8$, and all three vanish as the scales diverge, which is the energy-balance behavior that distinguishes the score from the cosine.

Chapter 7

Phase Transformations and Semantic Operators

The wave-based proposal names four semantic operators and applies them to base terms. Negation, written NEG, is described as aligning a pattern with antiphased opposites. A controlled phase shift, SHIFT+, is described as producing stable offset clusters. Two amplitude operators, INT_UP and INT_DOWN, are described as producing amplitude-driven zones [1]. The text of the source does not give formulas, phase offsets, or scaling rules for any of them, so the operators cannot be reproduced from the paper alone. The monograph therefore adopts candidate definitions, consistent with the one-line descriptions, and states exactly what each implies for the score. The definitions are a modeling choice made here and are not claimed to be the source's.

Definition 7.1 (Candidate operators)**.** For a pattern $\psi \in \mathcal{H}$, a phase offset $\delta \in (0, \pi)$, and a scale $\gamma > 1$, define $\mathrm{NEG}(\psi) = e^{\mathrm{i}\pi}\psi = -\psi$, $\mathrm{SHIFT}^{+}_{\delta}(\psi) = e^{\mathrm{i}\delta}\psi$, $\mathrm{INT_UP}_{\gamma}(\psi) = \gamma\psi$, and $\mathrm{INT_DOWN}_{\gamma}(\psi) = \gamma^{-1}\psi$.

All four are multiplication by a nonzero complex scalar $\lambda = \gamma e^{\mathrm{i}\delta}$, and the family of such multiplications is the multiplicative group $\mathbb{C}^{\times}$ acting on $\mathcal{H}$ by scalars. The composition law $U_g U_h = U_{gh}$ required of a group action holds, and the maps are invertible isometries up to scale. The closed form of Theorem 6.1 then gives the score between a pattern and its image in one line.

Proposition 7.2 (Score of an operator image)**.** *For $\psi \neq 0$ and $\lambda = \gamma e^{\mathrm{i}\delta}$, $S(\psi, \lambda\psi) = \frac{\rho}{2}(1 + \rho\cos\delta)$ with $\rho = 2\gamma/(1+\gamma^2)$. Consequently $S(\psi, \mathrm{NEG}\,\psi) = 0$, $S(\psi, \mathrm{SHIFT}^{+}_{\delta}\psi) = \frac{1}{2}(1 + \cos\delta)$, and $S(\psi, \mathrm{INT_UP}_{\gamma}\psi) = S(\psi, \mathrm{INT_DOWN}_{\gamma}\psi) = \frac{\rho}{2}(1+\rho)$ with the same ρ.*

Proof. Here $\|\lambda\psi\| = \gamma\|\psi\|$, so $\rho = 2\gamma/(1+\gamma^2)$ and $\mathrm{Re}\langle\psi, \lambda\psi\rangle = \|\psi\|^2 \,\mathrm{Re}\,\lambda = \|\psi\|^2\gamma\cos\delta$, giving $c = \cos\delta$. Substitute into (6.4); the values for the four operators follow with $(\gamma, \delta) = (1, \pi), (1, \delta), (\gamma, 0)$, and $(\gamma^{-1}, 0)$, noting $\rho(\gamma) = \rho(\gamma^{-1})$. □

Two consequences are worth stating. First, the score cannot distinguish the amplitude-raising operator from the amplitude-lowering operator with reciprocal factors when each is compared against the unmodified base pattern, since the two images are equally distant in the sense of S. Whether the two can be distinguished at all depends on the presence of a reference scale elsewhere in the store, which is a property of the encoder and not of S. Second, because every candidate operator is a complex scalar, each acts identically on every coordinate pair of the realified space, as a rotation by δ combined with a scaling by γ in each two-dimensional block. This is a one-parameter-pair family inside the real-linear maps of $\mathbb{R}^{2n}$, and a real-valued model that is allowed a single shared rotation-scaling block can represent every one of these operators exactly.

The semantic status of negation requires a separate argument, because the equation $\mathrm{NEG}(\psi) = -\psi$ is a mathematical convenience and not an analysis of negation. Four distinct things have been called negation in the literature on semantics and retrieval. A phase rotation by π is an involution, $\mathrm{NEG} \circ \mathrm{NEG} = \mathrm{id}$, which agrees with the classical law that double negation returns the proposition. It is also a linear isometry, which means it

maps antonyms to a set of patterns determined entirely by the geometry of the original ones: whatever is close to ψ is sent to something far from ψ and close to $-\psi$. This conflates contrary and contradictory opposition. The contrary of hot is cold, which is a particular alternative, whereas the contradictory of hot is not-hot, which is the whole complement, including lukewarm and cold and a great many unrelated things. A single vector $-\psi$ can stand for the contrary and cannot stand for the complement, because the complement contains patterns that are mutually orthogonal and no single point is close to all of them under any inner-product score. Negation understood as set complement is therefore not a point-to-point map but a point-to-set map, and the pointwise rotation is at best a representative of one of its members. Alternatively, negation may be modeled as a context-dependent transformation in which the effect on the pattern depends on what is being negated and why, which suggests a learned operator indexed by context; or as a relation between propositions in a separate relational structure, with the vector layer holding only the proposition content. The monograph treats the choice among these as an open modeling question to be settled by the discrimination experiments of Part IX, and it distinguishes in every case between operators that are invertible isometries, such as phase rotations, and lossy semantic operations, such as projections onto a subspace or collapse of a set of alternatives to a representative, which are idempotent and not injective.

Chapter 8

ResonanceDB as a Baseline Architecture

The software implementation accompanying the wave-based proposal is called ResonanceDB. As described in the source, patterns are stored in memory-mapped binary segments as real-valued amplitude-phase pairs and converted to complex form on demand; the similarity kernel is deterministic and retrieval is an exact scan, so there is no index and no approximation, with cost $O(NL)$ for N patterns of length L [1]. The reported configuration uses pattern lengths $L \in \{512, 1024\}$, dataset sizes from 10^4 to 5×10^5 patterns, and top-k retrieval with $k \in \{1, 10, 100\}$. Execution is on a 12-core CPU at 3.6 GHz with 32 GB of memory under Java 22, cold cache, and a fixed 12-thread executor. Two backends are described, a scalar kernel used for all reported latencies and an experimental vectorized kernel that is excluded from the results. The numerical precision of the stored values is not stated. The source reports that no training is involved, since phases are supplied by the encoder or specified procedurally.

The reported latencies at $L = 1024$ and 5×10^5 patterns, with a heap near 512 MB, are approximately 135 ms mean and 152 ms at the 95th percentile for top-1, 138 ms and 156 ms for top-10, and 145 ms and 163 ms for top-100. The source's text describes latency as scaling linearly with dataset size, while the latency table is reported at one dataset size only, so the scaling claim is not supported by the numbers presented and is treated here as a hypothesis to be tested.

A dimensional check is useful because it indicates what a faithful reproduction must control. A full scan at the reported configuration touches $5 \times 10^5 \times 1024 = 5.12 \times 10^8$ pattern coordinates. If each coordinate is stored as an amplitude and a phase in 32-bit floating point, the scan reads $5.12 \times 10^8 \times 8 = 4.1$ gigabytes, and a mean latency of 135 ms then corresponds to an effective read rate of about 30 GB/s; with 64-bit values the rate doubles. These figures are conditional on the unstated precision and on the whole file being read per query, and they are offered only to show that the result is within the range where memory bandwidth, not arithmetic, is the binding constraint, which suggests that implementation overhead and mathematical retrieval cost are not easily separated in this measurement. A second consideration concerns the arithmetic. Evaluating the interference sum for amplitude-phase storage requires $\cos(\phi_1 - \phi_2)$ for each coordinate, which is a transcendental evaluation when phases are continuous and a sign when phases are restricted to $\{0, \pi\}$. The source's mapping from real embeddings assigns phase 0 to nonnegative coordinates and phase π to negative ones. Whether the reported latency was measured on patterns with continuous phases, and whether the scalar kernel avoids the trigonometric evaluation, cannot be determined from the text, and a latency measured on binary-phase data would not transfer to dense-phase data. The reproduction protocol of Chapter 46 therefore requires latency to be reported separately for binary-phase and continuous-phase stores.

The baseline algorithm is specified here so that every later variant can be defined as a modification of it. In notation, the exact scan computes the score of the query against every stored pattern and returns the indices of the k largest scores.

```
function resonance_topk(store, q, k):
    Eq <- sum_x |q[x]|^2
    heap <- empty min-heap of size k
    for i in 1..N:
        psi <- store[i]                       # amplitude-phase pairs -> complex
        E   <- sum_x |psi[x]|^2
        if E + Eq == 0:  s <- 0
        else:
            ip <- sum_x Re( conj(psi[x]) * q[x] )
            s  <- (E + Eq + 2*ip) * sqrt(E*Eq) / (E + Eq)^2
        push (s, i) onto heap, keeping the k largest
    return heap sorted descending
```

Two separations are made explicit in every later comparison. Mathematical retrieval cost is the count of arithmetic operations and memory reads that the algorithm requires on an idealized machine, here $\Theta(NL)$ reads and multiply-accumulate operations for the exact scan, and implementation overhead is everything else: input and output, memory mapping, thread scheduling, garbage collection, and conversion between storage and compute formats. Approximate indexing changes the first and can change the second in either direction, and an approximate method must be compared with the exact one at matched recall and not at matched latency alone. The energies $E = \sum_x |\psi[x]|^2$ of stored patterns can be cached at write time, which removes one of the three reductions per pattern from the inner loop, and any reported latency should state whether this was done.

Chapter 9

The Representational Equivalence Problem

The claim that phase-aware retrieval offers an advantage over real-valued vector retrieval admits several readings, and the first task is to decide which of them is available. A categorical claim, that complex vectors can express something that real vectors cannot, is false for the following elementary reason.

Theorem 9.1 (Realification)**.** *Define $\iota : \mathbb{C}^n \to \mathbb{R}^{2n}$ by $\iota(z) = (\operatorname{Re} z, \operatorname{Im} z)$. Then ι is an $\mathbb{R}$-linear bijection with $\langle \iota z, \iota w \rangle_{\mathbb{R}} = \operatorname{Re}\langle z, w \rangle_{\mathbb{C}}$ for all z, w, so $\mathbb{C}^n \cong \mathbb{R}^{2n}$ as real inner-product spaces. Multiplication by i corresponds under ι to the orthogonal matrix $J = \left(\begin{smallmatrix} 0 & -I \\ I & 0 \end{smallmatrix}\right)$, which satisfies $J^2 = -I$. A real-linear map A of $\mathbb{R}^{2n}$ is of the form $\iota \circ B \circ \iota^{-1}$ for a complex-linear map B of $\mathbb{C}^n$ if and only if $AJ = JA$.*

Proof. Linearity and bijectivity are immediate. For $z = x + \mathrm{i}y$ and $w = u + \mathrm{i}v$, $\overline{z}\, w = (xu + yv) + \mathrm{i}(xv - yu)$ in each coordinate, so $\operatorname{Re}\langle z, w \rangle = \sum_k (x_k u_k + y_k v_k) = \langle \iota z, \iota w \rangle$. Multiplication by i sends $x + \mathrm{i}y$ to $-y + \mathrm{i}x$, hence $(x, y) \mapsto (-y, x) = J(x, y)$, and $J^2(x, y) = (-x, -y)$. If B is complex-linear then $B(\mathrm{i}z) = \mathrm{i}B(z)$, which reads $AJ = JA$ for $A = \iota B \iota^{-1}$. Conversely, an $\mathbb{R}$-linear A commuting with J commutes with multiplication by i and is therefore $\mathbb{C}$-linear. □

Corollary 9.2 (Dimension count for the inductive bias)**.** *The space of complex-linear maps of $\mathbb{C}^n$ has real dimension $2n^2$, while the space of real-linear maps of $\mathbb{R}^{2n}$ has real dimension $4n^2$. A complex-valued linear layer is therefore a real layer with the constraint $AJ = JA$, which halves the number of free parameters.*

The relationship between the score and the realification is immediate from Corollary 6.2: since ρ and c are functions of norms and of the real inner product, $S(z, w) = S_{\mathbb{R}}(\iota z, \iota w)$ where $S_{\mathbb{R}}(x, y) = \frac{\rho}{2}(1 + \rho \langle x, y \rangle / (\|x\| \|y\|))$ is the same formula on $\mathbb{R}^{2n}$. Retrieval over a complex store by S is retrieval over the realified store by $S_{\mathbb{R}}$, with identical rankings.

The comparison made in the source between resonance retrieval and cosine retrieval must therefore be read as a comparison between feature maps, norm treatment, and operators, not as one between number systems. Four candidate sources of advantage can be distinguished, and each has a different experimental signature. The first is the energy-balance factor ρ, which is an amplitude-sensitivity that cosine discards and which could be added to a real-valued score without changing the representation. The second is the feature map. The source's mapping from real embeddings v sets $A(x) = |v(x)|$ and $\phi(x) = 0$ or π according to the sign of $v(x)$, so the complex pattern equals $v(x)$ exactly, a real vector regarded as an element of $\mathbb{C}^n$.

Proposition 9.3 (Sign-phase patterns are real vectors)**.** *If $\psi(x) = |v(x)| e^{\mathrm{i}\phi(x)}$ with $\phi(x) = 0$ for $v(x) \geq 0$ and $\phi(x) = \pi$ otherwise, then $\psi(x) = v(x)$ for every x. For two such patterns ψ_u, ψ_v generated from real vectors u, v, the phase-aware cosine equals the ordinary cosine of u and v, and $S(\psi_u, \psi_v) = \frac{\rho}{2}(1 + \rho \cos(u, v))$.*

Proof. For $v(x) \geq 0$, $|v(x)| e^0 = v(x)$; for $v(x) < 0$, $|v(x)| e^{\mathrm{i}\pi} = -|v(x)| = v(x)$. The inner product of two real-valued complex vectors is their real inner product, so $c = \cos(u, v)$. □

On data generated this way, the only difference between S and cosine is the balance factor ρ, and for unit-norm embeddings there is no difference at all in ranking. Any measured advantage of resonance over cosine on such data must come either from ρ acting on unequal norms or from the operators, which are the transformations that introduce phases other than 0 and π. The third source is the family of operators, which as shown in Chapter 7 are complex scalars and so are a structured, low-dimensional subfamily of real-linear maps. A cosine baseline applied to the original real vectors cannot represent a phase offset that the real vectors do not contain, but a baseline applied to the realified vectors can, so a comparison that feeds the baseline only the original embeddings of the base terms and the complex model the operator-transformed patterns is a comparison of information supplied, not of retrieval rules. The fourth source is numerical organization and learning dynamics: the shared structure of complex arithmetic can improve optimization or conditioning when parameters are learned, which is the argument advanced in the literature on complex-valued neural networks [5], and which cannot arise for a retrieval rule without training.

Theorem 9.4 (Matched-baseline principle)**.** *Let $\mathcal{D}$ be a dataset of complex patterns and queries and let any retrieval task be defined by the set of rankings produced by S. Then the realified dataset $\iota(\mathcal{D})$ with the score $S_{\mathbb{R}}$ produces identical rankings. Consequently, a statement that resonance retrieval outperforms a real-valued method on $\mathcal{D}$ is equivalent to a statement that $S_{\mathbb{R}}$ on $\mathbb{R}^{2n}$ outperforms that method on $\iota(\mathcal{D})$, and no property of complex numbers is involved beyond the construction of $\mathcal{D}$.*

The protocol consequence is a requirement of matched capacity. A complex pattern of length L carries $2L$ real degrees of freedom, so the real-valued baselines in Part IX are of dimension $2L$ and are given the same encoder outputs, together with versions given only the original embeddings, so that the contribution of the operator-generated phases can be separated from the contribution of the retrieval rule. This chapter is the place where the monograph forecloses the claim of intrinsic mathematical superiority of complex embeddings as a premise. What remains possible is that the inductive bias of a structured operator family, the numerical behavior under learning, or the interpretability of phase as a relational variable provides practical advantages, and those are empirical hypotheses to be tested against matched baselines.

Part III

Distributed Oscillatory Memory

Chapter 10

Empirical Foundations of Macroscale Resonance

The empirical motivation for treating distributed temporal coordination as a candidate substrate of memory comes from the study of Cabral, Fernandes, and Shemesh, which recorded functional magnetic resonance imaging from rat brains at a temporal resolution far finer than is customary and analyzed the resulting signals as superpositions of oscillatory modes [22]. The design, as reported in the published article, was as follows. Six genetically similar female rats were each scanned twice in each of three conditions, giving twelve scans per condition, with ten-minute scans acquired at 9.4 tesla with a temporal resolution of 38 milliseconds, a sampling rate of 26.3 Hz. The three conditions were defined by isoflurane concentration on a background of medetomidine sedation: none, which the authors call sedation, one percent, called light anesthesia, and three percent, called deep anesthesia. Signals from about fourteen hundred brain voxels in a coronal slice were band-pass filtered below 0.3 Hz, and scans of a post-mortem rat served as a baseline against which power and component variance were judged. Principal component analysis of the covariance of the filtered signals yielded, under sedation, ten components above the post-mortem baseline whose spatial patterns showed fixed phase relationships across distinct cortical and subcortical subsystems, varying gradually and symmetrically across space, with regions oscillating in phase or in antiphase with each other, which the authors describe as characteristic of standing waves. The temporal signatures of these components, the time courses of their amplitudes, showed transient oscillations with fluctuating amplitude, and the recorded signals could be reconstructed as the linear superposition of the ten components scaled by their temporal signatures. Power extended up to about 0.3 Hz, peaked between 0.20 and 0.25 Hz in cortical voxels under medetomidine and between 0.10 and 0.15 Hz in the striatum, and was modulated by anesthetic level. With one percent isoflurane only six components remained above baseline, with altered spatial configuration, shorter-lived oscillations, and power peaking at lower frequencies, and at three percent the variance above baseline was captured by a single component in which cortex and striatum moved together in phase at very low frequency, consistent with the authors' statement that only global aperiodic fluctuations remain under deep anesthesia. The resonance quality factor of the oscillations, which counts the cycles before the autocorrelation decays to $1/e$, was highest under medetomidine and decreased with isoflurane. An extension to a whole-brain acquisition with lower temporal resolution, repetition time 350 ms across twelve slices in three rats, observed oscillations with overlapping phase relationships.

The authors interpret the findings with a mechanistic model in which the spatial modes are damped resonators operating in the subcritical range of a Hopf bifurcation and driven by background noise, the stochastic resonance of standing waves, whose superposition reproduces the sliding-window correlation structure of long-range functional connectivity. This regime is relevant to the specification below, because it is a damped, noise-driven regime in which the uniform state is stable and oscillations are transient responses, and not the self-sustained regime in which a continuum oscillator model would oscillate autonomously. The distinction is taken up in Chapter 13.

The observations are strong evidence for one claim and no evidence for several that

are adjacent to it. They support the statement that spontaneous activity measured by fMRI in sedated rats is well described by a small repertoire of spatial patterns whose amplitudes oscillate transiently, with phase relationships that are stable within a condition and that depend on the sedative state. They do not establish that those patterns carry information from one region to another, which is a claim about causal transmission that requires interventions or directed measures the study does not provide. They do not establish a binding mechanism, which would require showing that the phase relationships select and group the content represented in the participating regions. They do not establish a role in memory: the recordings were made at rest, under sedation or anesthesia, in six animals, and no encoding or recall of any episode was involved. The authors themselves note that the study did not attempt to separate neural activity from neurovascular coupling. They observe that the temporal signatures do not follow the canonical hemodynamic response, that the frequencies lie above the range typically associated with the blood-oxygenation-level-dependent signal and below cardiac and respiratory rhythms, and that earlier work found coherence between local field potentials and fMRI in this range, and they conclude that a more direct neuronal contribution cannot be excluded. The relationship of these rhythms of a few tenths of a hertz to the theta and gamma rhythms usually invoked in models of hippocampal and cortical memory [16, 23], which are more than an order of magnitude and about two orders of magnitude faster respectively, is therefore an open empirical question and not an identification. The components obtained by principal component analysis are, finally, defined by variance and orthogonality, which are properties of the method, so the interpretation of each component as a physical standing wave needs the additional tests discussed in Chapter 11, some of which the authors carry out through the post-mortem control and the comparison with spatially randomized patterns.

For the purposes of this monograph the study is treated as motivation of a modeling stance: macroscale activity can be organized by a small number of coordinated oscillating patterns whose phase relationships are stable and condition-dependent, and the number, frequency, and damping of the modes are state-dependent quantities. The stance suggests that a memory architecture should be allowed to use phase coordination across distributed components, and it identifies the quantities a model should expose to experiment, namely mode counts, mode frequencies, quality factors, and phase relationships, as functions of a state variable analogous to anesthetic depth. It does not constrain the mechanism by which such coordination would subserve recollection, and every architectural claim that goes beyond the stance is labeled in Chapter 54 as a hypothesis.

Chapter 11

Spectral Decomposition of Distributed Activity

Let a distributed field be given on a finite index set $\mathcal{X}$ of locations by

$$\Psi(x,t) = \sum_{k=1}^{K} a_k(t)\,\varphi_k(x), \tag{11.1}$$

where the φ_k are spatial modes and the a_k are time-dependent coefficients. Which modes to use and how to estimate them are separate questions with separate answers. Laplacian eigenmodes are obtained from a graph or mesh on $\mathcal{X}$ with symmetric nonnegative weights W, degree matrix D, and Laplacian $L = D - W$. The matrix L is symmetric positive semidefinite and has an orthonormal eigenbasis $L\varphi_k = \lambda_k\varphi_k$ with $0 = \lambda_1 \le \lambda_2 \le \cdots$, and the graph Fourier transform of a signal f on $\mathcal{X}$ is the vector of coefficients $\hat{f}_k = \langle \varphi_k, f\rangle$. A wave equation $\partial_t^2\Psi = -c^2 L\Psi$ on the graph has the standing-wave solutions $\Psi(x,t) = \varphi_k(x)\cos(c\sqrt{\lambda_k}\,t + \alpha)$, each with the fixed spatial profile of an eigenmode and angular frequency $c\sqrt{\lambda_k}$, so geometry alone supplies a family of candidate resonant modes whose frequencies are determined by the spectrum of the connectivity. Principal component analysis, in contrast, chooses spatial modes that maximize explained variance of the observed data, with no reference to a generative dynamics. Dynamic mode decomposition fits a linear propagator $x_{t+1} \approx Ax_t$ to the data and uses the eigenpairs (μ_j, ϕ_j) of A, with $\mu_j = e^{(\sigma_j + \mathrm{i}\omega_j)\Delta t}$, so that the growth rate σ_j and the angular frequency ω_j of each mode are read directly from the eigenvalue and the spatial structure, including its phase map, from the eigenvector.

The distinction that matters for the use of the word oscillatory is the one between standing and traveling spatial organization.

Proposition 11.1 (Standing versus traveling modes)**.** *Let $\phi = u + \mathrm{i}v \in \mathbb{C}^N$ and $x(t) = \mathrm{Re}(\phi\, e^{\mathrm{i}\omega t}) = u\cos\omega t - v\sin\omega t$. Then $x(t)$ is a standing wave, meaning that $x(t)$ lies on a fixed line through the origin for all t, if and only if $\phi = e^{\mathrm{i}\alpha}w$ for a real vector w and a real angle α, equivalently if and only if the phase differences $\arg\phi_j - \arg\phi_k$ lie in $\{0,\pi\}$ modulo 2π for all j,k at which both are nonzero. If u and v are linearly independent, $x(t)$ traverses an ellipse in the plane they span and the spatial pattern changes shape over the period.*

Proof. If $\phi = e^{\mathrm{i}\alpha}w$ then $x(t) = w\cos(\omega t + \alpha)$, which lies on the line through w. Conversely, if $x(t)$ stays on a line for all t then u and v are linearly dependent, so $v = \kappa u$ or $u = 0$, and then $\phi = u(1 + \mathrm{i}\kappa) = \sqrt{1+\kappa^2}\,e^{\mathrm{i}\arctan\kappa}u$, or $\phi = \mathrm{i}v$, each of the stated form. The phase-difference statement holds because $\phi_j/\phi_k = w_j/w_k$ is real. When u, v are independent, $x(t)$ is a nondegenerate ellipse. □

A fixed phase relationship across distant locations that is either zero or π, as reported for the modes of the rat study, is therefore exactly the signature of a standing wave in this sense, and a traveling wave requires a phase map that varies continuously across locations. A principal component pair in quadrature is the typical PCA signature of a traveling wave, since a single traveling wave occupies two components with equal variance and a quarter-period offset.

The legitimacy of calling a spatial mode an oscillatory mode requires conditions beyond the existence of the decomposition. Principal components of any stationary random process with a spectrum concentrated near a frequency will show peaked power spectra, so a spectral peak in a component is not by itself evidence of a generating oscillator. The conditions adopted here are four. The mode must correspond to a complex-conjugate pair of eigenvalues of a fitted propagator with $|\mu|$ close to one, and with ω separated from its neighbors in the spectrum by more than the estimation error. The mode must be reproducible across disjoint data segments within a condition. The spectral peak must exceed that of surrogate data that preserve the second-order statistics of the original, for example the covariance matrix and the power spectrum of each channel, while randomizing the cross-channel phase relationships. And the spatial phase map must conform to the standing-wave criterion of Proposition 11.1 within estimation error, or be explicitly reported as traveling. Only modes passing these tests are used in Mem8 as oscillatory modes, and modes failing them are used, if at all, as statistical components.

Chapter 12

Coupled Oscillators and Phase Synchronization

A minimal model of distributed phase coordination assigns a phase θ_i to each of N oscillators and couples them by a sinusoidal interaction,

$$\dot{\theta}_i = \omega_i + \sum_j K_{ij} \sin(\theta_j - \theta_i - \delta_{ij}) + u_i(t), \tag{12.1}$$

where ω_i is a natural frequency, K_{ij} a coupling strength, δ_{ij} a phase lag, and u_i an external drive. For the case of identical natural frequencies, symmetric coupling $K_{ij} = K_{ji}$, no lag, and no drive, the model is a gradient system, and its behavior is governed by a potential.

Theorem 12.1 (Gradient structure of the Kuramoto system)**.** *Let $\omega_i \equiv \omega$, $\delta_{ij} = 0$, $u \equiv 0$, and K symmetric. In the rotating frame $\vartheta_i = \theta_i - \omega t$, system* (12.1) *is $\dot{\vartheta} = -\nabla V(\vartheta)$ with*

$$V(\vartheta) = -\frac{1}{2} \sum_{i,j} K_{ij} \cos(\vartheta_j - \vartheta_i). \tag{12.2}$$

Consequently V is nonincreasing along trajectories, with $\dot{V} = -\sum_i \dot{\vartheta}_i^2$, every trajectory converges to the set of critical points of V, and, since V is real-analytic, to a single critical point. For the explicit Euler discretization $\vartheta^{k+1} = \vartheta^k - \Delta t \, \nabla V(\vartheta^k)$, V is nonincreasing whenever $\Delta t < 1/\max_i \sum_j |K_{ij}|$.

Proof. Differentiating, $\partial V/\partial \vartheta_i = -\sum_j K_{ij} \sin(\vartheta_j - \vartheta_i)$, using the symmetry of K to combine the two terms in which ϑ_i appears. Hence $\dot{\vartheta}_i = -\partial V/\partial \vartheta_i$ and $\dot{V} = \sum_i (\partial V/\partial \vartheta_i)\dot{\vartheta}_i = -\sum_i \dot{\vartheta}_i^2 \le 0$. The state space is a torus, hence compact, so the omega-limit set is nonempty and consists of critical points by LaSalle's principle, and convergence to a single point follows from the Łojasiewicz gradient inequality for analytic potentials. For the discretization, the Hessian of V has entries bounded by Gershgorin's theorem, giving $\|\nabla^2 V\| \le 2 \max_i \sum_j |K_{ij}| =: \Lambda$, so ∇V is Λ-Lipschitz, and the descent lemma gives $V(\vartheta^{k+1}) \le V(\vartheta^k) - \Delta t(1 - \Delta t \Lambda/2)\|\nabla V(\vartheta^k)\|^2$, which is a decrease for $\Delta t < 2/\Lambda = 1/\max_i \sum_j |K_{ij}|$. □

The theorem is the structural basis for the phase block of the minimal model in Appendix L, whose update is exactly the Euler step above with zero natural-frequency spread. A deterministic check on a random symmetric coupling matrix with $N = 30$ oscillators and a step 0.9 times the bound found V nonincreasing at every one of 400 steps and the order parameter equal to one to numerical precision at the end (Appendix L); this checks the statement and does not test any memory property.

Synchronization is not automatically beneficial. The all-to-all case with identical frequencies and uniform positive coupling $K_{ij} = K/N$ is the extreme: it is known that the fully synchronized state is the only stable equilibrium and that the set of initial conditions not converging to it has measure zero [47, 42, 21, 43]. In that regime the final phase pattern is the same for every generic initial pattern, so the map from initial relative phases to final relative phases carries no information about the initial pattern, and any code that stores

differences among traces in the relative phases of the oscillators is erased. Excess global coupling therefore reduces selectivity, because it collapses the very distinctions the phases were meant to carry, and it increases interference, because a single global phase pattern is shared by every trace read through it. The opposite extreme, no coupling, leaves the phases independent and provides no coordination. The useful regime lies between: structured coupling K_{ij} that is strong within coherence clusters and weak between them, so that the dynamics has several stable or metastable cluster states, each of which can serve as a context for a different subset of traces. Metastability, meaning the persistence of one cluster configuration for a long time followed by a transition to another, arises in the presence of heterogeneity of natural frequencies, delays, or noise, and the transitions between cluster states are the mechanism by which the system can move among dynamically accessible memory states. The monograph measures cluster structure by order parameters of subsets, $R_{\mathcal{S}} = |\frac{1}{|\mathcal{S}|}\sum_{i\in\mathcal{S}} e^{\mathrm{i}\theta_i}|$, and by pairwise phase-locking values between oscillators.

Chapter 13

Oscillatory Modes as a Computational Substrate

A continuum model of coordinated oscillation is supplied by the complex Ginzburg–Landau equation, written here with a drive I as

$$\partial_t \Psi = (D + \mathrm{i}\Omega)\Delta\Psi + (\alpha + \mathrm{i}\omega)\Psi - (\beta + \mathrm{i}\chi)|\Psi|^2\Psi + I(x,t). \tag{13.1}$$

The equation is treated as a proposed Mem8 model and not as an equation established by the neuroscience; its coefficients are free parameters of the specification, and the study discussed in Chapter 10 gives no basis for assigning them. The equation is the normal form near a Hopf bifurcation of a spatially extended medium [19], which is the reason it is a natural candidate substrate: it is the generic description of a distributed system of weakly nonlinear oscillators with diffusive coupling.

Two regimes of (13.1) need to be kept apart because the empirical study fixes only one of them. For $\alpha > 0$ the uniform state is unstable and the medium oscillates autonomously, which is the supercritical regime treated in Proposition 13.1. For $\alpha < 0$ the uniform state is stable, the linearization about it has damped oscillatory modes with decay rates set by $\alpha - Dk^2$ and frequencies set by $\omega - \Omega k^2$, and oscillations appear only as transient responses to a drive I or to noise, with a quality factor that grows as $\alpha \to 0^-$. The model of Cabral and colleagues is in the second regime, a subcritical range of a Hopf bifurcation with noise-driven resonance [22], so the supercritical regime is a design choice of Mem8 for which the study offers no support, while the damped regime is the one in which a cue can be treated as a resonant drive and the reconstruction as the ringing of the mode that the cue excites. Both are carried through the specification, and the experiments of Chapter 51 include the damped regime as a baseline.

Four features of its solution set bear on memory. The first is the family of plane-wave solutions.

Proposition 13.1 (Plane waves)**.** *For $I \equiv 0$, $D > 0$, and $\beta > 0$, equation* (13.1) *has the solutions $\Psi(x,t) = A\, e^{\mathrm{i}(k\cdot x - \nu t)}$ with*

$$A^2 = \frac{\alpha - Dk^2}{\beta}, \qquad \nu = \Omega k^2 - \omega + \chi A^2, \tag{13.2}$$

which exist when $Dk^2 < \alpha$. The case $k = 0$ is a spatially uniform oscillation with $A^2 = \alpha/\beta$.

Proof. For this ansatz $\Delta\Psi = -k^2\Psi$ and $|\Psi|^2 = A^2$. Substituting, the left side is $-\mathrm{i}\nu\Psi$ and the right side is $[-(D + \mathrm{i}\Omega)k^2 + (\alpha + \mathrm{i}\omega) - (\beta + \mathrm{i}\chi)A^2]\Psi$. The real part gives $0 = -Dk^2 + \alpha - \beta A^2$, and the imaginary part gives $-\nu = -\Omega k^2 + \omega - \chi A^2$. □

The second feature is stability. After rescaling, the uniform oscillation is stable to long-wavelength perturbations when $1 + (\Omega/D)(\chi/\beta) > 0$, the Benjamin–Feir–Newell criterion [19]; when the criterion fails the uniform state gives way to phase turbulence or to defect-mediated turbulence, which would correspond to unstable reconstruction. The third is the structure of the non-oscillatory subcase. When $\Omega = \chi = \omega = 0$ and I is time-independent,

equation (13.1) is the gradient flow $\partial_t \Psi = -\delta \mathcal{F} / \delta \overline{\Psi}$ of the energy

$$\mathcal{F}[\Psi] = \int \Big(D|\nabla \Psi|^2 - \alpha |\Psi|^2 + \tfrac{\beta}{2} |\Psi|^4 - \big(\bar{I} \Psi + I \overline{\Psi} \big) \Big) dx, \tag{13.3}$$

whose local minima are stationary attractors, so the non-oscillatory medium behaves as an energy-minimizing associative memory of Hopfield type, with fixed points playing the role of stored patterns and the drive I playing the role of a cue. The oscillatory terms Ω, χ, ω are the non-gradient terms: they are what allows propagating waves, limit cycles, and phase-coded states, and they are also what removes the guarantee that the dynamics settles. Whether these terms improve reconstruction under matched resources is exactly the question posed by the ablation experiments, and the formal separation just given makes the experiment well posed, since the nonoscillatory implementation required in Chapter 39 is the gradient subcase. The fourth feature concerns localized solutions and boundary conditions: with a spatially localized drive the medium supports localized structures, which would represent individual traces, and the choice of boundary condition, periodic, no-flux, or absorbing, determines which modes are admissible in the way that the geometry determines Laplacian eigenmodes in Chapter 11. Mode competition arises when two or more modes have positive growth rates and nonlinear saturation couples them, so that the amplitude of one suppresses the others; this is the continuum analogue of the softmax selection used in the discrete architecture.

Chapter 14

Temporal Binding and Phase-Coded Associations

A memory of an episode requires that items encountered at different times be associated in a way that preserves what is needed about their temporal relationship. Four relationships need to be distinguished because they are routinely run together. Temporal order is the relation $t_a < t_b$. Simultaneity is $|t_a - t_b| < \epsilon$ for a resolution ϵ. Causal dependence is a relation of intervention, in which the occurrence of a changes the occurrence or character of b. Mere coactivation is that a and b were represented at the same time in the system, for any reason. The first two are functions of timestamps, whereas the third and fourth are not, and the difference has a precise consequence: no encoding that depends only on timestamps can distinguish two histories that differ only in their causal structure. Two histories in which a causes b and in which a common cause produces both at the same times have identical timestamps, so by the sufficiency criterion of Chapter 32 a timing-only memory cannot support the recollection function that returns the causal structure.

Phase offers a way to encode temporal order and interval. Associate with each item a phase variable that advances at a carrier angular frequency ω, and record at encoding the phase of the item. The phase difference between two items encoded at times t_a and t_b is $\omega(t_b - t_a)$ modulo 2π, so the lag is recovered only modulo the period $T = 2\pi/\omega$: a single frequency yields an aliased measurement of lag. The aliasing is resolved by a bank of frequencies.

Proposition 14.1 (Multifrequency phase code)**.** *Let lags be integers $\ell \in \{0, \dots, L-1\}$ in units of a time step, and let $T_1, \dots, T_K$ be pairwise coprime integer periods with $\prod_k T_k \geq L$. The phase differences $\omega_k \ell \bmod 2\pi$ with $\omega_k = 2\pi/T_k$ determine ℓ uniquely. Conversely, for any set of periods whose least common multiple is smaller than L, two distinct lags yield identical phase codes.*

Proof. The phase difference at frequency ω_k determines $\ell \bmod T_k$. By the Chinese remainder theorem the residues $\ell \bmod T_k$ for pairwise coprime moduli determine ℓ modulo $\prod_k T_k$, hence uniquely in $\{0, \dots, L-1\}$. If $\operatorname{lcm}(T_1, \dots, T_K) = \Lambda < L$, then lags 0 and Λ both lie in $\{0, \dots, L-1\}$ and agree modulo every T_k. □

Static positional encodings in sequence models, which use sines and cosines at a geometric ladder of frequencies [2], are phase codes of exactly this kind, assigned by the position index and not generated by dynamics. This observation fixes the right test for whether phase carries information beyond a static positional encoding. A memory in which the phases are produced by interaction of the oscillators with the input history, and in which retrieval is read from the relative phases, must be compared with a memory that receives the same positional codes as inputs. Two controls are specified. A phase-shuffle control randomizes the assignment of phases to events while preserving their marginal distribution; if sequence reconstruction is unchanged, phase was not being used. A frozen-phase control fixes the oscillator phases to their initial values; if performance is unchanged, the dynamics were not contributing. These are the controls on which the ablation program of Chapter 51 is based.

The reconstruction of an ordered sequence from such a code proceeds by reading relative

phases of the stored items at the carrier frequencies and sorting by the decoded lag. The decoded lag error under phase noise of standard deviation σ_k at frequency ω_k is of order σ_k / ω_k for the lowest frequency that resolves the coarse order, and the highest frequency in the bank sets the fine resolution but is the most sensitive to noise in the same ratio, which gives the standard coarse-to-fine tradeoff of residue-number codes. The experimental protocol of Chapter 48 reports reconstruction error as a function of phase noise for this reason.

Part IV

Ecphoric Reconstruction

Chapter 15

The Ecphoric Interaction Operator

Recall is defined in this monograph as a conditional dynamical process and not as a ranking. Given a cue c_t, persistent memory M_t, a context Z_t, and the current field and phase state Ψ_t, the reconstruction is

$$r_t = \mathcal{E}(c_t, M_t, Z_t, \Psi_t), \tag{15.1}$$

where $\mathcal{E}$ is realized as the terminal state of a reconstruction trajectory $r^{(0)}, r^{(1)}, \dots$ generated by an iteration $r^{(k+1)} = \mathcal{T}(r^{(k)}, c_t, M_t, Z_t, \Theta^{(k)})$, with $\mathcal{E}$ the limit when it exists and the stopped iterate $r^{(K)}$ under a budget otherwise. The explicit trajectory, as opposed to a single ranking operation, has three uses. It makes the dependence on the dynamical state, the phases in particular, an object of analysis instead of a hidden implementation detail. It allows the intermediate states to be inspected, which is how false attractors and premature convergence are diagnosed. And it allows the cost of recollection to be accounted for as the number of steps and the work per step, which is the quantity that Chapter 25 trades against accuracy.

The first question is under what conditions a partial cue initiates convergence toward the historically associated state. Let $x^\star$ denote the target, the organization that the trace was formed to preserve, and suppose that for a fixed cue the iteration is a self-map $F_c(\cdot) = \mathcal{T}(\cdot, c, M, Z, \Theta)$ of a complete metric space with Lipschitz constant $\lambda < 1$ in its first argument. Then there is a unique fixed point $r^\star(c)$ and geometric convergence, and two statements bear on the target.

Theorem 15.1 (Convergence to a cue-determined state and its distance from the target)**.** *Let F_c be a λ-contraction on a complete metric space $(\mathcal{R}, d)$ with $\lambda < 1$. Then (i) F_c has a unique fixed point $r^\star(c)$ and $d(r^{(k)}, r^\star) \le \frac{\lambda^k}{1-\lambda} d(r^{(1)}, r^{(0)})$ for every initialization; (ii) for every $x \in \mathcal{R}$, $d(x, r^\star) \le d(x, F_c(x))/(1-\lambda)$, so the distance from the fixed point to the target is bounded by the residual $d(x^\star, F_c(x^\star))$ of the target under the cue-driven map; and (iii) if $c \mapsto F_c(r)$ is L_c-Lipschitz uniformly in r, then $d(r^\star(c), r^\star(c')) \le \frac{L_c}{1-\lambda}\, d(c, c')$.*

Proof. Part (i) is the Banach fixed-point theorem with its standard a priori estimate. For (ii), $d(x, r^\star) \le d(x, F(x)) + d(F(x), F(r^\star)) \le d(x, F(x)) + \lambda d(x, r^\star)$ and rearrangement gives the bound. For (iii), $d(r^\star(c), r^\star(c')) = d(F_c(r^\star(c)), F_{c'}(r^\star(c'))) \le d(F_c(r^\star(c)), F_{c'}(r^\star(c))) + d(F_{c'}(r^\star(c)), F_{c'}(r^\star(c'))) \le L_c d(c, c') + \lambda d(r^\star(c), r^\star(c'))$. □

Part (ii) is the precise sense in which a cue initiates convergence toward an historically associated state: the cue-driven map must have the target as an approximate fixed point, with residual small relative to $1 - \lambda$. Contraction ensures convergence to a fixed point and says nothing by itself about whether that fixed point is the target; the answer is entirely in the residual. Part (iii) gives the stability of recollection under cue perturbation, so that small changes of the cue produce small changes of the reconstruction, with a constant that blows up as $\lambda \to 1$. A cue that is too weak or too ambiguous makes the residual large, and then the unique fixed point may lie far from the target, which is the formal content of a spurious recollection.

Chapter 16

Associative Completion and Attractor Geometry

Completion from incomplete observations is modeled by recurrent dynamics whose fixed points are intended to coincide with stored patterns. The continuous associative memory of Ramsauer and colleagues provides the cleanest instance for the purposes of analysis [18]. For stored unit traces $m_1, \dots, m_P$ in the realified space and an inverse temperature $\beta > 0$, the update map is

$$F(r) = \sum_{i=1}^{P} w_i(r)\, m_i, \qquad w_i(r) = \frac{e^{\beta \langle m_i, r \rangle}}{\sum_j e^{\beta \langle m_j, r \rangle}}, \tag{16.1}$$

with $\langle \cdot, \cdot \rangle$ the real inner product, and its stationary points are those of the energy $\mathcal{E}(r) = \frac{1}{2}\|r\|^2 - \beta^{-1} \log \sum_j e^{\beta \langle m_j, r \rangle}$, since $\nabla \mathcal{E}(r) = r - F(r)$. The relation of this update to Bayesian reconstruction is stated in Proposition 21.2, where it appears as a monotone algorithm for a posterior mode. The geometry of the basins can be described directly. Fixed-point, limit-cycle, and metastable attractors need to be distinguished. A fixed point corresponds to a single stable reconstruction. A limit cycle, which cannot occur for the gradient dynamics above but can occur with the non-gradient terms of Chapter 13, corresponds to a recollection that is a cyclically evolving state, which may represent a temporally extended episode replayed in a loop. A metastable attractor is a state in which the dynamics lingers for a long time before leaving, and corresponds to an uncommitted or blended recollection.

Proposition 16.1 (Basin condition and one-step cleanup)**.** *Let $\|r\| \le 1$ and suppose that for some index 1 one has $\langle m_1, r \rangle \ge \langle m_i, r \rangle + \Delta$ for all $i \ne 1$. Then the weight mass outside trace 1 satisfies $\varepsilon := \sum_{i \ne 1} w_i(r) \le (P-1) e^{-\beta \Delta}$ and*

$$\|F(r) - m_1\| \le 2\varepsilon \le 2(P-1) e^{-\beta \Delta}. \tag{16.2}$$

Moreover, with $\mu = \max_{i \ne j} \langle m_i, m_j \rangle < 1$, the gap at the new state satisfies $\langle m_1, F(r) \rangle - \langle m_i, F(r) \rangle \ge (1 - \mu) - \varepsilon(3 - \mu)$ for each $i \ne 1$.

Proof. Each $w_i / w_1 = e^{\beta(\langle m_i, r \rangle - \langle m_1, r \rangle)} \le e^{-\beta \Delta}$ for $i \ne 1$, so $\varepsilon = \sum_{i \ne 1} w_i \le (1 - \varepsilon)(P - 1) e^{-\beta \Delta} \le (P-1) e^{-\beta \Delta}$. Then $F(r) - m_1 = \sum_{i \ne 1} w_i (m_i - m_1)$ and $\|m_i - m_1\| \le 2$. For the gap, $\langle m_1 - m_i, F(r) \rangle = \sum_j w_j \langle m_1 - m_i, m_j \rangle \ge w_1 (1 - \mu) - \sum_{j \ne 1} w_j \cdot 2 \ge (1 - \varepsilon)(1 - \mu) - 2\varepsilon$, using $\langle m_1 - m_i, m_j \rangle \ge -2$ for the remaining terms and $\langle m_1 - m_i, m_1 \rangle = 1 - \langle m_i, m_1 \rangle \ge 1 - \mu$. $\square$

The proposition says that if a state is separated from competitors by a margin Δ larger than $\beta^{-1} \log(P - 1)$ by a sufficient amount, then one update step places the state within an exponentially small distance of the stored trace, and the margin at the new state is of the order of $1 - \mu$, which exceeds the original margin requirement. The corresponding cue thresholds follow: the cue must supply enough correlation with the target trace, relative to its correlation with all other traces, to exceed $\beta^{-1} \log\big((P-1)/\varepsilon\big)$. False attractors arise in two ways. At low β the weights are nearly uniform for a wide range of states, $F(r)$ approaches the mean of the traces, and the dynamics has a blend state near the centroid of the store; as β increases this state loses stability through the criterion of Theorem 33.2 of Chapter 33 and is replaced by metastable states between clusters of similar traces and

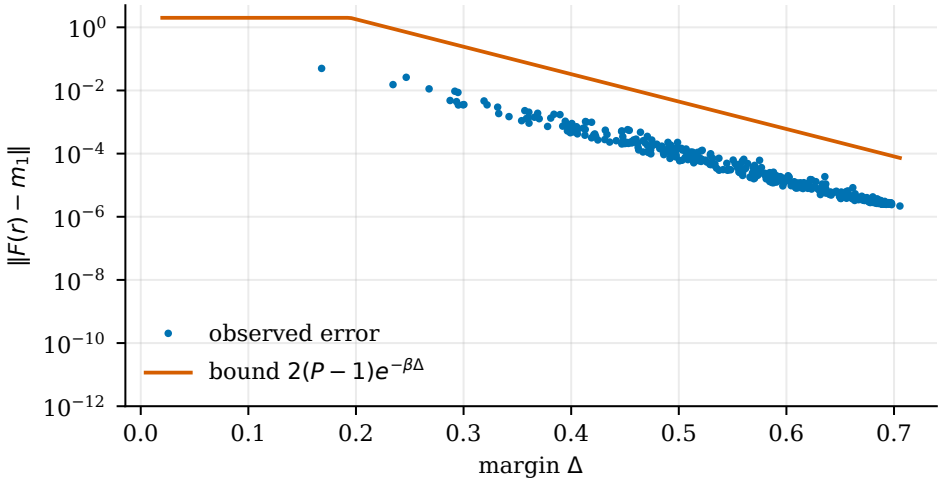


Figure 16.1: One-step cleanup error of the softmax update against the margin Δ for $P = 50$ random unit traces in $\mathbb{R}^{64}$ and $\beta = 20$, with the bound of Proposition 16.1. Computed by the script `figs.py` with seed 8; an illustration of the stated analytic result and not an evaluation of the architecture.

finally by single-trace attractors. At any β, spurious mixtures of traces with high mutual overlap are stable, because a state equidistant from two near-duplicates cannot be resolved by a margin. Spurious recollections in the sense of the previous chapter are the fixed points that are not near any stored trace, and their presence as a function of β, P, and the overlap structure is the object of the capacity analysis in Chapter 34.

Figure 16.1 plots the observed one-step error against the margin for cues of increasing corruption. Every observation lies below the bound, which is loose by a modest factor in the exponentially small regime, and the error stays near its maximum until the margin exceeds $\beta^{-1} \ln(P-1)$.

Chapter 17

Episodic Memory and Temporal Reinstatement

Episodes are represented as trajectories and not as isolated vectors:

$$e_i = \{x_i(\tau) : \tau \in [t_i^-, t_i^+]\}, \tag{17.1}$$

where $x_i(\tau)$ is the organized state of the episode at internal time τ. The difference between recovering what happened and recovering how it was organized in time is the difference between the content of an episode and its temporal organization. For a finite episode, content is the set or multiset of states visited, and temporal organization comprises the order in which they were visited, the intervals between them, the nesting of subepisodes, and the boundaries. Proposition 1.1 showed that the two come apart in the encoding. They come apart in evaluation as well, which requires separate error measures. Let x and $\hat{x}$ be a true and a reconstructed episode with item sequences $\sigma, \hat{\sigma}$ over a common item set. A content error is a distance between the multisets, for example the symmetric difference size. An order error is the normalized Kendall tau distance $\tau_K(\sigma, \hat{\sigma}) = \binom{n}{2}^{-1} \#\{\text{discordant pairs}\}$. An interval error is the root-mean-square difference of inter-item lags. A boundary error compares the sets of boundary times, for example by the F-score of boundaries matched within a tolerance. A reconstruction with zero content error and large order error is possible and is the expected failure of an encoding that stores a bag of items.

Event boundaries are defined here by a surprise criterion that follows the event-segmentation tradition in cognitive psychology [27]: given a predictive model $p(x_{\tau+1} \mid x_{\leq\tau})$, a boundary is placed at τ when $-\log p(x_{\tau+1} \mid x_{\leq\tau})$ exceeds a threshold. Boundaries partition an episode into events and support two operations that matter for reconstruction: encoding at event level, in which each event is stored as a unit with within-event structure reconstructed from a generative model, and retrieval at event level, in which a cue first selects an event and then unfolds it. Temporal compression is the observation that recollection can proceed faster than the original episode. Let the replay map be $\tilde{x}(s) = x(\kappa s)$ with compression factor $\kappa > 1$; for sampled trajectories this is subsampling or temporal pooling, and it reduces the cost of reconstruction by the factor κ at the price of losing structure faster than the new sampling resolution, a loss that is bounded by the Nyquist condition on the retained structure.

Variable-resolution reconstruction is formalized by a nested family of subspaces $V_1 \subseteq V_2 \subseteq \cdots \subseteq V_L$ of the representation space and the corresponding orthogonal projections π_ℓ. A reconstruction at level ℓ is $\pi_\ell \hat{x}$. Because the subspaces are nested, the squared error $\|x - \pi_\ell x\|^2$ is nonincreasing in ℓ, and the cost of reconstruction increases with ℓ, so a coarse reconstruction at low cost can be refined as the task requires. The same family underlies the multiscale consolidation of Chapter 27.

Chapter 18

Context-Sensitive Reconstruction

Context is modeled as an operator acting on cues and candidate memories, so that the reconstruction under context $Z_t = z$ is

$$r_t = \mathcal{E}_z(c_t, M_t), \tag{18.1}$$

where the operator U_z acts on the cue before scoring and, in some formulations, on the stored traces as well. For a context group acting by unitary operators with $U_z U_{z'} = U_{z+z'}$, the effective score of a trace m against a cue c under context z is $S(m, U_z c)$. This single construction accounts for four properties that are otherwise treated separately. Context-dependent accessibility arises because the score of a given trace against a given cue depends on z through U_z. Cue ambiguity arises because the same c under different contexts may match different traces. Competing interpretations arise because several (m_i, z) pairs may give comparable scores. The effective retrieval landscape, the function $m \mapsto S(m, U_z c)$ over the store, changes with z even though the store is unchanged.

Encoding specificity takes an exact form in the phase-context case. Let content be a unit vector s and let the context be a global phase, $U_z s = e^{\mathrm{i}z} s$.

Proposition 18.1 (Encoding specificity as phase mismatch)**.** *If a trace is stored as $m = U_{z_0} s$ and the cue is the same content presented under context z, $c = U_z s$, then $S(m, c) = \frac{1}{2}\big(1 + \cos(z - z_0)\big)$. The score equals one for matched context and zero for contexts differing by π, independent of the content.*

Proof. Both vectors have unit norm, so by (6.5) $S = \frac{1}{2}(1 + \mathrm{Re}\langle e^{\mathrm{i}z_0} s, e^{\mathrm{i}z} s\rangle) = \frac{1}{2}(1 + \mathrm{Re}\, e^{\mathrm{i}(z - z_0)})$. □

The proposition formalizes the principle of Tulving and Thomson that the effectiveness of a retrieval cue depends on its overlap with the encoding context [11], with phase as the carrier of the context. Two corollaries follow from the structure of U_z. Because U_z is invertible, the loss of accessibility caused by a context mismatch is not a loss of information: a system that can estimate the encoding context can undo the mismatch by applying $U_{z_0} U_z^{-1}$ to the cue, and the recovery of a trace by context reinstatement is exactly the application of this inverse. Second, when the context is not known, it can be marginalized or estimated: the posterior over context given the cue and the store is $p(z \mid c) \propto p(z) \sum_i e^{\beta S(m_i, U_z c)}$, and the system may either report the retrieval under the maximum a posteriori context or retain the mixture. The second option is the one required for a faithful treatment of ambiguity, since a hard assignment of context discards the information that the cue was compatible with several.

Chapter 19

Recollection, Familiarity, and Source Discrimination

Recognition of familiar content and recovery of its historical context are different operations and must be distinguishable in a formal account. Let each trace carry a content part ψ_i and a context-and-provenance part κ_i (Chapter 38). Familiarity is a function of the content scores alone, for example $f(c) = \max_i S(\psi_i, c)$ or a smooth maximum $\beta^{-1} \log \sum_i e^{\beta S(\psi_i, c)}$. Recollection involves, in addition, the scores of the context parts, and the recollective output is not merely a number but a reconstruction with an attribution of source and context.

Proposition 19.1 (Content-only familiarity cannot discriminate source)**.** *Suppose two traces have identical content $\psi_A = \psi_B$ and different context-provenance parts $\kappa_A \neq \kappa_B$. Then any familiarity function that depends on the trace set only through the content scores takes equal values on the two traces, and no such function can assign different support to the hypotheses that the content was encountered in context A or in context B.*

Proof. A function of $\{S(\psi_i, c)\}_i$ treats traces A and B symmetrically because their content scores coincide for every cue c. □

Source attribution requires scoring the provenance parts, and the natural construction is a posterior over sources. With provenance labels ν_i and trace weights w_i from the retrieval stage, define the attribution distribution $p(\nu \mid c) = \sum_i w_i \mathbb{1}[\nu_i = \nu]$, and define source-attribution confidence as the probability mass on the most probable source, with a margin over the second. Two failure types can then be separated: content retrieval that is accurate with source misattribution, in which the content weights concentrate on traces whose content matches but whose provenance label differs from the true source because the encoding contexts were similar or because the cue carries stronger contextual overlap with the wrong source, and source attribution that is accurate for a content that is wrongly reconstructed. Provenance reconstruction, as distinct from provenance lookup, is the process of inferring the context of the original event from the stored context parts and the current state, and it is subject to all the interference effects of Chapter 20. The consequence for architecture is that the provenance part of a trace must be a first-class component of the stored record, and that source discrimination is a test on which static content-addressable stores fail by construction, which makes it a meaningful discriminating experiment in Chapter 49.

Chapter 20

Memory Interference and False Recollection

Interference is the degradation of recollection of one item by the presence of others. Four varieties are distinguished: proactive interference, in which earlier traces degrade the acquisition or retrieval of later ones; retroactive interference, in which later traces degrade earlier ones; trace competition, in which several traces compete for the same cue; and source confusion, in which the content is retrieved with the wrong attribution. In the softmax retrieval model, the effect of an additional trace on the weight of a target is exact.

Proposition 20.1 (Dilution by an added trace)**.** *Let a store give a target trace weight $w_t = e^{\beta s_t}/Z$ with $Z = \sum_j e^{\beta s_j}$, and add a trace with score s_n against the same cue. The new weight is $w_t' = w_t\, Z/(Z + e^{\beta s_n})$, and the fractional loss $1 - w_t'/w_t = e^{\beta s_n}/(Z + e^{\beta s_n})$ equals the weight that the new trace itself receives.*

Proof. The partition function becomes $Z' = Z + e^{\beta s_n}$ and the numerator is unchanged. □

The proposition shows that the interference caused by a trace is proportional to the probability mass it captures, which is largest when its score approaches or exceeds the target's, and negligible when $\beta(s_n - s_t)$ is very negative. In this model proactive and retroactive interference are arithmetically the same at retrieval: the order of storage does not enter. Distinguishing them requires an additional mechanism that depends on age or on write order, such as the amplitude decay discussed in Chapter 29 through the balance factor ρ, or a write rule that orthogonalizes a new trace against existing ones, so that the proactive and retroactive effects are caused by different parts of the architecture and are separately identifiable.

False recollection is a reconstruction that matches strongly but is historically wrong. A classical case is the production of a critical lure after studying a list of its associates [17], and a model of the effect follows from the geometry of the store.

Proposition 20.2 (Centroid lure)**.** *Let $m_1, \dots, m_P$ be unit vectors with common pairwise real inner product $\mu \in [0, 1)$, and let $g = \frac{1}{P}\sum_i m_i$. Then $\|g\|^2 = (1 + (P-1)\mu)/P$ and the normalized centroid $\hat{g} = g/\|g\|$ satisfies $\langle \hat{g}, m_j \rangle = \sqrt{(1 + (P-1)\mu)/P} =: \kappa_P < 1$ for every j. For a reconstruction $r = \hat{g}$, $S(r, \hat{g}) = 1 > S(r, m_j) = \frac{1}{2}(1 + \kappa_P)$, so the lure resonates more strongly than any presented item.*

Proof. $\|g\|^2 = P^{-2}\big(P + P(P-1)\mu\big)$ and $\langle g, m_j \rangle = P^{-1}(1 + (P-1)\mu)$, so $\langle \hat{g}, m_j \rangle = \langle g, m_j \rangle / \|g\| = \sqrt{(1 + (P-1)\mu)/P}$, which is below one because $\mu < 1$. The scores follow from (6.5). □

The proposition shows that a high resonance score cannot by itself certify historical support, since a composite assembled from many traces can be closer to a hypothetical item than to any actual one. The architecture therefore requires a quantity that is separate from resonance and that measures support by the ledger of Chapter 41.

Definition 20.3 (Resonance and historical support)**.** Let a reconstruction r be given with resonance $\mathrm{res}(r) = \max_i S(m_i, r)$ against the store and let the ledger H contain the inscription records. The reconstruction is δ-supported if there is an inscription record $h \in H$ with

$d(r, \mathrm{dec}(h)) \leq \delta$, where $\mathrm{dec}(h)$ is the content recorded at inscription, and is a constructed reconstruction otherwise. A recollection is reported as supported only if it is δ-supported or if it is declared as a composite with the identified supporting records and their weights.

With this separation the model can report a highly resonant but constructed reconstruction as such, and the evaluation in Chapter 49 measures the false-recollection rate as the fraction of reported-supported recollections that fail the ledger check against the ground truth of the experiment.

Part V

Predictive Inference and Memory

Chapter 21

Memory as Inference over Latent Causes

Recollection is formulated as inference under a generative model in which the past organization is a latent cause of the present cue. Let x denote the latent past organization, M the persistent memory, Z the current context, and c the cue. The recollection problem is to compute or approximate

$$p(x_{t_0} \mid c_t, M_t, Z_t) \;\propto\; p(c_t \mid x_{t_0}, Z_t)\, p(x_{t_0} \mid M_t). \tag{21.1}$$

The factor $p(x \mid M)$ is the prior over past organizations that the persistent memory induces, and the factor $p(c \mid x, Z)$ is the cue likelihood, which says how a cue is produced from the past organization under the present context. The decomposition places the contribution of memory and the contribution of evidence in separate terms and makes explicit that recollection draws on both.

The softmax retrieval rule that appears throughout the monograph is the exact Bayesian posterior mean under a simple instance of this model. Let the prior be a mixture of isotropic Gaussians centered on the stored traces, $p(x \mid M) = \sum_i \pi_i \mathcal{N}(x; m_i, \sigma_x^2 I)$, and let the cue be a noisy full observation, $c = x + \xi$ with $\xi \sim \mathcal{N}(0, \sigma_c^2 I)$, all in the realified space.

Proposition 21.1 (Softmax readout as posterior mean)**.** *Under this model the posterior is a mixture with weights $w_i \propto \pi_i \exp\!\big(-\|c - m_i\|^2/(2s^2)\big)$, $s^2 = \sigma_x^2 + \sigma_c^2$, and posterior mean*

$$\mathbb{E}[x \mid c] = (1 - g) \sum_i w_i m_i + g\, c, \qquad g = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_c^2}. \tag{21.2}$$

If the traces and the cue have equal norms and the prior is uniform, then $w_i = \mathrm{softmax}_i(\beta \langle c, m_i \rangle)$ with $\beta = 1/s^2$, the weight rule of (16.1)*.*

Proof. Marginalizing x within component i, the cue is distributed as $\mathcal{N}(m_i, s^2 I)$, which gives the mixture weights by Bayes' rule. Within component i the posterior of x is Gaussian with mean $(\sigma_c^2 m_i + \sigma_x^2 c)/s^2 = (1 - g)m_i + gc$. Averaging over components gives (21.2). For equal norms, $\|c - m_i\|^2 = \|c\|^2 + \|m_i\|^2 - 2\langle c, m_i \rangle$, and the terms that do not depend on i cancel in the normalization, leaving $w_i \propto \exp(\langle c, m_i \rangle / s^2)$. □

The resemblance between this expression and the recurrent relaxation step needs care, because two different operators are involved. The posterior mean (21.2) is a one-shot reconstruction operator: its weights are evaluated at the cue, with inverse temperature $\beta_c = 1/(\sigma_x^2 + \sigma_c^2)$, and the coefficient g weights the observation itself, so that a precise cue ($\sigma_c \to 0$) gives $g \to 1$ and the reconstruction follows the cue. The relaxation step of the minimal model, $r^{(k+1)} = (1 - \eta) r^{(k)} + \eta \sum_i w_i(r^{(k)}) m_i$, is an iteration whose weights are evaluated at the current state and whose coefficient η is a step size and not a reliability ratio. Both are convex combinations of a memory-weighted readout with another vector, and that is all the resemblance establishes. A formal relationship requires an objective whose stationary points the iteration computes, and such an objective exists. Take unit-norm traces, a uniform prior, and write $\beta_x = 1/\sigma_x^2$ and $\mathcal{E}_\beta$ for the energy of Chapter 16 at inverse

temperature β. The negative log of the unnormalized posterior over the latent organization x given the cue is, up to an additive constant,

$$J(x) = \beta_x \, \mathcal{E}_{\beta_x}(x) + \frac{\|x - c\|^2}{2\sigma_c^2}, \tag{21.3}$$

since $\log \sum_i e^{-\|x-m_i\|^2/2\sigma_x^2} = -\frac{1}{2}\beta_x \|x\|^2 + \log \sum_i e^{\beta_x \langle m_i, x\rangle}$ + const for unit-norm traces.

Proposition 21.2 (Cue-anchored recurrence computes posterior modes)**.** *Let $L = \beta_x + \sigma_c^{-2}$, let $g = \sigma_x^2/(\sigma_x^2 + \sigma_c^2)$, and define $T(x) = (1-g)F_{\beta_x}(x) + g\,c$, where F_{β_x} is the softmax update* (16.1) *at inverse temperature β_x. Then $\nabla J(x) = L\,(x - T(x))$, so the stationary points of J are exactly the fixed points of T. For every $\eta \in (0, 2)$ the damped iteration $x^+ = (1-\eta)x + \eta\, T(x)$ satisfies*

$$J(x^+) \le J(x) - L\,\eta\Big(1 - \frac{\eta}{2}\Big)\|x - T(x)\|^2, \tag{21.4}$$

so J decreases along the iteration, and every limit point of the iteration is a fixed point of T. When $\sigma_c \to \infty$ the iteration reduces to the cue-free relaxation $x^+ = (1-\eta)x + \eta F_{\beta_x}(x)$, which descends the energy $\mathcal{E}_{\beta_x}$.

Proof. The gradient is $\nabla J(x) = \beta_x(x - F_{\beta_x}(x)) + (x - c)/\sigma_c^2$. Since $L(1-g) = \beta_x$ and $Lg = \sigma_c^{-2}$, this equals $L(x - T(x))$. For the decrease, write $J(x) = \frac{1}{2}L\|x\|^2 - \langle x, c\rangle/\sigma_c^2 - \ell(x) +$ const with $\ell(x) = \log \sum_i e^{\beta_x \langle m_i, x\rangle}$, which is convex with $\nabla \ell(x) = \beta_x F_{\beta_x}(x)$. Linearizing ℓ at x gives the majorizer $\tilde{Q}(y) = \frac{1}{2}L\|y\|^2 - \langle y, c\rangle/\sigma_c^2 - \ell(x) - \langle \beta_x F_{\beta_x}(x), y - x\rangle \ge J(y)$ with equality at $y = x$. The function $\tilde{Q}$ is a quadratic of curvature L minimized at $y = T(x)$, with $\nabla \tilde{Q}(x) = -L(T(x) - x)$. Along the segment $x + \eta(T(x) - x)$, $\tilde{Q}$ equals $\tilde{Q}(x) - L(\eta - \eta^2/2)\|T(x) - x\|^2$, and $J \le \tilde{Q}$ with $J(x) = \tilde{Q}(x)$ gives the stated inequality. The function J is bounded below and coercive, because $\log \sum_i e^{\beta_x \langle m_i, x\rangle} \le \log P + \beta_x \|x\|$. The summed decreases are therefore finite, $\|x - T(x)\| \to 0$, and continuity of T shows that limit points are fixed points. The final statement is the case $L = \beta_x$, $g = 0$. □

The proposition is the precise relationship between the two operators. The recurrent iteration is a monotone algorithm for the posterior mode, an instance of the concave–convex procedure, and in the cue-free case it is the mean-shift or expectation-maximization mode-finding iteration for the Gaussian mixture density of variance σ_x^2 [32]. The posterior mean (21.2) is a different object, a minimum-mean-square-error estimate evaluated at the cue with the flatter weights $\beta_c < \beta_x$. The two agree when the memory holds a single trace, since then $F \equiv m_1$ and both reduce to $(1-g)m_1 + gc$, and in general they differ: the mode of a multimodal posterior is a stored trace blended with the cue, while the posterior mean is an average over modes. The monograph therefore treats them as two reconstruction operators that answer different questions, and the sharpness parameter that governs selection in the recurrent iteration is $\beta_x = \sigma_x^{-2}$, the precision of the prior around each trace, and not the total precision of the cue and trace.

Three forms of inference over time should be distinguished. Filtering estimates the present state from past observations and is the mode used in forward prediction. Smoothing estimates a past state from observations that include later ones, and recollection of an earlier episode is naturally a smoothing problem, because the cue at recall time, and what has been learned since, are evidence about the earlier state. Prediction extrapolates forward from an estimated state. Reconstruction and prediction are therefore posteriors over latent states

at different times under a common generative model, which accounts for the empirical coupling between remembering the past and imagining the future without identifying the two. Amortized inference replaces the iterative computation of the posterior by a trained map $q_\phi(x \mid c, M, Z)$; the penalty is the amortization gap, the expected divergence $\mathbb{E}_c \, \mathrm{KL}(q_\phi(\cdot \mid c) \,\|\, p(\cdot \mid c))$ between the learned approximation and the true posterior, and Chapter 25 treats when it is worth paying.

Chapter 22

Variational Ecphory

Let q be an approximate posterior over the latent organization. The variational free energy of the recollection problem is

$$\mathcal{F}[q] = \mathbb{E}_q\big[-\log p(c, x \mid M, Z)\big] - H(q). \tag{22.1}$$

Proposition 22.1 (Decomposition of the free energy)**.** *For every distribution q with support contained in that of the posterior, $\mathcal{F}[q] = \mathrm{KL}\big(q \,\|\, p(\cdot \mid c, M, Z)\big) - \log p(c \mid M, Z)$. Hence $\mathcal{F}[q] \geq -\log p(c \mid M, Z)$ with equality if and only if q is the exact posterior.*

Proof. Write $p(c, x \mid M, Z) = p(x \mid c, M, Z)\, p(c \mid M, Z)$. Then $\mathbb{E}_q[-\log p(c, x)] = \mathbb{E}_q[-\log p(x \mid c)] - \log p(c)$, and $-H(q) = \mathbb{E}_q[\log q]$, so the sum is $\mathbb{E}_q[\log q - \log p(x \mid c)] - \log p(c) = \mathrm{KL}(q \| p(\cdot \mid c)) - \log p(c)$. Nonnegativity of the divergence gives the inequality and its equality case. □

Two sources of uncertainty enter any recollection and the framework requires that they be kept apart. The first is uncertainty about the historical episode that remains even when the posterior is computed exactly, measured by the posterior entropy $H(p(\cdot \mid c, M, Z))$ and, for point estimates, by the posterior covariance. It is a property of what the memory and the cue together determine. The second is uncertainty caused by present retrieval dynamics: the system does not hold the exact posterior but a truncated or biased approximation q_k after k iterations, and its excess relative to the exact posterior is the inference gap $\mathcal{F}[q_k] - \mathcal{F}[p] = \mathrm{KL}(q_k \| p) \geq 0$. For a point estimate $\hat{x}$ that is a function of the cue, the corresponding decomposition of squared error is exact.

Proposition 22.2 (Irreducible and reducible error)**.** *For any estimator $\hat{x}(c)$ and posterior mean $\mu(c) = \mathbb{E}[x \mid c]$, $\mathbb{E}\big[\|x - \hat{x}(c)\|^2 \mid c\big] = \operatorname{tr}\mathrm{Cov}(x \mid c) + \|\mu(c) - \hat{x}(c)\|^2$. The first term does not depend on the estimator and is the irreducible uncertainty about the episode given cue and memory; the second is the reducible part attributable to the retrieval procedure.*

Proof. Add and subtract $\mu(c)$ inside the norm; the cross term vanishes because $\mathbb{E}[x - \mu(c) \mid c] = 0$. □

The monograph reports reconstruction error in this decomposed form wherever the posterior is available, so that an improvement in retrieval dynamics cannot be mistaken for a reduction in intrinsic uncertainty and a system that is confidently wrong can be distinguished from one that is appropriately uncertain.

Chapter 23

Precision Weighting and Resonance Selection

Precision-sensitive cue integration weights each source of evidence by its reliability, and the architecture contains three places where an oscillatory coherence signal could plausibly act. It could change the gain on incoming evidence, so that coherent states give more weight to the cue. It could change the selection of memory traces, so that coherence sharpens or flattens the competition among them. It could change the transition dynamics of the reconstruction process, so that coherence speeds or slows convergence. These are distinct hypotheses with different observable signatures, and the first task is to determine which of them are distinguishable at all from outcomes.

Consider unit-norm traces and several cues $c_1, \dots, c_J$ with precisions $\Pi_1, \dots, \Pi_J$, each cue observing the target with independent Gaussian noise of variance $1/\Pi_j$. By the calculation of Proposition 21.1 extended to several cues, the trace weights are

$$w_i \;\propto\; \pi_i \exp\Big(\sum_{j=1}^{J} \Pi_j \langle c_j, m_i \rangle\Big). \tag{23.1}$$

A uniform gain γ on all evidence multiplies every Π_j by γ, and a change of the selection temperature, $\beta \mapsto \gamma\beta$, multiplies the whole exponent by γ as well.

Proposition 23.1 (Identifiability of the three loci)**.** *(i) A uniform gain on evidence and a change of the selection inverse temperature produce identical trace weights, and cannot be distinguished by the weights, the reconstruction at convergence, or the accuracy. (ii) They can be distinguished by cue-specific precision, because gain changes act on $\Pi_j/\Pi_{j'}$ only if the gain depends on the cue, whereas a temperature change leaves relative cue weighting fixed. (iii) A change of the relaxation coefficient η in $r^{(k+1)} = (1-\eta)r^{(k)} + \eta \sum_i w_i m_i$ leaves the fixed point unchanged and changes only the rate, with error $(1-\eta)^k \|r^{(0)} - r^\star\|$ at step k.*

Proof. (i) is immediate from (23.1): the exponent is scaled by γ in both cases. For (ii), if the gain applies uniformly then $\Pi_j/\Pi_{j'}$ is unchanged, and the weight vector is a function only of the common scale and the ratios, so a uniform gain cannot be told from a temperature change; a cue-specific gain changes the ratios, which no temperature change can do. For (iii), the fixed point of $r \mapsto (1-\eta)r + \eta v$ with $v = \sum_i w_i m_i$ is v for every $\eta \in (0, 1]$, and $r^{(k)} - v = (1-\eta)^k (r^{(0)} - v)$. □

The implication for empirical discrimination is a design requirement. To separate the gain and selection hypotheses the experiment must manipulate the relative reliability of two or more cues while manipulating coherence; to separate transition effects from the other two, it must record the time course of convergence in addition to the endpoint. A coherence manipulation that changes only endpoint accuracy cannot discriminate among the three, and the program in Chapter 51 is specified to include these manipulations.

Chapter 24

Active Inference and Prospective Memory

Reconstruction extends to action selection and future simulation by embedding the recollected state in a model of the future. The standard formalism, active inference, evaluates candidate policies π by their expected free energy,

$$G(\pi) = \mathbb{E}_{q(o,s|\pi)}\big[\log q(s \mid \pi) - \log p(o, s \mid \pi)\big] \tag{24.1}$$

which, under a prior preference distribution $p(o)$ over outcomes, decomposes into a risk term, the divergence of predicted outcomes from preferred outcomes, and an ambiguity term, the expected uncertainty of outcomes given states [29, 46]. Remembered states enter through the predictive prior: a recollection $q(x_{t_0} \mid c, M, Z)$ informs beliefs about the present and future state through the transition model $p(s_{t+1} \mid s_t, a)$ for states that depend on the remembered organization. Remembering is not thereby identified with planning, because the two computations condition on different things. Recollection conditions on a cue about the past and on persistent memory, and its posterior (21.1) contains no term for preferences or policies. Planning conditions on candidate policies and on preferences, and its posterior over policies is proportional to $\exp(-\gamma G(\pi))$. They are connected through the shared generative model and distinct in the quantities that define their objectives.

It follows that the value of a future continuation is not equivalent to the historical accuracy of a recollection, and a system optimizing one can sacrifice the other. A reconstruction biased toward outcomes that are favorable under the current preferences can have higher expected value for action selection and lower fidelity to the history, and an accurate reconstruction of a history irrelevant to the present can be of no value. Mem8 therefore keeps two separately defined objectives. The recollection objective is the variational free energy of Chapter 22, with historical fidelity measured against the ledger, and the decision objective is the expected free energy (24.1) or a utility from Chapter 37. A specification that merged them into a single scalar would make it impossible to detect the case in which fidelity has been traded for value, and the architecture rules out silent trade by treating fidelity as a constraint, not as one term of a sum, in Chapter 37.

Chapter 25

Amortized Deliberation and Learned Retrieval Policies

Reconstruction costs computation, and the cost is not always justified. The system faces a recurring choice among three actions when a cue arrives: to reuse a cached reconstruction produced earlier for a similar cue and context, to initiate a full ecphoric process, or to acquire further evidence before reconstructing. The distinction between costly reconstruction and cached retrieval is the memory-specific instance of a general distinction between deliberation whose cost is paid at decision time and responses amortized from earlier deliberation, with the cost spread over many uses; the same decision-theoretic structure applies whether the cached object is an intuitive judgment or a stored reconstruction, and what varies is only the price and the validity conditions.

The contraction analysis of Chapter 15 supplies a principled validity condition for the cache and a principled stopping rule for the iteration.

Proposition 25.1 (Cache validity radius and iteration budget)**.** *Let the reconstruction map be a λ-contraction with cue-Lipschitz constant L_c as in Theorem 15.1. A reconstruction cached for cue c' is within ε of the fixed point for cue c whenever $d(c, c') \le \varepsilon(1-\lambda)/L_c$. Starting from any $r^{(0)}$, the iteration is within ε of its fixed point after*

$$k \;\ge\; \frac{\ln\!\big(\varepsilon(1-\lambda)/d(r^{(1)}, r^{(0)})\big)}{\ln \lambda} \tag{25.1}$$

steps. In the relaxation map of the minimal model with fixed weights, $\lambda = 1-\eta$ and the error after k steps equals $(1-\eta)^k$ times the initial error.

Proof. The first statement is part (iii) of Theorem 15.1 rearranged. The second solves $\lambda^k d(r^{(1)}, r^{(0)})/(1-\lambda) \le \varepsilon$ for k, noting that $\ln\lambda < 0$ reverses the inequality. The last statement is part (iii) of Proposition 23.1. □

A policy for the choice among reuse, reconstruction, and evidence acquisition is a map from the observable state, which includes the cue, the context, the age and provenance of the cached item, and the current uncertainty, to an action. The optimal policy compares net values: reuse has cost near zero and a value reduced by the expected error at cache distance; reconstruction has cost proportional to the iteration count of Proposition 25.1 and a value corresponding to the residual error; evidence acquisition has the cost of the observation and a value equal to the expected reduction of posterior uncertainty, which the variational account of Chapter 22 makes computable as an expected information gain. Under a myopic value-of-computation criterion, the iteration continues while the expected decrease in loss from one more step exceeds the marginal cost of the step, and stops otherwise. A learned retrieval policy amortizes this decision as a trained classifier over the observable state, which is itself an instance of the amortized deliberation that it governs: the policy's own cost is paid once at training and spread over all later decisions, and its validity depends on the stationarity of the distribution of cues, so its failure under distribution shift is a failure mode to be tested in Chapter 53.

Part VI

Persistence, Consolidation, and Transformation

Chapter 26

Encoding as the Creation of Recoverable Constraints

Encoding is treated here as an operation that changes the set of future states that can be recovered from partial cues, not as the deposit of a copy. A history h leaves a memory state $E(h)$, and what matters about that state is the set of reconstructions it makes available under a family of cues and contexts. Recoverability must therefore be defined relative to such a family, since a trace that no admissible cue can elicit is, for practical purposes, absent, and a trace that many cues can elicit is robust.

Definition 26.1 (Recoverability)**.** Fix a cue family Q, a context family $\mathcal{Z}_0$, a target recollection function $T : \mathcal{H}ist \times \mathcal{C} \to \mathcal{Y}$, a tolerance $\delta \geq 0$, and a metric d on $\mathcal{Y}$. A history h is $(Q, \mathcal{Z}_0, \delta)$-recoverable from the memory state m if

$$\sup_{c \in Q,\, z \in \mathcal{Z}_0} d\big(\mathcal{E}(m, c, z), T(h, c)\big) \leq \delta. \tag{26.1}$$

It is cue-wise recoverable at (c, z) if the inequality holds at that single pair.

The definition makes explicit that storage of a trace does not guarantee recollection. In the softmax memory of Chapter 16, a trace m_1 among P stored traces is recoverable to within 2ε at a cue only if the cue gives it a margin Δ over every competitor with $\Delta \geq \beta^{-1} \log\big((P-1)/\varepsilon\big)$ (Proposition 16.1), and for cues orthogonal to all traces the weights are uniform and the reconstruction is the centroid of the store, from which no individual trace can be read. Adding traces raises the required margin logarithmically in P, which is a precise sense in which each new encoding changes the recoverability of the old ones without altering them. Conversely, a well-chosen encoding can create recoverability that did not exist: writing a trace that links a previously inaccessible content to a cue that is commonly available enlarges the recoverable set for that content without touching the content itself. The monograph accordingly describes the encoding stage by its effect on the recoverable set, and it requires of any encoding rule a statement of the cue and context families for which it intends to create recoverability.

Chapter 27

Consolidation and Multiscale Memory

A single store with a single timescale cannot be both plastic enough to absorb new episodes quickly and stable enough to retain old structure without interference, and the standard resolution is a division of labor among systems with different rates, in which fast learning of specific episodes is combined with slow integration of shared structure through repeated, interleaved replay [26]. Mem8 adopts a layered form of this division. Transient activation, the fields and phases of Part III, lives on timescales of milliseconds to seconds and is not itself persistent. A fast store holds recent traces with high fidelity and orthogonalized, pattern-separated codes, together with the full provenance of each. A slow store holds overlapping distributed codes that capture regularities across many episodes and supports semantic completion, and it is updated by replay of traces from the fast store at a small learning rate. Episode-specific provenance remains in the ledger of Chapter 41 and is not folded into the slow store, so that consolidation changes where a content is accessible from without changing the record of where it came from.

The dynamics of transfer admit a closed form in the linear case. Let $m_f(t)$ be the strength of a trace in the fast store and $m_s(t)$ its strength in the slow store, with the fast store decaying at rate γ, replay transferring strength to the slow store at rate κ, and the slow store decaying at rate δ:

$$\dot{m}_f = -(\gamma + \kappa) m_f, \qquad \dot{m}_s = \kappa m_f - \delta m_s. \tag{27.1}$$

Proposition 27.1 (Two-store solution)**.** *With $a = \gamma + \kappa \neq \delta$ and initial condition $m_f(0) = m_0$, $m_s(0) = 0$, the solution is*

$$m_f(t) = m_0 e^{-at}, \qquad m_s(t) = \frac{\kappa\, m_0}{a - \delta}\big(e^{-\delta t} - e^{-at}\big). \tag{27.2}$$

If $a > \delta$, the slow-store strength increases to a maximum at $t^\star = \ln(a/\delta)/(a - \delta)$ and then decays at the slow rate δ.

Proof. The first equation integrates directly. For the second, multiply by $e^{\delta t}$ to obtain $\frac{d}{dt}(e^{\delta t} m_s) = \kappa m_0 e^{-(a-\delta)t}$, integrate from zero, and rearrange. Differentiating,

$$\dot{m}_s = \frac{\kappa m_0}{a - \delta}\big(a e^{-at} - \delta e^{-\delta t}\big),$$

and setting this to zero gives $a e^{-at} = \delta e^{-\delta t}$, hence $e^{(a-\delta)t} = a/\delta$. □

The shape is the one that systems consolidation requires: the fast-store strength falls quickly, the slow-store strength rises and persists, and the time of its maximum is controlled by the ratio of the rates. Total accessibility is the sum of the two contributions weighted by their access costs, which is why a trace that has been consolidated can be cheaper to reconstruct even when its total strength is lower. Replay in this architecture is not a literal recapitulation: it is the sampling of traces from the fast store with probability that depends on their recency, their prediction error, and their importance, followed by a small gradient

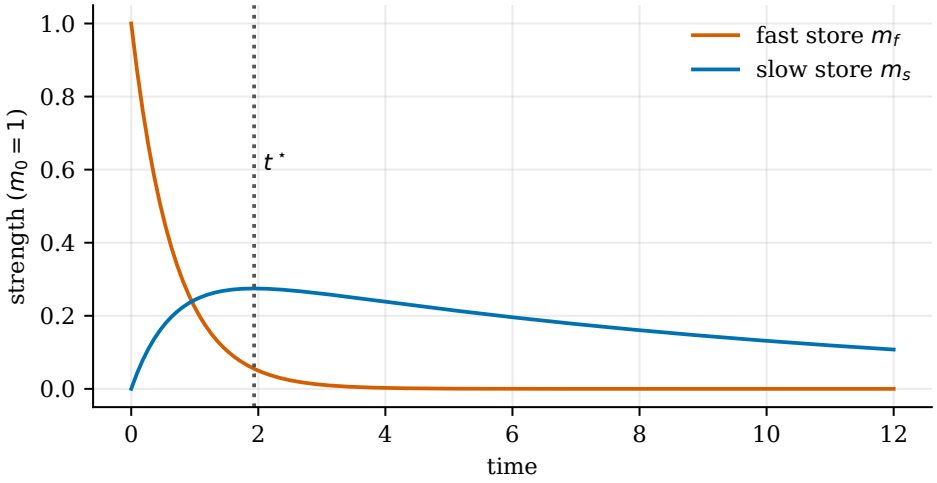


Figure 27.1: Closed-form solution of Proposition 27.1 for $\gamma = 1$, $\kappa = 0.5$, $\delta = 0.1$. The slow store peaks at $t^\star$ and then decays at the slow rate. Computed by the script `figs.py` with seed 8; an illustration of the stated analytic result and not an evaluation of the architecture.

or Hebbian update of the slow store, and the interleaving of old and new samples is what prevents the new material from overwriting the old, which is the catastrophic interference that sequential training would produce. Stability under continual learning is stated as an open problem in Chapter 59.

Figure 27.1 plots the solution of the two-store system. The transfer from the fast to the slow store produces the characteristic crossing, after which the slow store carries the trace.

Worked consolidation: the two-store peak and a three-store cascade

The closed form of Proposition 27.1 supports several quantities that can be computed exactly, and a cascade of more than two stores can be handled by the same method. This section gives the numbers for the parameters of Figure 27.1, treats the degenerate case that the proposition excludes, and states the cascade as a proposed extension whose proof is a partial-fraction computation.

Worked Example 27.1 (Peak time and transfer efficiency)**.** Take $\gamma = 1$, $\kappa = \frac{1}{2}$ and $\delta = \frac{1}{10}$, so that $a = \frac{3}{2}$. The slow store peaks at $t^\star = \ln(15)/1.4 = 1.934$, at which time the fast store retains $m_f(t^\star) = 0.0549\, m_0$ and the slow store holds its maximum $m_s(t^\star) = 0.2747\, m_0$. The maximum has the closed form

$$m_s(t^\star) = \frac{\kappa m_0}{a}\left(\frac{\delta}{a}\right)^{\delta/(a-\delta)}, \tag{27.3}$$

which follows by substituting $e^{-at^\star} = (\delta/a)\, e^{-\delta t^\star}$ and $e^{-\delta t^\star} = (\delta/a)^{\delta/(a-\delta)}$ into the solution, and it agrees with the numerical maximum to four decimals. The factor $\kappa/a = 0.333$ is the fraction of the initial strength that is ever transferred to the slow store, since the fast store loses strength at the total rate a and the share κ/a of that flow goes to replay; the peak is below this ceiling because the slow store is already decaying during the transfer. The consequence for design is that replay efficiency, $\kappa/(\gamma + \kappa)$, caps the slow-store strength before

the slow decay rate δ is considered, so raising κ and lowering γ are not interchangeable with lowering δ.

The case $a = \delta$ is excluded from Proposition 27.1 because the denominator vanishes. Taking the limit of the solution gives $m_s(t) = \kappa m_0\, t e^{-at}$, with maximum $\kappa m_0/(ae)$ at $t = 1/a$. For $\gamma = \kappa = \frac{1}{2}$ and $\delta = 1$ this is $0.1839\, m_0$ at $t = 1.0$, which the matrix exponential reproduces.

Extension 27.1 (Cascade of n stores)**.** Let store i lose strength at rate a_i and pass strength to store $i+1$ at rate $k_i \le a_i$, with $m_1(0) = m_0$ and $m_i(0) = 0$ for $i > 1$, and let the rates $a_1, \dots, a_n$ be distinct. Then

$$m_n(t) = m_0 \Big(\prod_{i=1}^{n-1} k_i \Big) \sum_{i=1}^{n} \frac{e^{-a_i t}}{\prod_{j \neq i} (a_j - a_i)}, \tag{27.4}$$

and the total strength ever delivered to store n is $m_0 \prod_{i=1}^{n-1} k_i / a_i$.

Proof. The Laplace transform of the cascade gives $\hat{m}_n(s) = m_0 \prod_{i<n} k_i / \prod_{i \le n} (s + a_i)$. Since the roots $-a_i$ are simple, partial fractions give residues $1/\prod_{j \neq i}(a_j - a_i)$ at $s = -a_i$, and inversion gives the sum. For the delivered strength, store i passes on the fraction k_i/a_i of what it loses, and the product of these fractions along the chain is the share that reaches store n. □

The formula is the Bateman solution for a decay chain, a classical result, and it is used here for the memory interpretation and not as a new theorem.

Worked Example 27.2 (Three-store cascade)**.** Take exit rates $(a_1, a_2, a_3) = (1.5, 0.4, 0.1)$ and transfer rates $(k_1, k_2) = (0.5, 0.3)$. The second store peaks at $t = \ln(a_1/a_2)/(a_1 - a_2) = 1.2$, and the third at $t = 5.4$ with strength $0.1557\, m_0$, so the deeper store peaks later and lower, as the sequence of timescales requires. The strength ever delivered to the third store is $k_1 k_2/(a_1 a_2) = 0.250\, m_0$, and the numerical integral of $a_3 m_3$ over the whole time axis reproduces it, which is the conservation statement in the extension. The cascade quantifies the cost of depth: each added store multiplies the available strength by the transfer fraction k_i/a_i and delays the peak, so a deep hierarchy is justified only when the recollections it serves are those that the shallow stores lose first.

Chapter 28

Reconsolidation and Memory Modification

The retrieval of a trace can render it labile and subject to modification before it is stabilized again, a process observed in experimental work on reconsolidation after retrieval [30]. In Mem8 the corresponding operation is a controlled post-retrieval update,

$$M_{t+1} = M_t + \eta\, \mathcal{U}(M_t, r_t, c_t, Z_t), \tag{28.1}$$

where $\mathcal{U}$ is an update direction computed from the retrieved reconstruction, the cue, and the context. Two requirements separate admissible updating from uncontrolled rewriting. The first is that the update be supported by evidence: $\mathcal{U}$ must depend on information that is external to the memory's own output, such as a new observation that conflicts with the retrieved reconstruction, and its magnitude must be gated by a statistic that measures the strength of that evidence. The second is that the record of what was previously stored be preserved, so that every revision can be audited and reversed.

The danger of an update that does not depend on external evidence can be stated exactly. Suppose that retrieval pulls each stored trace toward the reconstruction it helped produce, $\mathcal{U}_i = w_i(r - m_i)$, which is the natural form of an update driven by the memory's own output.

Proposition 28.1 (Self-reinforcing distortion)**.** *If the update $m_i^+ = m_i + \eta w_i(r - m_i)$ is applied with the reconstruction r and the weights w_i held fixed, then $m_i^+ - r = (1 - \eta w_i)(m_i - r)$. After n repetitions the deviation of trace i from r is $(1 - \eta w_i)^n (m_i - r)$, which tends to zero for every trace with $w_i > 0$.*

Proof. Subtract r from both sides of the update. Iterating gives the stated power. □

Repeated retrieval under such a rule therefore collapses the traces that participate in a reconstruction onto that reconstruction, erasing the distinctions among them: what is remembered becomes what was last reconstructed, which is a formalization of the drift toward schema-consistent reproduction observed in serial reproduction studies [13]. The fixed-weights assumption makes the statement exact, and the qualitative conclusion, that distinctness is lost through self-generated updates, does not depend on it. Evidence-supported updating replaces r in the update by an evidence-bearing target e and gates the update by a likelihood-ratio statistic $\Lambda(e, r)$, the ratio of the probability of the evidence under the hypothesis that the retrieved reconstruction is outdated to its probability under the hypothesis that it is correct: $\mathcal{U}_i = g(\Lambda)\, w_i(e - m_i)$ with $g(\Lambda) = 0$ below a threshold and increasing above it. A policy of this form changes a trace only when the world has told the system that it should.

The second requirement is satisfied by logging. If every update is recorded with the exact vector $\eta\, \mathcal{U}_s$ that was applied, then any earlier state is recoverable from a later one.

Proposition 28.2 (Reversibility of logged updates)**.** *If updates follow* (28.1) *and the ledger stores each increment $\Delta_s = \eta\, \mathcal{U}(M_s, r_s, c_s, Z_s)$ exactly, then $M_t = M_T - \sum_{s=t}^{T-1} \Delta_s$ for every $t \le T$.*

Proof. Telescoping the recursion $M_{s+1} = M_s + \Delta_s$ from $s = t$ to $T - 1$. □

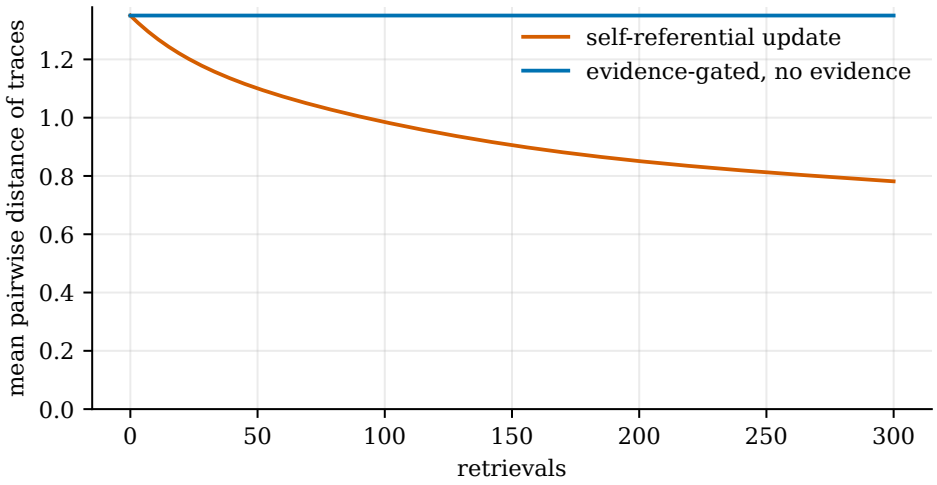


Figure 28.1: Mean pairwise distance among five stored traces under repeated retrieval with a self-referential update ($\eta = 0.05$, $\beta = 6$, fixed cue), compared with an evidence-gated update that receives no evidence. Computed by the script `figs.py` with seed 8; an illustration of the stated analytic result and not an evaluation of the architecture.

The proposition is elementary and its consequence is the design rule of Chapter 41: memory modification must not silently rewrite the record of what was previously stored, and a log of increments, together with the evidence that triggered each, makes the history of a memory a first-class and recoverable object.

Figure 28.1 simulates the update of Proposition 28.1 with weights and reconstruction recomputed at every step. Traces that participate in the reconstruction collapse toward one another, while the gated update, which changes nothing without external evidence, leaves their separation intact.

Drift and gating: numbers and a counterexample

Proposition 28.1 states the collapse qualitatively, and the evidence gate of the preceding discussion is a design rule whose content depends on how its statistic is computed. Two short examples fix the scale of the first and show a way in which the second can fail.

Worked Example 28.1 (Scale of self-reinforcing drift)**.** Hold the weight on a participating trace at $w_i = 0.4$ and take the update rate $\eta = 0.05$, so the contraction factor per retrieval is $1 - \eta w_i = 0.98$. After $n = 50$ retrievals the trace retains the fraction $0.98^{50} = 0.364$ of its original deviation from the reconstruction, and the deviation halves after 34.3 retrievals. The collapse is therefore not a fast phenomenon at these rates: a trace that is retrieved fifty times has lost nearly two thirds of its distinctness from its neighbours, and the policy matters in the regime of heavily rehearsed material, which is the regime in which memory is least often checked against the world.

Counterexample 28.1 (Correlated reports pass the gate)**.** Suppose the gate opens when the likelihood ratio statistic Λ reaches $\Lambda^* = 10$, and that Λ is computed as the product of the likelihood ratios of the individual reports, which is correct when the reports are conditionally independent given the hypothesis. A single report from a source that is

correct with probability 0.7 has likelihood ratio $\ell = 0.7/0.3 = 2.333$. Three conditionally independent reports give $\Lambda = \ell^3 = 12.704 > \Lambda^*$ and the gate opens. If the three reports are three copies of one observation, as when one source is relayed three times, the same product still gives 12.704, whereas the correct likelihood ratio of the single underlying observation is $2.333 < \Lambda^*$ and the gate should stay closed. The statistic is inflated by the factor ℓ^{k-1} for k copies, which is $\ell^2 = 5.44$ here.

The counterexample is about the computation of the gate and not about the gate. The minimum number of conditionally independent reports of ratio ℓ needed to open a gate with threshold Λ^* is $\lceil \ln \Lambda^* / \ln \ell \rceil$, which equals three for the values above because $\ln 10 / \ln(7/3) = 2.718$, and the count is meaningful only if the reports are independent. The ledger of Chapter 41 supplies the required discipline, since every evidence record carries its source, so the gate can aggregate by source and count each at most once. This is a design constraint on the architecture and it is not a theorem about the world: sources can be correlated without being identical, and the treatment of partial dependence is left to the protocols of Part IX.

Chapter 29

Forgetting, Interference, and Resource Constraints

Forgetting is defined in three ways, and the three are not equivalent. It may be a decline of recoverability, so that Definition 26.1 fails for cue families that formerly satisfied it. It may be an increase of reconstruction cost, so that the trace remains recoverable only through a longer iteration or a more specific cue. It may be a deliberate release of a representational commitment, in which the system removes content, leaving in the ledger a record that the removal occurred. A theory that recognized only the first would treat all forgetting as loss, whereas the second and third are often the efficient and the required behavior.

The retrieval score gives amplitude decay a specific and sometimes unexpected consequence. If a stored trace of unit direction decays in amplitude as $m(t) = e^{-\gamma t}\hat{m}$ and the query is the aligned unit vector, then $\rho = 2e^{-\gamma t}/(1 + e^{-2\gamma t}) = \operatorname{sech}(\gamma t)$ and by Theorem 6.1

$$S\big(m(t), \hat{m}\big) = \frac{\operatorname{sech}(\gamma t)}{2}\big(1 + \operatorname{sech}(\gamma t)\big), \tag{29.1}$$

which falls monotonically from one at $t = 0$ toward zero, whereas the cosine of the same pair does not change with amplitude at all. Under the resonance score an older, weaker trace is therefore penalized relative to a newer one of equal alignment, producing a recency preference as a direct consequence of the energy-balance factor and without any dedicated mechanism. Whether this preference is desirable is a matter of policy: it is appropriate when recency predicts relevance and inappropriate when the old trace is the better evidence. The corollary for architecture is that amplitude must be treated as a managed quantity, either normalized at write time and given an explicit age feature, or allowed to carry age deliberately, and that the choice must be recorded as part of the specification.

Resource-sensitive forgetting is the policy of choosing which traces to release when a budget is binding. The comparison with indiscriminate temporal decay is clean in the simplest case.

Proposition 29.1 (Optimal retention under a budget)**.** *Suppose each of N items will be queried in the future independently with probabilities $p_1, \dots, p_N$, a query for a retained item succeeds and a query for a released item fails, and at most B items can be retained. The retention set that maximizes the expected number of successful queries is the set of B items with the largest p_i. A policy that retains the B youngest items attains the optimum if and only if the access probability is nonincreasing in age, in the sense that the youngest B items are among the B largest p_i.*

Proof. The expected number of successes is $\sum_{i \in R} p_i$ over the retained set R, which is maximized under the cardinality constraint by selecting the B largest values. The age-based policy selects the B youngest, which coincides with this set exactly when those items carry the B largest probabilities. □

The proposition does not say that decay is wrong; it says that decay is optimal precisely when age is a sufficient statistic for future access, and that where other predictors are available, such as the number of distinct contexts in which an item was previously retrieved, the provenance of the item, or its dependence by other retained items, a better policy

exists. The monograph compares learned and rule-based retention policies against age-based decay in Chapter 52 at matched budgets, and deliberate release is implemented by tombstone records that remove the content while preserving in the ledger the facts that the trace existed and that it was released, with the content removed in full when erasure is required.

Retention and recency: a counterexample and a closed form

Counterexample 29.1 (Youngest-first retention is not optimal)**.** Let six items, listed from youngest to oldest, have future access probabilities $(0.10, 0.50, 0.05, 0.30, 0.02, 0.40)$ and let the budget be $B = 3$. The age-based policy retains the three youngest and obtains an expected 0.65 successful queries, while the optimal policy of Proposition 29.1 retains the items with probabilities 0.50, 0.40 and 0.30 and obtains 1.20. The regret is 0.55 expected queries, nearly half of the optimum. The example is built so that age is a poor predictor, and the proposition says exactly when this happens: the age policy is optimal if and only if the youngest B items are the B most probable. Whether any real memory has access probabilities of this form is a matter of measurement, and the example makes no claim about it.

Worked Example 29.1 (Recency from the energy-balance factor)**.** For a trace decaying as $m(t) = e^{-\gamma t}\hat{m}$ and an aligned unit query, equation (29.1) gives scores that depend only on γt. At $\gamma t = 1$ one has $\operatorname{sech}(1) = 0.648$ and $S = 0.534$, at $\gamma t = 2$ one has 0.266 and 0.168, and at $\gamma t = 3$ one has 0.099 and 0.055, while the cosine of the same pair stays at one throughout. The score falls to one half when $\rho(1 + \rho) = 1$, that is, when ρ is the reciprocal of the golden ratio, and the corresponding age is $\gamma t = \operatorname{arcosh} \varphi = 1.0613$. The resonance score therefore imposes a definite recency scale on a store in which amplitude decays, and the scale is set by γ alone: a trace loses half its score at about one decay time, regardless of its alignment. This is a consequence of the closed form of Theorem 6.1 and not a finding about memory.

Chapter 30

Memory Compression and Future Admissibility

Compression of a store is acceptable not when it reduces storage by some factor but when it preserves what future reconstruction needs. Let $\mathcal{Q}$ be a family of admissible future queries, let C be a compression map applied to the memory state, and let d be a distance on reconstructions. The objective is a bound

$$\sup_{q \in \mathcal{Q}} d\big(\mathcal{E}(q, M), \mathcal{E}(q, C(M))\big) \le \varepsilon, \tag{30.1}$$

which ties the permitted loss directly to preserved reconstructive capacity and not to a storage criterion. Two limiting cases clarify the content of the bound. If C factors through the quotient by recollective equivalence, in the sense that $C(M)$ determines the class of M under $\sim_{\mathcal{Q}}$, then the supremum is zero for any $\varepsilon \ge 0$, and the compression is lossless with respect to $\mathcal{Q}$; this is the sufficiency theorem of Chapter 32 in the language of compression. At the other extreme, a compression that discards any distinction that some query in $\mathcal{Q}$ depends on has a strictly positive supremum, whatever its storage savings.

For the softmax reconstruction the bound can be made explicit.

Proposition 30.1 (Distortion bound for the softmax readout)**.** *Let $M = (m_1, \dots, m_P)$ and $M' = (m'_1, \dots, m'_P)$ be stores with $\|m_i\|, \|m'_i\| \le 1$ and $\max_i \|m_i - m'_i\| \le \delta$, and let $F_M(c) = \sum_i w_i(c) m_i$ with $w = \mathrm{softmax}(\beta \langle m_i, c \rangle)$ for a query c with $\|c\| \le 1$. Then $\|F_M(c) - F_{M'}(c)\| \le (1 + 2\beta)\delta$.*

Proof. The scores differ by at most δ in each coordinate, since $|\langle m_i - m'_i, c \rangle| \le \delta$. The derivative of the softmax is $\beta(\mathrm{diag}(w) - ww^\top)$, and for a perturbation v with $\|v\|_\infty \le \delta$ its image has ℓ_1 norm at most $\beta \sum_i w_i |v_i - \bar{v}| \le 2\beta\delta$, with $\bar{v} = \sum_j w_j v_j$. Integrating along the segment between the two score vectors gives $\|w - w'\|_1 \le 2\beta\delta$. Then $\|F_M - F_{M'}\| \le \|\sum_i (w_i - w'_i) m_i\| + \|\sum_i w'_i (m_i - m'_i)\| \le 2\beta\delta + \delta$. □

The bound shows that the distortion permitted to each trace is inversely proportional to $1 + 2\beta$, so sharper selection, which is useful for discrimination, makes the readout more sensitive to compression error, and therefore trades against compressibility. For the storage cost of achieving a given per-trace distortion, the rate–distortion function of a Gaussian source with independent components of variance σ^2 under squared error is $R(D) = \frac{1}{2} \log_2(\sigma^2 / D)$ bits per component for $D < \sigma^2$ [45]. For a trace with d real components and a target distortion $\|m - m'\|^2 \le \delta^2$, with $D = \delta^2 / d$ per component, this gives

$$R(\delta) = \frac{d}{2} \log_2 \frac{d\sigma^2}{\delta^2} \text{ bits per trace}, \qquad \delta^2 < d\sigma^2, \tag{30.2}$$

as a lower bound on the representation size under this source model, and the budget of Proposition 30.1 then fixes δ from the desired reconstruction accuracy. For low-rank compression of a store matrix, the Eckart–Young theorem states that the best rank-k approximation in the Frobenius norm has error $\big(\sum_{j>k} \sigma_j^2\big)^{1/2}$ in terms of the singular values of the store, and each trace is perturbed by at most that amount, so the singular spectrum of the

store determines how far it can be compressed under a given ε [3]. These are bounds under stated source models, and the monograph treats the sufficiency-based compression, which discards by equivalence class and not by magnitude, as the quantity of primary interest, since it can compress by much more than any magnitude-based method when most of the stored structure is irrelevant to the admissible queries.

How tight is the distortion bound?

Proposition 30.1 bounds the change of the softmax readout under a perturbation of the store by $(1 + 2\beta)\delta$. Two questions about this bound can be answered by computation: how loose it is on typical stores, and whether its linear growth in β is an artifact of the proof.

Worked Example 30.1 (Looseness on random stores)**.** In 300 random stores with $3 \le P \le 19$ traces in dimensions 4 to 29, inverse temperatures drawn uniformly from $[0.5, 15]$, perturbations of size up to 0.1, and five random cues each, the largest observed ratio of $\|F_M(c) - F_{M'}(c)\|$ to $(1 + 2\beta)\delta$ was 0.258. The bound therefore held in every instance and was typically loose by a factor of several, as expected of a worst-case bound on random data. This is an algebraic check of the proposition, with a fixed seed, and it is not a statement about any memory system.

Extension 30.1 (The linear dependence on β is necessary)**.** There are stores and queries for which $\|F_M(c) - F_{M'}(c)\| \ge \frac{\sqrt{3}}{2} \tanh(\beta\delta/2)$ with $\max_i \|m_i - m_i'\| = \delta$. Consequently $\sup \|F_M - F_{M'}\|/\delta \ge \frac{\sqrt{3}}{4}\beta(1 + o(1))$ as $\delta \to 0$ for each fixed β, and the factor $1 + 2\beta$ in Proposition 30.1 cannot be replaced by any function that is $o(\beta)$.

Proof. In $\mathbb{R}^2$ take the cue $c = (1, 0)$ and the traces $m_1 = (\frac{1}{2}, \frac{\sqrt{3}}{2})$ and $m_2 = (\frac{1}{2}, -\frac{\sqrt{3}}{2})$, which are unit vectors with equal scores $\frac{1}{2}$ against c. Perturb m_1 to $m_1' = m_1 + \delta c$. Then $w_1' = 1/(1 + e^{-\beta\delta})$ and $w_2' = 1 - w_1'$, and $F_{M'} - F_M$ has second coordinate $\frac{\sqrt{3}}{2}(w_1' - w_2') = \frac{\sqrt{3}}{2} \tanh(\beta\delta/2)$, since the unperturbed weights are equal and the second coordinates of m_1 and m_2 are opposite. The norm is at least the absolute value of this coordinate, and $\tanh(x) \ge x(1 + o(1))$ as $x \to 0$. □

For $\beta = 10$ and $\delta = 0.01$ the exact ratio $\|F_M - F_{M'}\|/\delta$ is 4.36 against the proposition's bound of 21, and for $\beta = 40$ and $\delta = 0.0025$ it is 17.31 against 81. The ratios to the bound are 0.208 and 0.214, so the bound is tight up to a factor of about five in this family and the proof's linear dependence on β reflects a real sensitivity. A store sharp enough for selective recall in the sense of Chapter 33 is therefore a store whose readout is correspondingly sensitive to quantization or pruning of the traces, and the compression budget ε of (30.1) must be spent at a rate that scales inversely with β.

Synthesis of Part VI

The examples of this part share a structure: each chapter fixes a rate or a sensitivity, and the examples show that the rate is controlled by a small number of quantities that can be read off the model. In consolidation the controlling quantities are the exit rates and transfer fractions, which determine both when the deeper store peaks and how much strength it ever receives (Extension 27.1), so that the delivered fraction $k_1 k_2/(a_1 a_2) = 0.250$ is fixed

by the rates before any content is considered. In reconsolidation the controlling quantity is the contraction factor $1 - \eta w_i$ of the self-referential update, which sets the half-life of distinctness at 34.3 retrievals for the example rates, together with the independence of the evidence that opens the gate (Counterexample 28.1). In forgetting it is the match between the retention rule and the access probabilities (Counterexample 29.1) and the recency scale arcosh φ that the energy-balance factor builds in (Worked Example 29.1). In compression it is the product of the inverse temperature and the perturbation size (Extension 30.1).

The results connect through the temperature and through the sufficiency criterion. A store that is sharp enough for selective retrieval, with β above the thresholds of Chapter 33, is also a store that amplifies perturbations linearly in β, so the compression that consolidation performs when it moves strength into a slower, more overlapping store has a cost that grows with the selectivity it is meant to support. Reconsolidation determines whether the distinctions the slow store receives are distinctions of the world or of the system's own earlier reconstructions, and the gate computation determines whether the evidence for a revision is counted once or several times. Forgetting then decides, under a budget, which of the surviving traces are kept, and Proposition 29.1 is the statement that this decision should be made on predicted access and not on age unless age is a sufficient statistic. All four operations are changes to the stored state, and the criterion of Theorem 32.2 applies after each of them: an operation is admissible with respect to the committed recollections only if it does not merge histories that those recollections distinguish. Part VII makes this the organizing principle, and the viability-preserving equivalence of Extension 37.1 is the form that principle takes when the operations are repeated.

The same caveats hold as for Part VII. The rates are those of linear models chosen so that the solutions are in closed form, the retention probabilities and the report likelihoods are constructed to expose a logical point, and none of these numbers is an estimate of a quantity in a biological or deployed memory. The numerical statements were produced by the script `examples.py` and cross-checked against the matrix exponential or by direct simulation as noted at each example.

Part VII

Formal Theory of Reconstructive Memory

Chapter 31

State Spaces and Transition Systems

The formal theory begins by fixing what is a state, what is a transition, and what an observer can see. The system state is the tuple $X_t = (M_t, \Psi_t, \Theta_t, Z_t, H_t, B_t)$ of equation (4.1), taking values in a measurable space $\mathcal{X}$, and the stochastic variant of the dynamics is the controlled Markov kernel

$$X_{t+1} \sim P_\vartheta(\,\cdot \mid X_t, c_t, a_t), \tag{31.1}$$

where $c_t \in \mathcal{C} \cup \{\varnothing\}$ is the cue presented at step t (or none), a_t is an action drawn from a finite or compact action set that includes encoding an observation, changing the context, and allocating resources, and ϑ collects the parameters. Three maps complete the specification. The storage map $\Sigma : \mathcal{X} \times \mathcal{O} \to \mathcal{X}$ incorporates an observation o into the persistent component, the observation function $\mathrm{obs} : \mathcal{X} \to \mathcal{W}$ reports whatever an external observer is permitted to read directly, and the ecphoric operator $\mathcal{E} : \mathcal{M} \times \mathcal{C} \times \mathcal{Z} \to \mathcal{Y}$ produces a reconstruction. The deterministic variant is obtained when $P_\vartheta(\cdot \mid x, c, a)$ is a point mass $\delta_{f_\vartheta(x,c,a)}$, and the finite-dimensional approximation replaces $\mathcal{X}$ by a finite-dimensional manifold or by a finite set with the cardinalities fixed by the representation sizes of Chapter 38.

The distinction between the physical state and the accessible mnemonic state is the one that organizes the remaining chapters of the theory. The physical state is X_t in full. The accessible state is its class under recollective equivalence, in the sense of Definition 2.2: two physical states are mnemonically equivalent if no cue in the admissible family can separate them. For this reading to be coherent the equivalence classes must evolve consistently, so that two equivalent states are still equivalent after the same cue and action. When that holds the accessible state is itself a Markov system, and when it fails the accessible state is not a function of the accessible state at the previous step, so any prediction of recollection from the accessible state alone is incomplete. The condition is exactly the lumpability condition of Chapter 36, and its failure is not a pathology: a memory in which two behaviorally equivalent states respond differently to a future encoding is a memory with hidden state.

Proposition 31.1 (Accessible state is a Markov system exactly under lumpability)**.** *Let $\pi : \mathcal{X} \to \mathcal{A}$ be the quotient map to accessible states under an equivalence $\sim$ on a finite state space, and let P be a transition kernel for fixed (c, a). The process $\pi(X_t)$ is a Markov chain with transition kernel $\bar{P}(\alpha, \alpha') = \sum_{x' \in \alpha'} P(x, x')$ for every initial distribution if and only if the sum on the right depends on x only through $\pi(x)$.*

Proof. This is the strong lumpability criterion of Kemeny and Snell [24]. Sufficiency: if the block-sum is a function of $\pi(x)$, the conditional law of $\pi(X_{t+1})$ given $X_0, \dots, X_t$ depends only on $\pi(X_t)$ and equals $\bar{P}$. Necessity: if two states x, y in the same block have different block-sums into some block α', an initial distribution concentrated on x and one concentrated on y give the same $\pi(X_0)$ and different laws of $\pi(X_1)$, so $\bar{P}$ cannot be defined. □

Chapter 32

Observability and Reconstructive Identifiability

This chapter contains the central result of the monograph. It is deliberately elementary, in the sense that its proof is a few lines, because its force is in what it makes precise: the difference between a memory that preserves every distinction among histories and a memory that preserves exactly those distinctions that a specified family of future recollections needs.

Let $\mathcal{H}ist$ be a set of histories, $\mathcal{M}$ a set of persistent memory states, $\mathcal{C}$ a set of admissible cues, and $\mathcal{Y}$ a set of reconstructions. An encoding is a map $E : \mathcal{H}ist \to \mathcal{M}$. A target recollection function $T : \mathcal{H}ist \times \mathcal{C} \to \mathcal{Y}$ specifies what should be reconstructed from the history h when the cue c arrives. An ecphoric decoder is a map $D : E(\mathcal{H}ist) \times \mathcal{C} \to \mathcal{Y}$, and its defining feature, which the theorem makes essential, is that it sees the memory state and the cue and nothing else. In particular it does not see the history.

Definition 32.1 (Ecphoric sufficiency)**.** The encoding E is *sufficient* for T on $\mathcal{C}$ if $E(h_1) = E(h_2)$ implies $T(h_1, c) = T(h_2, c)$ for all $c \in \mathcal{C}$. It is *realized* by a decoder D if $D(E(h), c) = T(h, c)$ for all $h \in \mathcal{H}ist$ and $c \in \mathcal{C}$.

Theorem 32.2 (Ecphoric sufficiency criterion)**.** *An encoding E is realized by some decoder if and only if it is sufficient for T on $\mathcal{C}$. When it is realized, the decoder is unique on $E(\mathcal{H}ist) \times \mathcal{C}$.*

Proof. Suppose D realizes E and $E(h_1) = E(h_2) = m$. Then $T(h_1, c) = D(m, c) = T(h_2, c)$ for every c, so E is sufficient. Conversely, suppose E is sufficient. For (m, c) with $m \in E(\mathcal{H}ist)$ choose any h with $E(h) = m$ and set $D(m, c) = T(h, c)$. By sufficiency the value does not depend on the choice of h, so D is well defined, and $D(E(h), c) = T(h, c)$ by construction. Uniqueness on $E(\mathcal{H}ist) \times \mathcal{C}$ holds because any decoder realizing E must take the value $T(h, c)$ at $(E(h), c)$. □

The theorem makes no use of structure on $\mathcal{M}$, $\mathcal{C}$ or $\mathcal{Y}$, and this is a feature. The oscillatory, resonant, attractor and inferential mechanisms of Parts II to V are each a candidate way of computing D within resource limits, and the theorem says that none of them can succeed where the criterion fails. The criterion itself is a purely informational statement about E and T. Its content becomes visible in the minimal sufficient memory.

Definition 32.3 (Recollective profile and equivalence of histories)**.** The profile of a history is the function $\tau_h = T(h, \cdot) : \mathcal{C} \to \mathcal{Y}$. Two histories are recollectively equivalent, $h_1 \approx_T h_2$, if $\tau_{h_1} = \tau_{h_2}$.

Corollary 32.4 (Minimal sufficient memory)**.** *The encoding E is sufficient for T on $\mathcal{C}$ if and only if there is a map $\phi : E(\mathcal{H}ist) \to \mathcal{H}ist/\approx_T$ with $\phi \circ E = [\cdot]_T$, where $[\cdot]_T$ is the quotient map. Equivalently, $\approx_T$ is a coarsening of the kernel equivalence of E. In particular the quotient encoding $[\cdot]_T$ is sufficient and every sufficient encoding factors through it in the sense that it determines it; the cardinality $|\mathcal{H}ist/\approx_T|$ is the smallest number of distinct memory states a sufficient encoding can use.*

Proof. Sufficiency says that the kernel equivalence of E refines $\approx_T$, which is precisely the existence of ϕ with $\phi \circ E = [\cdot]_T$ (a map defined on the image of E by $\phi(E(h)) = [h]_T$, well

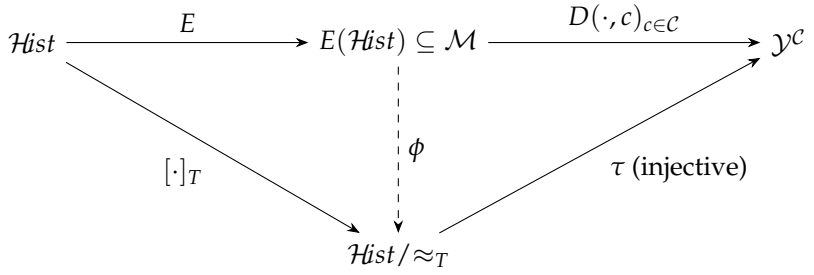


Figure 32.1: Sufficiency as factorization (Theorem 32.2 and Corollary 32.4). The profile map τ sends a class of histories to its recollective profile; E is sufficient exactly when the dashed map ϕ exists, and the decoder is then $D = \tau \circ \phi$.

defined by refinement). For the cardinality, ϕ is onto $\mathcal{H}ist / \approx_T$ because every class contains some h, hence $|E(\mathcal{H}ist)| \geq |\mathcal{H}ist / \approx_T|$. □

The corollary is the formal content of the phrase that memory preserves conditions for reinstatement and not the history itself. Two histories that differ in ways that no admissible cue can reveal through T need not be distinguished by the memory, and any storage that does distinguish them is surplus with respect to T. The same distinctions may of course matter for a different target T', which is why the target must be fixed as part of the specification and not left implicit. The example of Chapter 1 shows the failure in its smallest form: for $E(a, b) = e_a + e_b$ one has $E(a, b) = E(b, a)$, and a target that returns the first item of the pair has $T((a, b), c) \neq T((b, a), c)$ for every cue that does not reveal order, so no decoder exists however the retrieval dynamics are designed.

The exact criterion is brittle in one respect, since real decoders return approximations. The approximate version is stated for a metric d on $\mathcal{Y}$.

Theorem 32.5 (Approximate sufficiency)**.** *Let $(\mathcal{Y}, d)$ be a metric space and $\varepsilon \geq 0$. For $m \in E(\mathcal{H}ist)$ and $c \in \mathcal{C}$ write $A(m, c) = \{T(h, c) : E(h) = m\}$. A decoder with $d\big(D(E(h), c), T(h, c)\big) \leq \varepsilon$ for all h, c exists if and only if every set $A(m, c)$ is contained in a closed ball of radius ε. A necessary condition is* $\operatorname{diam} A(m, c) \leq 2\varepsilon$ *for all (m, c), and if every set of diameter δ in $\mathcal{Y}$ lies in a ball of radius $\delta/2$, which holds on the real line and, more generally, in hyperconvex spaces such as ℓ_∞^k, the condition* $\operatorname{diam} A(m, c) \leq 2\varepsilon$ *is also sufficient.*

Proof. Given such a decoder, the center $D(m, c)$ is within ε of every element of $A(m, c)$, so the set lies in the ball $\bar{B}(D(m, c), \varepsilon)$. Conversely, choose for each (m, c) a center $y_{m,c}$ of a closed ball of radius ε containing $A(m, c)$ and set $D(m, c) = y_{m,c}$. Necessity of the diameter bound follows from the triangle inequality through the center, giving $d(y, y') \leq 2\varepsilon$ for $y, y' \in A(m, c)$. For the final statement, if $\operatorname{diam} A \leq 2\varepsilon$ and the set lies in a ball of radius $\operatorname{diam} A/2 \leq \varepsilon$, the first part applies. □

The factor of two between the necessary and the sufficient conditions is real in general metric spaces and is the price of not assuming that sets of a given diameter have centers close to their elements. In Euclidean space Jung's theorem reduces the gap from a factor of two to a factor of $\sqrt{2}$, a refinement that does not change the structure of the argument.

The next result quantifies what the absence of sufficiency costs. It concerns the recovery of the history itself, which is the extreme target $T(h, c) = h$.

Proposition 32.6 (Identifiability limit)**.** *Let $\mathcal{H}ist$ be finite with prior p and let E be deterministic. The maximum over decoders $\hat{h} : E(\mathcal{H}ist) \to \mathcal{H}ist$ of the probability of recovering h is*

$$\max_{\hat{h}} \Pr[\hat{h}(E(h)) = h] = \sum_{m \in E(\mathcal{H}ist)} \max_{h : E(h) = m} p(h), \tag{32.1}$$

which equals $|E(\mathcal{H}ist)|/|\mathcal{H}ist|$ under the uniform prior. Moreover, for any estimator, Fano's inequality gives

$$\Pr[\hat{h}(E(h)) \neq h] \;\geq\; \frac{H(h \mid E(h)) - \log 2}{\log |\mathcal{H}ist|}. \tag{32.2}$$

Proof. Given m, the decoder must name one history in the fiber $E^{-1}(m)$, and the probability of being correct conditional on m is maximized by the history of greatest prior mass; summing over fibers weighted by the probability of the fiber gives the first expression. Under the uniform prior each fiber has maximum mass $1/|\mathcal{H}ist|$ and there are $|E(\mathcal{H}ist)|$ fibers. Fano's inequality is the standard inequality [45] applied to the variable h and its estimate through $E(h)$, and $H(h \mid E(h)) = H(h) - H(E(h))$ for deterministic E. □

Two aspects of this bound deserve emphasis. It is independent of the retrieval mechanism, so no amount of recurrent refinement, phase coordination, or learned inference can raise the accuracy above $|E(\mathcal{H}ist)|/|\mathcal{H}ist|$ when only the memory state is consulted. And it concerns recovery of the whole history, so it is an upper bound on the ability to answer every possible question, not on the ability to answer the questions in T; Theorem 32.2 is the relevant criterion when the target is narrower. The two together define the extent to which statements about the fidelity of a memory system are statements about the encoding and not about the decoder. A general lesson is that observation of poor reconstruction does not by itself implicate the retrieval dynamics, and a diagnostic that separates the two is to hold the encoding fixed and ask whether Theorem 32.2 is satisfied for the target of interest.

Two qualifications limit how the theorem may be used. The first is that the decoder is assumed to have no additional information; if the decoder can consult the ledger of Chapter 41 or any other store, then that store is part of the memory state and the criterion applies to the combined encoding. The second is that sufficiency is a property of the pair (E, T) together with the cue family, so the claim that a system is sufficient is empty until the cue family is stated, and a system can be sufficient for one family and insufficient for a larger one; this is the formal reason that a design must record the family of future recollections it commits to supporting.

A finite history space: destroyed and preserved distinctions

The criterion is easiest to appreciate on a history space small enough to enumerate. The material of this section is a worked example followed by a proposed extension to noisy encodings; the first is an exercise in Theorem 32.2, and the second is proved here from elementary decision theory and is not drawn from the memory literature.

Worked Example 32.1 (Three encodings of a six-element history space)**.** Let a history record what was encountered and where, $\mathcal{H}ist = \{a, b, c\} \times \{0, 1\}$, so that $h = (x, y)$ has $|\mathcal{H}ist| = 6$, and let the prior be uniform. A single future cue asks where the item was seen, so $\mathcal{C} = \{c_{\text{where}}\}$ and $T(h, c_{\text{where}}) = y$. Consider three encodings: $E_1(x, y) = x$, which keeps

the identity of the item and discards the location; $E_2(x, y) = y$, which keeps the location and discards the identity; and $E_3(x, y) = (\bar{x}, y)$ with $\bar{a} = \bar{b} = \mathsf{ab}$ and $\bar{c} = \mathsf{c}$, which keeps the location and merges two of the three items. The kernel classes are

$$\ker E_1 = \big\{\{(a,0),(a,1)\},\{(b,0),(b,1)\},\{(c,0),(c,1)\}\big\}, \tag{32.3}$$

$$\ker E_2 = \big\{\{(a,0),(b,0),(c,0)\},\{(a,1),(b,1),(c,1)\}\big\}, \tag{32.4}$$

$$\ker E_3 = \big\{\{(a,0),(b,0)\},\{(a,1),(b,1)\},\{(c,0)\},\{(c,1)\}\big\}, \tag{32.5}$$

with 3, 2 and 4 classes respectively. The recollective equivalence $\approx_T$ has exactly two classes, $\{y = 0\}$ and $\{y = 1\}$, since τ_h depends on h only through y. By Corollary 32.4 an encoding is sufficient exactly when its kernel is contained in $\approx_T$. The kernel of E_1 is not contained in $\approx_T$, because the class $\{(a,0),(a,1)\}$ joins two histories with different locations. The kernels of E_2 and E_3 are contained in $\approx_T$, so both are sufficient, and E_2 is the minimal one with $|\mathcal{H}ist/\approx_T| = 2$ memory states.

For the sufficient encodings the decoder is the one constructed in the proof of Theorem 32.2: choose a representative of each fiber and read off its location. For E_2 it is the table $D_2(0, c_{\text{where}}) = 0$ and $D_2(1, c_{\text{where}}) = 1$, and for E_3 it is $D_3((\mathsf{ab}, y), c_{\text{where}}) = D_3((\mathsf{c}, y), c_{\text{where}}) = y$. The tables are explicit and the check $D(E(h), c_{\text{where}}) = T(h, c_{\text{where}})$ passes on all six histories. For E_1 no table exists. The fiber $\{(a,0),(a,1)\}$ forces a single answer for two different targets, so any decoder errs on at least one of them, and under the uniform prior the best decoder answers correctly with probability one half.

The example also separates two quantities that are easily conflated. By Proposition 32.6 the accuracy of recovering the whole history is $|E(\mathcal{H}ist)|/|\mathcal{H}ist|$, which equals 0.500 for E_1, 0.333 for E_2 and 0.667 for E_3. The encoding that is worst for the stated target, E_1, is better than the best minimal encoding at recovering histories, and the ordering of the three encodings by identifiability is not the ordering by sufficiency. This is the quantitative form of the statement that fidelity to the history and sufficiency for a family of recollections are different properties.

Counterexample 32.1 (Sufficiency is relative to the cue family)**.** Enlarge the cue family by a second cue, c_{what}, with $T(h, c_{\text{what}}) = x$. The recollective equivalence on $\mathcal{H}ist$ is now the identity, since τ_h determines both coordinates. Then E_3, which was sufficient for c_{where}, fails: the class $\{(a,0),(b,0)\}$ joins histories with different answers to c_{what}. The best decoder for the new cue under the uniform prior has error 0.333, because in each of the two merged classes the decoder must guess between two equally likely items. The encoding E_2 fails for the same reason. Only the identity encoding is sufficient for both cues, and it needs all six states. The example shows that the claim "the encoding is sufficient" carries an implicit quantifier over the cue family, which is why the family must be recorded in the specification.

The three failures above were exact. A deployed memory is noisy, so the memory state is a random variable whose law depends on the history, and the right question is how small the error can be made when the criterion holds only up to a channel.

Extension 32.1 (Bayes risk of noisy ecphory)**.** Let $\mathcal{H}ist$ be finite with prior π, and let the encoding be a stochastic kernel $K(m \mid h)$ on a finite set $\mathcal{M}$. Fix a cue c and write $y = T(h, c)$.

The minimum over all decoders $D : \mathcal{M} \to \mathcal{Y}$ of $\Pr[D(m) \neq T(h,c)]$ is

$$R^*(c) \;=\; 1 - \sum_{m \in \mathcal{M}} \max_{y} \sum_{h:\, T(h,c)=y} \pi(h) K(m \mid h). \tag{32.6}$$

(i) For deterministic K one has $R^*(c) = 0$ if and only if E is sufficient for $T(\cdot, c)$. For general K, $R^*(c) = 0$ if and only if the conditional laws of m given different target values have disjoint supports. *(ii)* If the target is binary with classes $\mathcal{H}ist_0, \mathcal{H}ist_1$, class priors π_0, π_1 and class-conditional laws $P_j(m) = \Pr[m \mid h \in \mathcal{H}ist_j]$, then $R^* = \sum_m \min\{\pi_0 P_0(m), \pi_1 P_1(m)\}$ and

$$\tfrac{1}{2}\Big(1 - \sqrt{1 - 4\pi_0\pi_1\,\mathrm{BC}^2}\Big) \;\le\; R^* \;\le\; \sqrt{\pi_0\pi_1}\,\mathrm{BC}, \qquad \mathrm{BC} = \sum_m \sqrt{P_0(m) P_1(m)}. \tag{32.7}$$

(iii) Garbling cannot help: if $K' = GK$ for a further channel G from $\mathcal{M}$ to $\mathcal{M}'$, then $R^*_{K'}(c) \ge R^*_K(c)$.

Proof. Given m, the posterior probability that the target equals y is proportional to

$$\sum_{h:\, T(h,c)=y} \pi(h) K(m \mid h),$$

so the decoder that names the maximizer is optimal, which gives the first formula. For (i), $R^* = 0$ forces, for every m with positive mass, that all of the mass at m comes from a single target value, which is the disjoint-support statement; for deterministic K this says each fiber of E lies in one target class, which is Definition 32.1. For (ii), write $a_m = \pi_0 P_0(m)$ and $b_m = \pi_1 P_1(m)$, so $\sum_m (a_m + b_m) = 1$ and $\sum_m \min\{a_m, b_m\} = \tfrac{1}{2}\big(1 - \sum_m |a_m - b_m|\big)$. Since $|a - b| = \sqrt{(a+b)^2 - 4ab}$, the Cauchy–Schwarz inequality gives

$$\sum_m |a_m - b_m| \le \Big(\sum_m (a_m + b_m)\Big)^{1/2} \Big(\sum_m \Big[(a_m + b_m) - \tfrac{4 a_m b_m}{a_m + b_m}\Big]\Big)^{1/2} = \Big(1 - 4\sum_m \tfrac{a_m b_m}{a_m + b_m}\Big)^{1/2}. \tag{32.8}$$

A second application of Cauchy–Schwarz, $\big(\sum_m \sqrt{a_m b_m}\big)^2 \le \sum_m \frac{a_m b_m}{a_m + b_m} \sum_m (a_m + b_m)$, gives $\sum_m a_m b_m / (a_m + b_m) \ge \pi_0\pi_1 \mathrm{BC}^2$, which yields the lower bound. The upper bound follows from $\min\{a, b\} \le \sqrt{ab}$. For (iii), a decoder for m' composed with G is a randomized decoder for m, and randomization cannot beat the optimal deterministic decoder. $\square$

Part (i) is the noisy counterpart of Theorem 32.2: exact recollection requires that the memory state separate target classes without overlap, and the sufficiency criterion is the case in which the separation is deterministic. Part (iii) is the quantitative content of the remark that retrieval dynamics cannot repair an insufficient encoding: any noise or compression applied downstream of an encoding can only add to its floor. The lower bound in (32.7) depends on the encoding only through the Bhattacharyya coefficient of Chapter 35, which ties practical recoverability to the information geometry of the memory state, and the bound is of the Hellman–Raviv and Kailath type, proved above and checked numerically on two thousand randomly drawn binary channels (Appendix M).

Worked Example 32.2 (Noise on a sufficient encoding)**.** Apply the bit-flip channel with flip probability $p = 0.10$ to the minimal encoding E_2 of Worked Example 32.1. The two class-conditional laws are $(1 - p, p)$ and $(p, 1 - p)$, the priors are one half, and (32.7)-(ii)

gives, summing $\min\{\frac{1}{2}(1-p), \frac{1}{2}p\} = \frac{1}{2}p$ over the two memory symbols, $R^* = p = 0.10$. The decoder is the unchanged table D_2 read at the noisy symbol. A continuous version places the memory state on the real line, $m = \pm 1 + \sigma\varepsilon$ with ε standard normal and $\sigma = \frac{1}{2}$. The optimal decoder is the sign of m and the risk is $\Phi(-1/\sigma) = 0.02275$, which a Monte Carlo run of 10^6 draws reproduces as 0.0228. Here $\mathrm{BC} = e^{-1/(2\sigma^2)} = 0.1353$, and (32.7) brackets the risk between 0.0046 and 0.0677. The lower bound is loose by a factor of about five in this regime, which is typical of a bound that holds uniformly over channels, but its direction is the useful one: it shows that the risk cannot be made small without reducing the overlap of the two class-conditional laws.

The final quantity connects back to the insufficient encoding. For E_1 followed by any symmetric channel the risk remains at least one half, as part (iii) predicts, and the numerical check confirms it for noise levels 0, 0.1 and 0.3. The noisy analysis therefore does not soften the exact theorem; it states how fast failure sets in as the separation between target classes is eroded.

Dynamic ecphoric sufficiency

The criterion of Theorem 32.2 concerns a fixed encoding and a fixed family of cues. An adaptive memory changes its state by consolidation, compression, merging and release, and the family of questions it will be asked changes with it. Let $E_0 : \mathcal{H}ist \to \mathcal{M}_0$ be an initial encoding and let $G_t : \mathcal{M}_t \to \mathcal{M}_{t+1}$ be transformations of the memory state, chosen from a stream of external observations that does not depend on which history is stored, so that for a given stream the G_t are fixed functions. Write $E_t = G_{t-1} \circ \cdots \circ G_0 \circ E_0$ for the encoding at time t and $K_t = \ker E_t$ for its kernel equivalence. At each time the system is committed to a target T_t on a cue family Q_t, with recollective equivalence $\approx_t$ of Definition 32.3.

Theorem 32.7 (Dynamic ecphoric sufficiency)**.** *Let $\approx_{\geq t} = \bigcap_{\tau \geq t} \approx_\tau$ be the equivalence under the joint profile of all future targets, and let the horizon be $T_{\max} \leq \infty$. Then The following hold. (i) the kernels are monotone, $K_t \subseteq K_{t+1}$, so information about the history can only be lost. (ii) E_τ is sufficient for (T_τ, Q_τ) if and only if $K_\tau \subseteq \approx_\tau$. (iii) the system can answer every committed family from time t to $T_{\max}$ whatever transformations are applied in between, in the sense that E_τ is sufficient for (T_τ, Q_τ) at every $\tau \geq t$ for every admissible sequence of further transformations, if and only if $K_t \subseteq \approx_{\geq t}$ and the further transformations preserve $K_\tau \subseteq \approx_\tau$; in particular $K_t \subseteq \approx_{\geq t}$ is necessary, and it is sufficient when no further transformation is applied. (iv) a transformation G_t is admissible, meaning that it leaves the system able to answer all committed future families if no further compression is applied, exactly when $\ker(G_t \circ E_t) \subseteq \approx_{\geq t+1}$, and the most compressive admissible transformation is the quotient by $\approx_{\geq t+1}$ restricted to the image of E_t.*

Proof. (i) If $E_t(h_1) = E_t(h_2)$ then $E_{t+1}(h_1) = G_t(E_t(h_1)) = G_t(E_t(h_2)) = E_{t+1}(h_2)$. (ii) is Theorem 32.2 applied at time τ. (iii) Necessity: by (i) and (ii), sufficiency at every $\tau \geq t$ gives $K_t \subseteq K_\tau \subseteq \approx_\tau$ for all $\tau \geq t$, hence $K_t \subseteq \bigcap_{\tau \geq t} \approx_\tau$. Sufficiency in the case of no further transformation: then $K_\tau = K_t$ for all τ, and $K_t \subseteq \approx_{\geq t} \subseteq \approx_\tau$. (iv) Applying (iii) with $E_{t+1} = G_t \circ E_t$ and no further transformation gives the criterion. The quotient of the image of E_t by the relation induced by $\approx_{\geq t+1}$ is well defined exactly when $K_t \subseteq \approx_{\geq t+1}$, and any admissible G_t factors through it by Corollary 32.4. □

Two consequences deserve statement. The first is irreversibility of insufficiency: if E_t is

not sufficient for a target (T, Q), then by (i) no later transformation, however it is chosen, makes the memory sufficient for it, since the kernel only grows. Repair therefore requires information from outside the memory state. The second is the price of an open future.

Corollary 32.8 (Cost of commitment)**.** *Let $\mathfrak{T} \subseteq \mathfrak{T}'$ be two classes of possible future targets and let $\approx_{\mathfrak{T}}$ be the equivalence under their joint profile. Then $|\mathcal{H}ist/\approx_{\mathfrak{T}}| \leq |\mathcal{H}ist/\approx_{\mathfrak{T}'}|$, so the minimal number of memory states that keeps every target in the class answerable is nondecreasing in the class. If $\mathfrak{T}'$ contains the identity target $T(h, c) = h$, the minimal memory is the history set itself and no lossless compression exists.*

Proof. A larger class gives a finer joint-profile equivalence, since $h \approx_{\mathfrak{T}'} h'$ implies $h \approx_{\mathfrak{T}} h'$, so the quotient has at least as many classes. For the identity target the profile separates all histories. □

The corollary is a precise form of the statement that compression is possible only relative to a commitment about what will be asked: a system that wishes to remain answerable to any future question about its past cannot compress, and a system that compresses must have decided, and recorded, which questions it gives up. It is also where the ledger of Chapter 41 enters in a way that no single component supplies.

Proposition 32.9 (Repair by the ledger)**.** *Suppose the memory state is (M_t, L_t) with M_t a compressed trace store and L_t an append-only ledger whose inscriptions determine the history, so that* $\ker E_t^L = \mathrm{id}$ *for the combined encoding. Then for every target determined by the inscriptions the combined encoding is sufficient at all times, whatever transformations are applied to M_t, and a trace store that has lost sufficiency for a newly committed family can be rebuilt from L_t. The rebuilt store has the kernel $\approx_{\geq t}$ of the new family, and the cost of repair is the cost of the replay of Proposition 44.2.*

Proof. With the identity kernel, $K_\tau = \mathrm{id} \subseteq \approx_\tau$ for every target τ, so (ii) of Theorem 32.7 gives sufficiency, and the ledger is never compressed by the G_t that act on M_t alone. Rebuilding applies an encoding E' defined from the inscriptions, whose kernel is chosen as the quotient by the new joint profile and is therefore sufficient by Corollary 32.4. □

The proposition separates two roles that the ledger and the traces play and ties them to the two notions of Chapter 32. Information preservation is guaranteed by the ledger and never lost; resource-bounded accessibility is provided by the traces, which may lose sufficiency irreversibly but can be rebuilt at a cost. What the results do not establish is a rate: how often a trace store must be rebuilt, and what the optimal compression policy is when future families arrive stochastically, are open (Problem 59.7). The theorem also assumes that the stream of observations used by the transformations is independent of the stored history; transformations driven by the system's own retrievals, which depend on the history through the memory state, require a different analysis.

Chapter 33

Stability of Recurrent Reconstruction

Many of the decoders considered in the monograph are iterated maps. A recurrent reconstruction computes $r^{(k+1)} = F(r^{(k)}, c)$ from an initial state, and the questions of this chapter are whether the iteration converges, to what, and with what sensitivity to its inputs. The simplest sufficient condition is contraction. Let $(\mathcal{R}, d)$ be a complete metric space.

Lemma 33.1 (Banach bounds)**.** *If $d(F(x,c),F(y,c)) \le \lambda\, d(x,y)$ with $0 \le \lambda < 1$ for all x, y, then $F(\cdot, c)$ has a unique fixed point $r^\star(c)$ and for every initial state x one has $d(F^k(x), r^\star) \le \lambda^k d(x, r^\star)$, the a priori bound $d(F^k(x), r^\star) \le \frac{\lambda^k}{1-\lambda} d(x, F(x))$, and the residual bound $d(x, r^\star) \le \frac{d(x,F(x))}{1-\lambda}$. If in addition $d(F(x,c),F(x,c')) \le L_c\, d(c,c')$ for all x, then $d(r^\star(c), r^\star(c')) \le \frac{L_c}{1-\lambda}\, d(c,c')$.*

Proof. Existence, uniqueness and the first bound are the Banach fixed-point theorem. For the residual bound, $d(x, r^\star) \le d(x, F(x)) + d(F(x), F(r^\star))$, and the last term is at most $\lambda\, d(x, r^\star)$. The a priori bound follows by applying the residual bound at $F^k(x)$ and $d(F^k(x), F^{k+1}(x)) \le \lambda^k d(x, F(x))$. For the cue sensitivity, write $\Delta = d(r^\star(c), r^\star(c'))$ and use the triangle inequality through $F(r^\star(c), c')$:

$$\Delta \le d\big(F(r^\star(c), c), F(r^\star(c), c')\big) + d\big(F(r^\star(c), c'), F(r^\star(c'), c')\big) \le L_c\, d(c,c') + \lambda \Delta,$$

and rearrange. □

For the softmax update of Chapter 16 the contraction condition can be made concrete. The map $F(r) = \sum_i w_i(r) m_i$ is smooth with Jacobian $DF(r) = \beta\, \mathrm{Cov}_{w(r)}(m)$, the covariance matrix of the traces under the weights $w(r)$.

Theorem 33.2 (Contraction and local stability of the softmax update)**.** *Let $m_1, \dots, m_P$ be unit vectors in $\mathbb{R}^N$. Part (i): for $\beta < 1$ the map F is β-Lipschitz on $\mathbb{R}^N$, so it has a unique fixed point and the iteration converges to it from every initial state at rate β^k. Part (ii): at any fixed point $r^\star$ with weights $w^\star = w(r^\star)$, the iteration is locally asymptotically stable if $\beta\, \lambda_{\max}(\mathrm{Cov}_{w^\star}(m)) < 1$ and locally unstable if $\beta\, \lambda_{\max}(\mathrm{Cov}_{w^\star}(m)) > 1$.*

Proof. The Jacobian is $DF = \beta(\sum_i w_i m_i m_i^\top - \bar m \bar m^\top) = \beta\, \mathrm{Cov}_w(m)$ with $\bar m = \sum_i w_i m_i$, a positive semidefinite matrix. Its largest eigenvalue is at most its trace, $\mathrm{tr}\, \mathrm{Cov}_w(m) = \sum_i w_i \|m_i\|^2 - \|\bar m\|^2 \le 1$. Hence $\|DF\| \le \beta$ everywhere, and by the mean value inequality F is β-Lipschitz, which gives (i) by Lemma 33.1. For (ii), a fixed point of a C^1 map is locally asymptotically stable if the spectral radius of the Jacobian is below one and unstable if some eigenvalue has modulus above one; since $DF = \beta\, \mathrm{Cov}_{w^\star}$ is symmetric positive semidefinite, its spectral radius is $\beta \lambda_{\max}$. □

Part (i) says that for $\beta < 1$ the memory is a global contraction toward a single blend, since weights hardly depend on the state; this is the regime in which the store behaves as an averaging device. Part (ii) explains how the blend at low β loses stability as β grows: at a symmetric blend the covariance $\mathrm{Cov}_{w^\star}(m)$ approaches the covariance of the traces,

and the threshold $\beta\lambda_{\max} = 1$ is the point at which the blend becomes unstable and single-trace or cluster attractors take over. At a well-separated single-trace attractor the weights concentrate and $\mathrm{Cov}_{w^\star}$ is exponentially small, so the fixed point is strongly stable, which is the one-step cleanup of Proposition 16.1 seen from the dynamical side.

Three propositions are easily run together when a reconstruction is said to converge, and the monograph separates them. A recurrent reconstruction converges dynamically if the iterates approach a fixed point; this is a property of the map $F(\cdot, c)$ alone. It converges representationally if the fixed point lies within a stated distance of a stored trace; this is a property of the map and the store. It converges epistemically if the fixed point lies within a stated distance of the target recollection $T(h, c)$ for the history that produced the memory; this is a property of the map, the store, and the history, and it is the only one of the three that bears on fidelity. The implications run in one direction only: epistemic convergence to a target that is a stored trace entails representational convergence, which entails dynamical convergence of the iteration that reaches it, and none of the converses holds. The contraction regime of Theorem 33.2(i) in fact sits at the wrong end of this ordering for selective recall, as the following proposition shows.

Proposition 33.3 (The globally contractive regime is not selective)**.** *Let $m_1, \dots, m_P$ be unit vectors with maximal pairwise overlap $\mu = \max_{i \neq j} \langle m_i, m_j \rangle$, let $\beta < 1$, and let $r^\star$ be the unique fixed point of F. Then every weight satisfies $w_j(r^\star) \leq e^{2\beta}/P < e^2/P$, and for every stored trace*

$$\|r^\star - m_j\|^2 \;\geq\; 1 - 2\mu - 2(1-\mu)\,\frac{e^2}{P}. \tag{33.1}$$

Proof. The fixed point lies in the convex hull of the traces, so $\|r^\star\| \leq 1$ and every score $\langle m_i, r^\star \rangle$ lies in $[-1, 1]$. Hence $w_j = e^{\beta s_j}/\sum_i e^{\beta s_i} \leq e^{\beta}/(Pe^{-\beta}) = e^{2\beta}/P$. Then $\langle m_j, r^\star \rangle = w_j + \sum_{i \neq j} w_i \langle m_j, m_i \rangle \leq \mu + (1-\mu) w_j$, and $\|r^\star - m_j\|^2 = \|r^\star\|^2 + 1 - 2\langle m_j, r^\star \rangle \geq 1 - 2\mu - 2(1-\mu) w_j$. □

For nearly orthogonal traces (μ small) and a store of more than a few traces the fixed point is therefore at distance close to one from every trace: global convergence is bought by giving up selection. Selective recall requires the multistable regime $\beta\lambda_{\max} > 1$ at the blend, in which there are several basins and the guarantees of Proposition 16.1 and Theorem 34.2 replace the global contraction guarantee.

Contraction is not correctness. The monograph insists on this point because the convergence results above are easy to mistake for evidence of faithful recollection. A contraction converges to its unique fixed point, and the fixed point is a function of the cue and of the stored state, not of the history.

Example 33.4 (Contraction without fidelity)**.** The constant map $F(x, c) = x_0$ is a contraction with $\lambda = 0$ for every c, converges in one step, and is perfectly stable to cue perturbations, since $L_c = 0$. It reconstructs the same state for every history and every cue, and no sufficiency criterion is met for any nonconstant target.

The quantity that connects stability to correctness is the residual at the target. For a target reconstruction $y = T(h, c)$, Lemma 33.1 gives $d(y, r^\star(c)) \leq d(y, F(y, c))/(1-\lambda)$, so the iteration recovers the target to within $\delta/(1-\lambda)$ when the target is an approximate fixed point of the dynamics with residual δ. In words, contraction amplifies the target's residual by the factor $(1-\lambda)^{-1}$ and no more; design of the update must therefore make

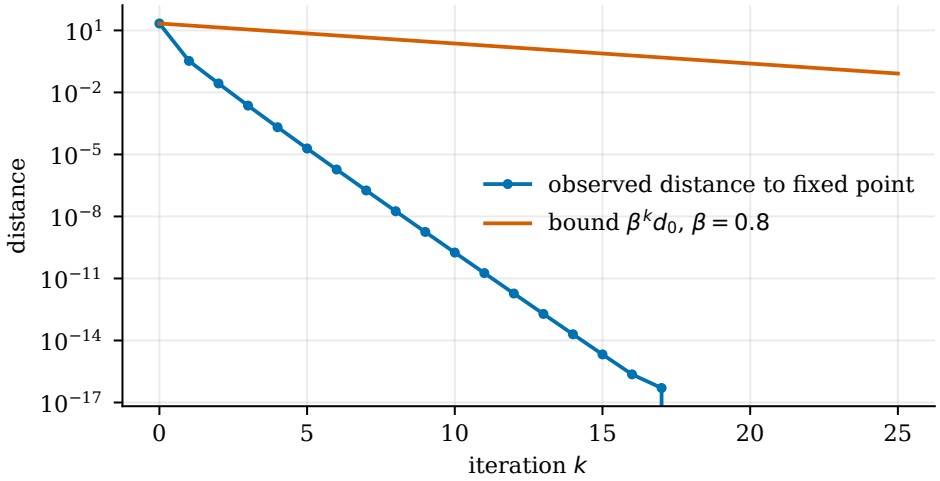


Figure 33.1: Distance to the fixed point for the softmax update with $\beta = 0.8$ on twelve unit traces in $\mathbb{R}^{64}$, with the bound $\beta^k d_0$ of Theorem 33.2(i). Computed by the script `figs.py` with seed 8; an illustration of the stated analytic result and not an evaluation of the architecture.

the targets approximate fixed points, which is a requirement on how traces are written, not on how the iteration is run. The same lemma bounds the sensitivity to cue perturbation by $L_c/(1-\lambda)$, so that robustness to degraded cues and sharpness of the attractor trade against each other through λ.

Other regimes need other tools. If the iteration is linear, $F(x) = Ax$, the long-run behavior is governed by the spectrum of A: convergence if the spectral radius is below one, sustained oscillation if there is a complex pair on the unit circle, and divergence beyond it. For a gradient system $\dot{r} = -\nabla V(r)$ the potential V is itself a Lyapunov function with $\dot{V} = -\|\nabla V\|^2 \leq 0$, as in the Kuramoto and softmax cases of Chapters 12 and 16, and convergence to the critical set of V follows from standard arguments when V is bounded below with compact sublevel sets. Non-gradient terms, such as the dispersive terms of Chapter 13, break this guarantee and admit limit cycles and chaotic reconstructions, which are admissible only if the specification assigns them a meaning.

Figure 33.1 confirms the geometric rate of Theorem 33.2. The observed convergence is faster than the bound, which holds uniformly over all initial states and all stores.

Three traces: the Jacobian, the spectrum, and the loss of stability

The statements of Theorem 33.2 are easiest to appreciate when every quantity can be computed in closed form. The smallest store with a nontrivial competition among traces consists of three orthonormal traces, and this section works it through. Everything below is a worked example of results already proved, together with one extension, the expansion of the unstable branch near the threshold, which is derived here by a direct series computation.

Worked Example 33.1 (Three orthonormal traces)**.** Let m_1, m_2, m_3 be the standard basis of $\mathbb{R}^3$ and consider the cue-free update $r \mapsto F(r) = \sum_i w_i m_i$ with $w = \mathrm{softmax}(\beta r)$, so that $F(r) = w$ and the state lives on the probability simplex. By Chapter 33 the Jacobian at a

point with weights w is

$$J(w) = \beta \operatorname{Cov}_w(m) = \beta\big(\operatorname{diag}(w) - ww^\top\big), \tag{33.2}$$

since $M^\top \operatorname{diag}(w) M = \operatorname{diag}(w)$ for orthonormal traces. At the uniform blend $w = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ the covariance has eigenvalues $\frac{1}{3}, \frac{1}{3}, 0$, the zero belonging to the direction $(1, 1, 1)$ that leaves the simplex, so the Jacobian has leading eigenvalue $\beta/3$ and the blend is locally stable exactly when $\beta < 3$, as Theorem 33.2(ii) requires with $\lambda_{\max}(\mathrm{Cov}) = \frac{1}{3}$. Above that temperature the blend is unstable, and the question is where the iteration goes.

The other fixed points have the symmetric form $r = (a, b, b)$ with $a + 2b = 1$ and $t = a - b$. Substituting into $a = 1/(1 + 2e^{-\beta t})$ gives the scalar equation

$$t = \frac{1 - e^{-\beta t}}{1 + 2e^{-\beta t}}, \tag{33.3}$$

whose solution $t = 0$ is the uniform blend, $t > 0$ is a state dominated by one trace and $t < 0$ is a blend of two traces with the third suppressed. A series expansion about $t = 0$ gives $\frac{1-e^{-\beta t}}{1+2e^{-\beta t}} = \frac{\beta t}{3}\big(1 + \frac{\beta t}{6} + O(t^2)\big)$, hence (33.3) reads $t(1 - \beta/3) = \frac{\beta^2}{18} t^2 + O(t^3)$ and the nonzero branch near the threshold is

$$t \;\approx\; \frac{6(3 - \beta)}{\beta^2}, \tag{33.4}$$

which changes sign at $\beta = 3$. The bifurcation at $\beta = 3$ is therefore transcritical and not a pitchfork, because the branch passes through the blend rather than emerging symmetrically from it. At $\beta = 2.9$ the approximation gives 0.071 and the exact root is 0.078.

Numerical solution of (33.3), checked against the iteration itself, gives the following picture. For $\beta < 2.75$ the blend is the only fixed point, and for $\beta = 2$ the equation has 0 nonzero roots. At $\beta \approx 2.75$ a saddle-node creates a pair of positive roots. On the window $2.75 < \beta < 3$ the blend (leading Jacobian eigenvalue 0.967 at $\beta = 2.9$) and a one-trace state ($t = 0.653$, leading eigenvalue 0.774) are both stable, separated by the unstable branch ($t = 0.078$, leading eigenvalue 1.030), so the system is bistable and the unstable fixed point lies on the boundary between the two basins. At $\beta = 3$ the unstable branch meets the blend, which loses its stability there, and for $\beta > 3$ the branch continues to $t < 0$ as the family of two-trace blends.

At $\beta = 6$ there are three kinds of fixed point besides the blend. The one-trace state $r = (0.9948, 0.0026, 0.0026)$ has Jacobian eigenvalues 0.0463 and 0.0155 (and 0 along the normal), so it is a contraction locally and an attractor. The two-trace blend $r = (0.4839, 0.4839, 0.0322)$ has eigenvalues 2.9034 and 0.2804; the leading one is β times the weight 0.4839 on either of the two surviving traces, with eigenvector $(1, -1, 0)$, and it exceeds one, so the blend is a saddle that separates the basins of the two traces it contains. The uniform blend has leading eigenvalue $2.0 = \beta/3$ and is repelling. The retrieval picture at high temperature is therefore three attractors, one per trace, separated by saddles at the pair blends and with a repeller at the center.

The example shows what the threshold in Theorem 33.2 does and does not control. Crossing $\beta = 3$ is the moment at which the blend ceases to be stable, but selection already becomes possible slightly earlier, on the bistable window, where the basin of the one-trace state grows from nothing. The theorem's condition $\beta\lambda_{\max}(\mathrm{Cov}_{w^\star}) < 1$ is evaluated at a

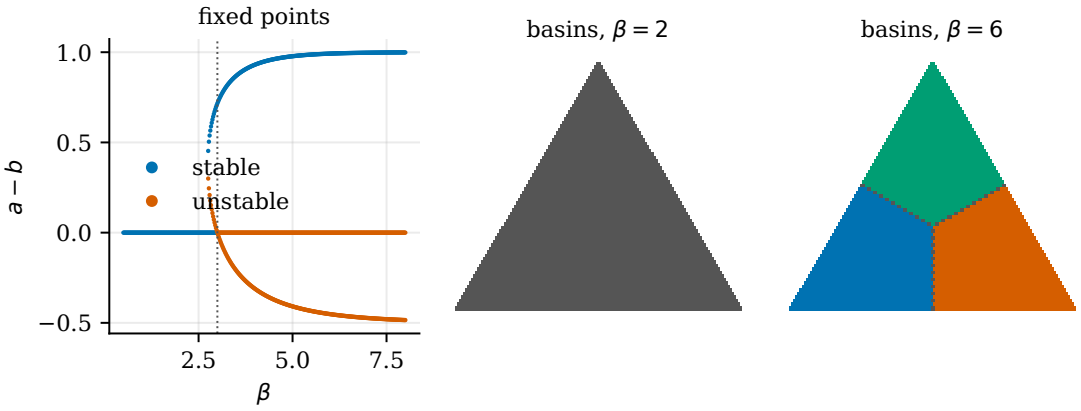


Figure 33.2: The three-orthonormal-trace example. Left: fixed-point branches of (33.3), expressed as the difference $a - b$, against β, with stable points in blue and unstable points in orange; the saddle-node near $\beta = 2.75$ and the crossing of the blend at $\beta = 3$ are visible. Center and right: basins of attraction of the update on the probability simplex for $\beta = 2$ and $\beta = 6$. Below the threshold every initial state reaches the uniform blend (grey); above it the simplex divides into three basins. Computed by `examples.py` with a fixed seed; an illustration of the analytic result.

fixed point, and the example shows that different fixed points of the same dynamics have different spectra, so a statement of the form "the system is stable at temperature β" is meaningless without naming the fixed point.

Extension 33.1 (Overlap shifts the threshold)**.** Let the three traces have pairwise overlap $\mu \in [0, 1)$ and equal norms. By symmetry the blend $r = \bar{m}$ is still a fixed point with uniform weights, and the covariance of the traces under uniform weights has leading eigenvalue $(1 - \mu)/3$. The blend is therefore stable if and only if $\beta < 3/(1 - \mu)$.

Proof. The Gram matrix of the traces is $G = \mu \mathbf{1}\mathbf{1}^\top + (1 - \mu)I$. For uniform weights $\mathrm{Cov}(m) = \frac{1}{3} M^\top (I - \frac{1}{3}\mathbf{1}\mathbf{1}^\top) M$, whose nonzero eigenvalues coincide with those of $\frac{1}{3}(I - \frac{1}{3}\mathbf{1}\mathbf{1}^\top)G$. On the subspace orthogonal to $\mathbf{1}$ the matrix G acts as $(1 - \mu)I$, so the nonzero eigenvalue is $(1 - \mu)/3$ with multiplicity two. □

For $\mu = \frac{1}{2}$ the threshold is 6, and the numerical Jacobian at the blend has leading eigenvalue 0.8333 at $\beta = 5$ and 1.1667 at $\beta = 7$, so it crosses one between them as the formula predicts. The consequence is that correlation among traces postpones the onset of selectivity: the more the stored traces overlap, the colder the system must be, in the sense of a larger β, before the blend loses stability. This is the dynamical counterpart of the capacity statement of Chapter 34, that the number of distinguishable traces is controlled by the overlaps.

Chapter 34

Resonance Selectivity and Capacity

The question of capacity is how many traces can be stored before retrieval degrades, and the answer depends on assumptions about the distribution of traces, the noise in the cue, and what counts as success. This chapter derives bounds under explicit assumptions. They are bounds under a random-pattern model and should not be read as measured properties of any particular dataset.

Random-pattern model. Let the stored traces $m_1, \dots, m_P$ be independent and uniformly distributed on the unit sphere of $\mathbb{C}^n \cong \mathbb{R}^{2n}$. The cap estimate that underlies everything below is that for two independent uniform unit vectors the real overlap satisfies

$$\Pr\big[\mathrm{Re}\langle u, v\rangle \geq a\big] \leq (1-a^2)^{n-1} \leq e^{-(n-1)a^2}, \qquad 0 < a < 1, \tag{34.1}$$

which follows from the standard bound on the measure of a spherical cap [28]. Equation (34.1) was verified numerically in the pilot calculations of Appendix M for $n \in \{16, 32, 64\}$.

Proposition 34.1 (False activation under a resonance threshold)**.** *Let ψ, q be unit vectors in $\mathbb{C}^n$, so that by Theorem 6.1 the resonance score is $S(\psi, q) = \frac{1}{2}\big(1 + \mathrm{Re}\langle \psi, q\rangle\big)$. If a trace is declared active when $S \geq \theta$ with $\theta \in (\frac{1}{2}, 1)$, then for a query q independent of P stored random unit traces, the expected number of false activations satisfies*

$$\mathbb{E}[\#\textit{active}] \leq P\, e^{-(n-1)(2\theta-1)^2}. \tag{34.2}$$

Proof. $S \geq \theta$ is equivalent to $\mathrm{Re}\langle \psi, q\rangle \geq 2\theta - 1$. Apply (34.1) with $a = 2\theta - 1$ to each trace and sum over P traces by linearity of expectation. □

The threshold needed to hold the expected number of false activations below one is therefore $2\theta - 1 \geq \sqrt{\ln P/(n-1)}$, which says that the usable score threshold rises toward one as the store grows at fixed dimension, and the capacity for a given threshold is exponential in the dimension, $P \leq e^{(n-1)(2\theta-1)^2}$. This is the exponential capacity in the dimension that is characteristic of dense associative memories [33, 18], obtained here for the threshold readout, and its dependence on the squared margin shows the cost of lowering the threshold to tolerate degraded cues.

For the recurrent softmax readout the relevant event is that the stored traces are well separated, so that Proposition 16.1 applies.

The result that follows must be read for what it is. It is a capacity theorem for associative retrieval by a softmax readout over random unit vectors, and the exponential scaling is the familiar consequence of high-dimensional concentration together with a sufficiently low temperature [33, 18]. It guarantees one-step proximity to a stored trace. It says nothing about recovering temporal organization, historical provenance, or an episode reconstructed from distributed constraints, and it should not be cited as establishing an exponential capacity for episodic recollection. Its role is to show that the retrieval component of the

architecture can match known forms of high-dimensional pattern separation, and its price, a temperature that grows linearly with the dimension, is stated in the theorem.

Theorem 34.2 (Baseline associative capacity of the softmax readout)**.** *Let $P = e^{\kappa(n-1)}$ traces be drawn as above, and let $a \in (0,1)$ satisfy $a^2 > 2\kappa$. Then with probability at least $1 - \frac{1}{2}e^{(n-1)(2\kappa - a^2)}$ the maximum overlap satisfies $\max_{i \neq j} \mathrm{Re}\langle m_i, m_j\rangle < a$. On this event, for every stored trace m_j and every cue r with $\|r\| \leq 1$ and $\|r - m_j\| \leq \delta$ for some $\delta < (1-a)/2$, one update of the softmax map* (16.1) *with inverse temperature β in the realified space satisfies*

$$\|F(r) - m_j\| \leq 2(P-1)\, e^{-\beta(1-a-2\delta)}. \tag{34.3}$$

In particular the error is below any prescribed $\eta > 0$ once $\beta(1-a-2\delta) \geq \kappa(n-1) + \ln(2/\eta)$.

Proof. There are $\binom{P}{2} \leq P^2/2$ pairs, and the union bound with (34.1) gives the probability that some pair has overlap at least a as at most $\frac{1}{2}P^2 e^{-(n-1)a^2} = \frac{1}{2}e^{(n-1)(2\kappa - a^2)}$. On the complement, for the cue r one has $\langle m_j, r\rangle = 1 + \langle m_j, r - m_j\rangle \geq 1 - \delta$ and, for $i \neq j$, $\langle m_i, r\rangle = \langle m_i, m_j\rangle + \langle m_i, r - m_j\rangle \leq a + \delta$. The margin is at least $\Delta = 1 - a - 2\delta > 0$, and Proposition 16.1 gives the stated bound with $\varepsilon \leq (P-1)e^{-\beta\Delta}$. The final statement solves $2Pe^{-\beta\Delta} \leq \eta$ for $\beta\Delta$ using $\ln P = \kappa(n-1)$. □

The theorem shows, for associative retrieval of stored traces and not for episodic reconstruction, that the number of recoverable traces can grow exponentially in the dimension, with the rate $\kappa < a^2/2$, provided the inverse temperature grows linearly with the dimension, $\beta \gtrsim \kappa(n-1)/(1-a-2\delta)$. The scaling of β is the hidden cost: a memory sharp enough to store exponentially many traces is a memory whose Jacobian $\beta\,\mathrm{Cov}_w(m)$ is large away from the attractors, so its basins are narrow and its blend states unstable (Theorem 33.2). Interference accumulates in a different manner under the dilution formula of Proposition 20.1: adding a competitor with score s_n rescales the weight of every earlier trace by $Z/(Z + e^{\beta s_n})$, which is small only when the competitor's score is low relative to the incumbents' normalizer. The random-pattern model is the best case for capacity, because it makes overlaps small. Structured stores with clusters of near-duplicates have larger μ in Proposition 16.1 and require larger margins, and the correlated-pattern and phase-correlated cases are reserved for the experimental program of Chapter 52.

Figure 34.1 compares the cap bound (34.1) with simulation for three dimensions. The bound holds throughout and has the correct exponential dependence on n and on a^2, which is what the capacity argument of Theorem 34.2 uses.

Why the margin hypothesis matters

The guarantee of Proposition 16.1, and through it Theorem 34.2, is conditional on a margin Δ between the cued trace and its competitors that exceeds $\beta^{-1}\ln(P-1)$. This section shows by example what happens when the hypothesis fails. It also separates two statements that are easily run together, namely that a sufficient condition has failed and that retrieval has failed. The second statement needs its own proof, and the examples give it by exact computation in a family where everything is available in closed form.

Counterexample 34.1 (Near-duplicate traces)**.** Take P unit traces with a common pairwise overlap μ, that is, with Gram matrix $G = \mu\mathbf{1}\mathbf{1}^\top + (1-\mu)I$, and cue with the first trace,

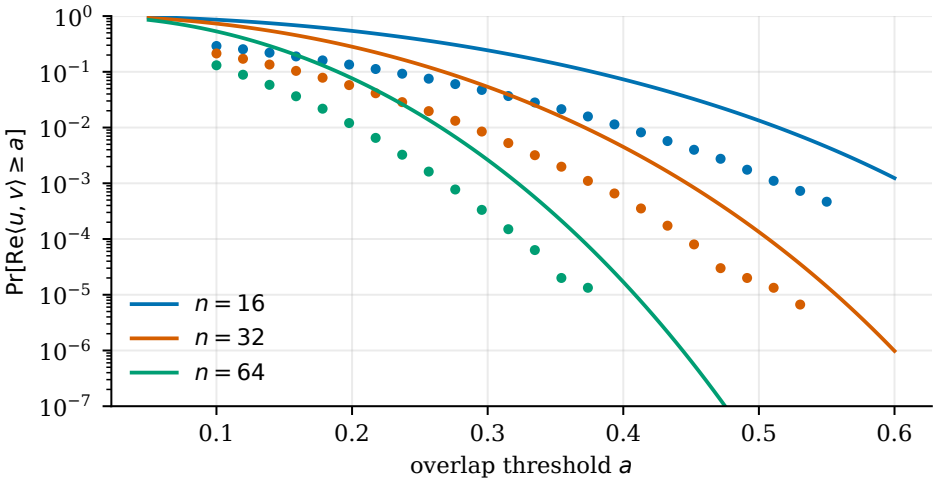


Figure 34.1: Empirical tail probability of the real overlap of two independent uniform unit vectors in $\mathbb{C}^n \cong \mathbb{R}^{2n}$ (points, 3×10^5 samples) with the cap bound $(1-a^2)^{n-1}$ (curves). Computed by the script `figs.py` with seed 8; an illustration of the stated analytic result and not an evaluation of the architecture.

$r = m_1$. The margin is $\Delta = 1 - \mu$ for every competitor, the weights are $w_1 = \big(1 + (P-1)e^{-\beta(1-\mu)}\big)^{-1}$ on the cued trace and $(1-w_1)/(P-1)$ on each other trace, and the one-step error is an exact expression,

$$\|F(m_1) - m_1\| = (1 - w_1)\sqrt{\frac{1 + (P-2)\mu}{P-1} + 1 - 2\mu}, \tag{34.4}$$

because $F(m_1) - m_1 = (1 - w_1)(\bar{c} - m_1)$ with $\bar{c}$ the centroid of the competitors. Consider $P = 21$ and $\mu = 0.9$. The weight on the cued trace equals one half exactly at

$$\beta^{*} = \frac{\ln(P-1)}{1-\mu} = 30.0, \tag{34.5}$$

and below this temperature the cued trace does not even carry a majority of the weight. At $\beta = 20$ one finds $w_1 = 0.270$ and a one-step error of 0.237 against a distance 0.324 from m_1 to the centroid of its competitors, so most of the displacement toward the centroid has taken place. The bound of Proposition 16.1 reads $\varepsilon \le (P-1)e^{-\beta\Delta} = 2.707$, which exceeds one and says nothing. At $\beta = 40$ the weight is 0.732, the error is 0.087 and the bound is $0.366 < 1$, so the hypothesis holds and the conclusion is informative. The figure shows the weight w_1 against β for four store sizes together with the threshold $\ln(P-1)/(1-\mu)$ for each, which grows with the logarithm of the store.

The counterexample is stated as a necessary scale and not only as a failure. If the hypothesis $\beta\Delta > \ln(P-1)$ is violated by a symmetric family, then $w_1 < \tfrac{1}{2}$, so the single-step update moves the state further than halfway to the competitor centroid; the margin condition is therefore not an artifact of the proof but the temperature at which a symmetric cluster stops being resolved. The same expression also shows why sharpening the temperature is a costly remedy. For this family the cue-free iteration $r \mapsto F(r)$ started at m_1 converges for

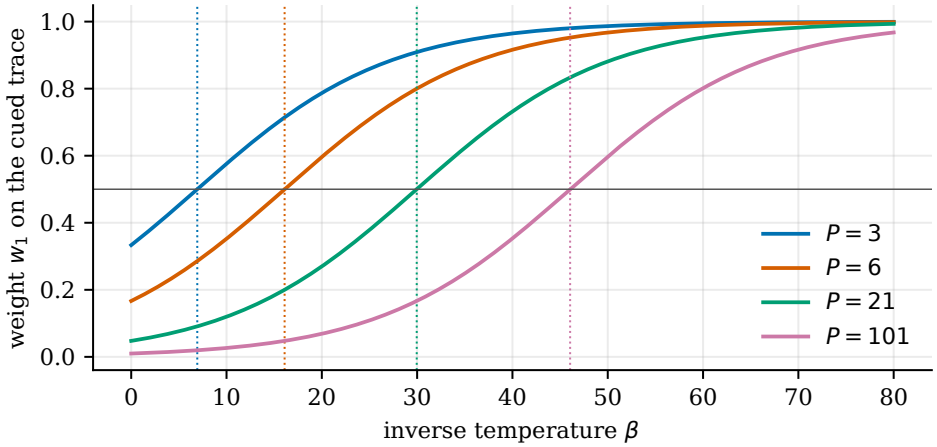


Figure 34.2: Weight on the cued trace against inverse temperature for the symmetric near-duplicate family with $\mu = 0.9$ and $P \in \{3, 6, 21, 101\}$. The dotted verticals mark $\beta^* = \ln(P-1)/(1-\mu)$, where the weight equals one half. Computed by `examples.py`; an illustration of the closed form in Counterexample 34.1.

every β tested (from 10 to 40) to the uniform blend, at distance 0.309 from m_1, because the blend remains stable until β reaches $1/\lambda_{\max} = P/(1-\mu) = 210$, where $\lambda_{\max} = (1-\mu)/P$ is the leading covariance eigenvalue at uniform weights, the computation of Extension 33.1 with P in place of three. In a store of near-duplicates the temperature at which the cue-free iteration becomes selective is therefore much higher than the temperature at which one step resolves the cue, and whether a cue-anchored iteration of the type in Proposition 21.2 holds the cued trace in this regime is not decided by the computation above.

Counterexample 34.2 (A vacuous bound without a failed retrieval)**.** A conclusion that the proposition cannot certify is not a conclusion that is false. Take two traces with overlap $\mu = 0.9$, so $\|m_1 - m_2\| = 0.447$, cue with $r = m_1$ and let $\beta = 10$. The margin is $\Delta = 0.1$, and the bound of Proposition 16.1 is $\|F(r) - m_1\| \le 2(P-1)e^{-\beta\Delta} = 0.74$. This exceeds the distance 0.447 between the two traces, which is the largest displacement possible for a convex combination of them, so the bound is vacuous. The exact computation gives a weight 0.269 on the competitor and a one-step error 0.120, nearly four times smaller than the distance between the traces. The output is closer to m_1 than to m_2, so the retrieval selects the right trace even though the proposition certifies nothing.

The two counterexamples together state what the margin hypothesis can and cannot be used for. Counterexample 34.2 shows that a margin too small to make the bound informative is compatible with correct identification of the nearest trace. Counterexample 34.1 shows that small margins are not harmless either: in the near-duplicate family the uninformative regime coincides with an exactly computed one-step error comparable to the whole displacement toward the competitor centroid, and with a drift of the cue-free iteration to the blend. The logical situation is that Proposition 16.1 is a sufficient condition, so an uninformative bound licenses no inference by itself, and a conclusion of retrieval failure must be supported by a separate computation, as it is in (34.4). For this reason the experimental program of Part IX reports the margin and the observed error together and does not treat the margin as a predictor.

Chapter 35

Information Geometry of Memory

A memory read out through a softmax or through a Bayesian posterior produces, for each cue, a probability distribution over traces or latent causes. The information geometry of these distributions provides a language for distinguishability that does not depend on the coordinates of the stored embeddings. This chapter develops the elementary parts of that language and states plainly what they do and do not establish.

The softmax weights form an exponential family indexed by the inverse temperature: $p_\beta(i \mid r) = \exp(\beta s_i - \Lambda(\beta))$ with $s_i = \langle m_i, r\rangle$ and log-partition $\Lambda(\beta) = \ln \sum_j e^{\beta s_j}$. For an exponential family the Fisher information for the natural parameter is the variance of the sufficient statistic, so that

$$\mathcal{I}(\beta) = \mathrm{Var}_w(s) = \Lambda''(\beta), \tag{35.1}$$

which is the amount of information that a single draw of the selected trace carries about the temperature, and which is zero for every β if the scores are all equal, tends to zero as $\beta \to \infty$ when the weights collapse onto one trace, and at $\beta = 0$ equals the variance of the scores under uniform weights. It can therefore peak at an intermediate temperature where several traces compete, as the example below shows, but it need not do so: the location of the maximum depends on the score profile. This has a consequence for the identifiability analysis of Chapter 23: precision and temperature can be distinguished from outcomes only where this variance is nonzero, and the cue-specific manipulations that make it nonzero are the ones that identify precision.

The space of distributions over a finite set carries a canonical metric, the Fisher–Rao metric $g(\dot{p}, \dot{p}) = \sum_i \dot{p}_i^2 / p_i$. The substitution $u_i = 2\sqrt{p_i}$ maps the open simplex isometrically onto the positive orthant of the sphere of radius two, since $du_i = dp_i / \sqrt{p_i}$ gives $\sum_i du_i^2 = \sum_i dp_i^2 / p_i$. The geodesic distance is therefore

$$d_{FR}(p, q) = 2 \arccos \mathrm{BC}(p, q), \qquad \mathrm{BC}(p, q) = \sum_i \sqrt{p_i q_i}, \tag{35.2}$$

where BC is the Bhattacharyya coefficient. The metric is, up to scale, the only Riemannian metric on the simplex that is invariant under sufficient statistics, a theorem of Čencov that is the usual justification for calling it canonical [41]. The geometry reflects the fact that the probability of confusing two distributions from finite samples is controlled by BC, with large BC meaning near-indistinguishability.

The monotonicity property of this geometry gives a clean bound on what can be recovered from a posterior.

Proposition 35.1 (Data-processing for recollective distinguishability)**.** *Let K be a stochastic kernel from a finite set $\mathcal{X}$ to a finite set $\mathcal{Y}$, and let p, q be distributions on $\mathcal{X}$. Then* $\mathrm{BC}(pK, qK) \geq \mathrm{BC}(p, q)$*, and consequently* $d_{FR}(pK, qK) \leq d_{FR}(p, q)$*.*

Proof. For each y, the Cauchy–Schwarz inequality gives

$$\sqrt{\Big(\sum_x p_x K_{xy}\Big)\Big(\sum_x q_x K_{xy}\Big)} \;\geq\; \sum_x \sqrt{p_x q_x}\, K_{xy}.$$

Summing over y and using $\sum_y K_{xy} = 1$ gives the first inequality, and the second follows because arccos is decreasing. □

If the cues c and c' induce posteriors p_c and $p_{c'}$ over latent causes of Chapter 21, and the reconstruction is a stochastic function of the cause, then the distributions of reconstructions satisfy the proposition's bound: two cues whose posteriors are close in the Fisher–Rao metric cannot produce reconstructions that are further apart in that metric. The geometry of posteriors thus bounds the distinguishability of recollections from above. The proposition does not say that the geometry of posteriors characterizes accessibility, since it is silent about the computational cost of realizing the kernel, and it does not remove the dependence on the stored embeddings, since the posteriors are computed from them. Whether accessibility admits a characterization in terms of the posterior geometry alone, rather than a bound, is left as Problem 35.2.

Worked Example 35.1 (Fisher information of the temperature and a data-processing check)**.** Take three traces with scores $s = (1.0,\, 0.2,\, 0)$ against a cue. The Fisher information $\mathcal{I}(\beta) = \mathrm{Var}_w(s)$ equals 0.187 at $\beta = 0$, where the weights are uniform, rises to a maximum of 0.204 near $\beta = 0.7$, and falls to 0.0000 (to four decimals) at $\beta = 30$, where the first trace carries essentially all the weight. A numerical second difference of the log-partition function reproduces $\mathcal{I}(\beta)$ to within 10^{-5} at $\beta \in \{0.5, 3, 10\}$, which checks (35.1). The practical consequence is the one stated above: a cue of this kind identifies the temperature only in an intermediate band, and the design of a cue family for identifying precision is the design of score profiles whose variance is large where the system is meant to operate. A different profile, for instance $(0.9, 0.5, 0.1)$, has its maximum at $\beta = 0$, which is why the location of the peak cannot be asserted in general.

For the distances, let $p = (0.5, 0.3, 0.2)$ and $q = (0.2, 0.3, 0.5)$ be the posteriors over three latent causes induced by two cues, and let the reconstruction be a stochastic function of the cause through the kernel with rows $(0.9, 0.1)$, $(0.5, 0.5)$ and $(0.1, 0.9)$. Then $\mathrm{BC}(p, q) = 0.9325$ and $d_{FR}(p, q) = 0.7393$, while the output laws satisfy $\mathrm{BC}(pK, qK) = 0.9708$ and $d_{FR}(pK, qK) = 0.4847$. The distance has shrunk, as Proposition 35.1 requires, and the shrinkage quantifies how much of the distinction between the two cues is lost by the stochastic reconstruction.

Open Problem 35.2 (Geometric characterization of accessibility)**.** Define a family of cost-bounded decoders and the induced notion of accessible recollective equivalence. Determine whether the equivalence classes of accessible equivalence are the level sets of a function of the Fisher–Rao geometry of the posteriors $\{p_c\}$ alone, or whether the cost of the decoder introduces invariants of the representation that the posterior geometry does not capture.

Chapter 36

Projection, Abstraction, and Sufficient Memory States

An abstraction is a map $\pi : \mathcal{X} \to \bar{\mathcal{X}}$ from detailed memory states to compressed representations. For a deterministic dynamics $F : \mathcal{X} \to \mathcal{X}$ the abstraction is dynamically adequate if the compressed state evolves autonomously, that is, if there is a map $\bar{F}$ on the compressed space with

$$\pi \circ F = \bar{F} \circ \pi. \tag{36.1}$$

Proposition 36.1 (Closure criterion)**.** *A map $\bar{F}$ satisfying* (36.1) *exists if and only if $\pi(x) = \pi(y)$ implies $\pi(F(x)) = \pi(F(y))$, that is, if the kernel equivalence of π is preserved by F. When it exists it is unique on $\pi(\mathcal{X})$.*

Proof. If (36.1) holds and $\pi(x) = \pi(y)$ then $\pi(F(x)) = \bar{F}(\pi(x)) = \bar{F}(\pi(y)) = \pi(F(y))$. Conversely, if the implication holds define $\bar{F}(\pi(x)) = \pi(F(x))$, which is well defined by hypothesis. □

For stochastic dynamics the same criterion is the strong lumpability of Proposition 31.1. The criterion is about closure of the projected dynamics, and closure alone says nothing about whether the projection keeps what a recollection needs. The following example separates the two ideas.

Example 36.2 (Closure without sufficiency)**.** Take any F and the projection $\pi(x) = 0$ onto a single point. Then $\bar{F}(0) = 0$ satisfies (36.1), and the compressed dynamics is trivially closed. It also discards every distinction among states, so no nonconstant target can be recovered. Conversely, the identity projection preserves everything and is closed, but achieves no compression. A projection that is useful must be closed and must keep the equivalence class that the future queries require.

Proposition 36.3 (Closure and sufficiency together answer future queries)**.** *Suppose $\pi \circ F = \bar{F} \circ \pi$ and the reconstruction factors through the projection, $\mathcal{E}(x, c) = \bar{\mathcal{E}}(\pi(x), c)$. Then for every initial state x_0, every number of steps k and every cue c, the reconstruction after k steps equals $\bar{\mathcal{E}}(\bar{F}^k(\pi(x_0)), c)$, so the compressed system answers every query that the detailed system answers along the trajectory.*

Proof. By induction, $\pi(F^k(x_0)) = \bar{F}^k(\pi(x_0))$, since $\pi(F^{k+1}(x_0)) = \bar{F}(\pi(F^k(x_0)))$. Then $\mathcal{E}(F^k(x_0), c) = \bar{\mathcal{E}}(\pi(F^k(x_0)), c)$. □

Exact closure is rare and approximate closure is the realistic notion. Suppose the commutation defect is bounded, $\|\pi F(x) - \bar{F}\pi(x)\| \le \varepsilon$ for all x, and $\bar{F}$ is $\bar{L}$-Lipschitz. Let $x_k = F^k(x_0)$ and $\bar{x}_k = \bar{F}^k(\pi(x_0))$.

Proposition 36.4 (Drift under approximate commutation)**.** *Under these assumptions $\|\pi(x_k) - \bar{x}_k\| \le \varepsilon \sum_{j=0}^{k-1} \bar{L}^j$, which is $\varepsilon(1 - \bar{L}^k)/(1 - \bar{L})$ for $\bar{L} \neq 1$ and $k\varepsilon$ for $\bar{L} = 1$.*

Proof. Let $e_k = \|\pi(x_k) - \bar{x}_k\|$, with $e_0 = 0$. Then $e_{k+1} \le \|\pi F(x_k) - \bar{F}\pi(x_k)\| + \|\bar{F}\pi(x_k) - \bar{F}(\bar{x}_k)\| \le \varepsilon + \bar{L} e_k$, and the stated sum follows by induction. □

The compressed model is therefore reliable on a horizon of order $1/\varepsilon$ when $\bar{L} \geq 1$, and on all horizons, with error at most $\varepsilon/(1-\bar{L})$, when the compressed dynamics is a contraction. The qualitative implication, that contractive compressed dynamics tolerate imperfect closure indefinitely while expansive dynamics do not, bears on the design of consolidation: the slow store of Chapter 27 should have contractive replay dynamics if it is to remain a faithful abstraction of the fast store.

A linear example: projection onto the span of the traces. The projection that discards cue components orthogonal to the traces is an instance of closure that can be checked exactly. Let $V = \operatorname{span}\{m_1, \ldots, m_P\} \subset \mathbb{R}^N$ and let Π be the orthogonal projector onto V. Because each $m_i \in V$, the scores satisfy $\langle m_i, r\rangle = \langle m_i, \Pi r\rangle$ for every r, so the softmax weights depend on r only through Πr, and the update obeys

$$F(r) = F(\Pi r), \qquad F(r) \in \operatorname{conv}\{m_i\} \subset V, \qquad \text{hence} \qquad \Pi F = F = F\Pi. \tag{36.2}$$

The projection $\pi = \Pi$ is therefore exactly closed with $\bar{F} = F|_V$, and the update is blind to the component of the cue normal to V. If the cue is corrupted by isotropic noise ξ with $\mathbb{E}[\xi\xi^\top] = \sigma^2 I_N$, then $\mathbb{E}\|\Pi\xi\|^2 = \sigma^2 \dim V$, so the fraction of the noise energy that survives the projection is $\dim V / N$, and the remaining fraction $1 - \dim V/N$ is removed in one step. When the stored traces span a low-dimensional subspace of a high-dimensional ambient space, most of the noise is discarded.

This is the linear counterpart of the tangent–normal decomposition developed in the author's companion manuscripts on manifold-aligned prediction [15, 14]. In that framework the semantic manifold $\mathcal{M}$ has at each point a tangent space and its orthogonal complement, the normal space; admissible explanatory updates lie in the tangent space, and components of an update in the normal space are those that depart from the lawful structure of the data, which the companion work identifies with hallucination. The span V here is the globally flat special case in which the tangent space is the same at every point and the normal space is its complement. Equation (36.2) says that Mem8's softmax reconstruction has the no-normal-component property in this special case by construction, so that normal-direction content in an output could only enter through additive noise in the dynamics or from a decoder that is not the softmax readout. The curved case requires local tangent spaces and a projection that depends on the base point, and neither the closure property nor the noise-removal fraction extends to it without curvature corrections; those extensions are stated in the companion work and are not derived here. The claims of the companion manuscripts concerning the empirical behavior of recent generative models are cited as claims of those manuscripts and not asserted independently.

Figure 36.1 checks the linear tangent–normal statement of equation (36.2). The fraction of noise energy surviving the projection matches $\dim V / N$, so a trace span of low dimension discards most of an isotropic perturbation in one step.

Two ways to fail: closure without sufficiency and sufficiency without closure

Example 36.2 separates closure from sufficiency with a degenerate projection that keeps nothing. Two finite constructions make the separation in both directions with projections

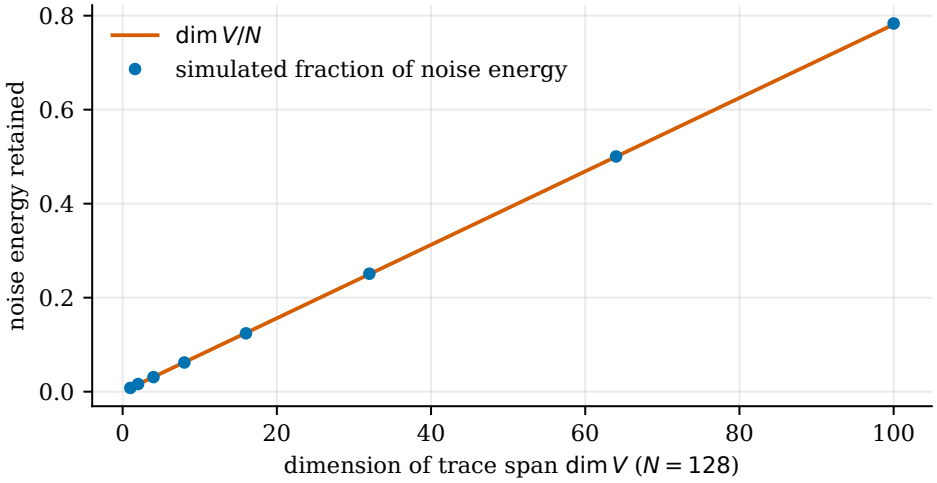


Figure 36.1: Fraction of isotropic noise energy retained by projection onto the span of the traces in $\mathbb{R}^{128}$, against the dimension of the span, with the value $\dim V/N$ of equation (36.2). In each run the softmax update was verified numerically to satisfy $F(r) = F(\Pi r)$. Computed by the script `figs.py` with seed 8; an illustration of the stated analytic result and not an evaluation of the architecture.

that are nontrivial, and a stochastic version places it in the language of transition kernels. These are worked examples and a proposed extension; the criteria they exercise are Propositions 36.1 and 31.1 and Theorem 32.2.

Counterexample 36.1 (Closed but not sufficient)**.** Let $\mathcal{X} = \{1, 2, 3, 4\}$ with the dynamics F that swaps within the pairs, $F(1) = 2$, $F(2) = 1$, $F(3) = 4$, $F(4) = 3$, and let π identify each pair, $\pi(1) = \pi(2) = \mathsf{A}$ and $\pi(3) = \pi(4) = \mathsf{B}$. Then $\pi(x) = \pi(y)$ implies $\pi(F(x)) = \pi(F(y))$, so by Proposition 36.1 the projection is closed, and in fact $\bar{F}$ is the identity on $\{\mathsf{A}, \mathsf{B}\}$. Let the target be the parity of the state within its pair, $T(1) = T(3) = 1$ and $T(2) = T(4) = 0$. The projection is not sufficient for T by Theorem 32.2, because the class $\{1, 2\}$ contains states with different targets, and the failure is robust: no decoder that sees only the compressed state can answer the parity query at any time, since the compressed trajectory is constant and carries no information about it. The distinction destroyed by π is dynamically invisible, which is exactly why closure does not detect its loss. The compressed system is a perfect autonomous model of the block label and an uninformative memory for the parity.

Counterexample 36.2 (Sufficient now but not closed)**.** Let $\mathcal{X} = \{1, 2, 3, 4\}$ with $F(1) = 3$, $F(2) = 1$, $F(3) = 3$, $F(4) = 4$, let $T(1) = T(2) = 1$ and $T(3) = T(4) = 0$, and take the projection $\pi = T$, which has classes $\{1, 2\}$ and $\{3, 4\}$. The projection is sufficient for T by construction, since the reconstruction is the label itself. It is not closed: $\pi(1) = \pi(2)$ but $\pi(F(1)) = \pi(3) = 0$ while $\pi(F(2)) = \pi(1) = 1$. Consequently the compressed system cannot predict its own future label from its present one. The same failure is visible in the language of recollection: the target after one update is $T \circ F$ with $T(F(1)) = 0$ and $T(F(2)) = 1$, so π is not sufficient for $T \circ F$, and a memory that is sufficient for the recollections required today is not sufficient for the recollections required after one step of its own dynamics.

The two counterexamples correspond to two independent defects. Closure is a property of the pair (π, F) and says that the kernel of π is a congruence for F. Sufficiency is a property of the pair (π, T) and says that the kernel of π refines $\approx_T$. A projection supports recollection along a trajectory only when the kernel of π refines $\approx_{T \circ F^k}$ for every k in the horizon, and the coarsest such kernel is the congruence refinement of the targets, which Extension 37.1 of Chapter 37 constructs for a system with inputs. The first counterexample shows that closure is not enough and the second that sufficiency at one time is not enough.

For a stochastic dynamics the same two properties are statements about kernels and conditional laws. Let P be a transition kernel on a finite set $\mathcal{X}$, let π partition $\mathcal{X}$ into blocks, and let $L(x, B') = \sum_{y \in B'} P(x, y)$ be the lumped row. The partition is strongly lumpable if $L(x, \cdot)$ depends on x only through $\pi(x)$, in which case the lumped kernel $\bar{P}(\pi(x), \cdot) = L(x, \cdot)$ is the transition kernel of $\pi(X_k)$ for every initial law, which is Proposition 31.1. Sufficiency for a target T at time k is the statement that the conditional law of $T(X_k)$ given $\pi(X_k)$ is degenerate, which is equivalent to the vanishing of the Bayes risk of Extension 32.1 with the memory state $\pi(X_k)$ and the history X_k.

Worked Example 36.1 (A four-state chain that is lumpable and uninformative)**.** Let $\mathcal{X} = \{0, 1, 2, 3\}$ with blocks $\{0, 1\}$ and $\{2, 3\}$, and let the kernel have rows $(0.35, 0.35, 0.15, 0.15)$ for states 0 and 1 and $(0.10, 0.10, 0.40, 0.40)$ for states 2 and 3. The lumped rows are equal within each block, so the partition is strongly lumpable, with $\bar{P} = \left(\begin{smallmatrix} 0.70 & 0.30 \\ 0.20 & 0.80 \end{smallmatrix}\right)$. The target is the identity of the state within its block, $T(x) = x \bmod 2$. From the uniform initial law the law after one step is $(0.225, 0.225, 0.275, 0.275)$. Given the block label, the two members of each block are equally likely, so the Bayes risk of the parity query is 0.500, which is the worst possible for a binary target, whereas given the full state it is zero. The compressed chain is therefore an exact autonomous model of the block process and carries no information whatever about the target; the closure of the lumped kernel and the insufficiency of the lumped state hold simultaneously, the stochastic form of Counterexample 36.1.

Exact lumpability is also fragile, and the realistic notion is the approximate one. For the stochastic case the drift bound of Proposition 36.4 takes the following form, in which the Lipschitz constant of the compressed dynamics is replaced by the Dobrushin contraction coefficient $\kappa(\bar{P}) = \max_{a,b} \mathrm{TV}(\bar{P}(a, \cdot), \bar{P}(b, \cdot))$.

Extension 36.1 (Drift under approximate lumpability)**.** Let P be a kernel on a finite set with a partition π, and let $\bar{P}$ be a kernel on the blocks with defect $\varepsilon = \max_x \mathrm{TV}\big(L(x, \cdot), \bar{P}(\pi(x), \cdot)\big)$. Let ν_k be the law of $\pi(X_k)$ under P from any initial law of X_0, and let $\bar{\nu}_k = \bar{\nu}_{k-1} \bar{P}$ with $\bar{\nu}_0 = \nu_0$. Then

$$\mathrm{TV}(\nu_k, \bar{\nu}_k) \;\le\; \varepsilon \sum_{j=0}^{k-1} \kappa^j \;=\; \varepsilon \, \frac{1 - \kappa^k}{1 - \kappa} \;\le\; k\varepsilon, \qquad \kappa = \kappa(\bar{P}), \tag{36.3}$$

with the middle expression read as $k\varepsilon$ if $\kappa = 1$.

Proof. Let μ_k be the law of X_k and $e_k = \mathrm{TV}(\nu_k, \bar{\nu}_k)$, so $e_0 = 0$. Then $\nu_{k+1} = \sum_x \mu_k(x) L(x, \cdot)$, and the triangle inequality gives

$$e_{k+1} \le \mathrm{TV}\Big(\sum_x \mu_k(x) L(x, \cdot), \; \sum_x \mu_k(x) \bar{P}(\pi(x), \cdot)\Big) + \mathrm{TV}(\nu_k \bar{P}, \bar{\nu}_k \bar{P}) \le \varepsilon + \kappa \, e_k, \tag{36.4}$$

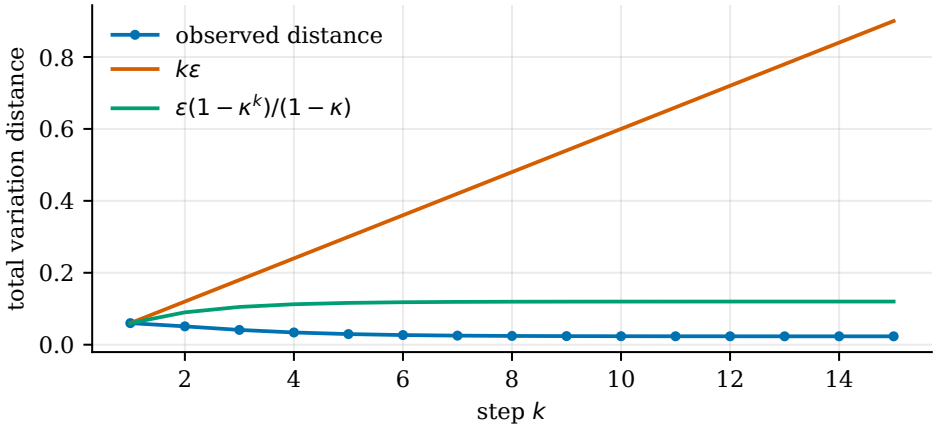


Figure 36.2: Total variation distance between the true block law and the compressed prediction for the perturbed four-state chain of Worked Example 36.2, with the bounds $k\varepsilon$ and $\varepsilon(1-\kappa^k)/(1-\kappa)$ of Extension 36.1. Computed by `examples.py`; an illustration of the stated bound.

since the middle term is an average of per-state defects and $\sum_x \mu_k(x)\bar{P}(\pi(x),\cdot) = \nu_k\bar{P}$, and since $\bar{P}$ contracts total variation by at most the Dobrushin coefficient. Induction from $e_0 = 0$ gives the claim. □

Worked Example 36.2 (Drift in a perturbed chain)**.** Perturb state 1 of Worked Example 36.1 to the row $(0.29, 0.35, 0.21, 0.15)$. Its lumped row becomes $(0.64, 0.36)$ and the defect against the reference rows of the unperturbed states is $\varepsilon = 0.060$. The Dobrushin coefficient of the reference kernel is $\kappa = 0.500$. Starting in the perturbed state, the observed total variation distance between the true block law and the compressed prediction is 0.0295 at $k = 5$, against the bound $k\varepsilon = 0.300$ and the sharper bound $\varepsilon(1-\kappa^k)/(1-\kappa) = 0.1163$, and it settles at 0.0233 at long horizons. The sharper bound is the informative one: because the compressed kernel contracts, errors do not accumulate linearly, and the asymptotic bound is $\varepsilon/(1-\kappa) = 0.12$. The figure shows both bounds against the observed distance.

The point of the pair of results is the separation of the roles. The bound controls how well the compressed state tracks the block process, which is closure. It does not control what the block process can say about the target, which is sufficiency and has to be certified by the separate criterion of Theorem 32.2, or by its noisy form in Extension 32.1. A system can have an excellent drift bound and a useless memory, as in Worked Example 36.1, or a sufficient memory with an uncontrolled drift, as in Counterexample 36.2.

Chapter 37

Memory as Preservation of Viable Continuation

A memory is useful when it keeps the system's future options open under its objectives. Stated precisely, a memory policy is a rule that chooses encoding, update and release actions, and its quality is measured by the trajectories it makes possible. Let $f : \mathcal{X} \times \mathcal{A} \to \mathcal{X}$ be a deterministic transition law (stochastic versions replace membership by probability), let $K \subseteq \mathcal{X}$ be a set of admissible states, and call a trajectory viable if it remains in K for all time.

Definition 37.1 (Viability kernel)**.** The viability kernel of K is the set $\mathrm{Viab}(K)$ of initial states from which some sequence of actions yields a trajectory remaining in K forever.

In the memory setting K encodes constraints such as sufficient remaining resources, bounded interference, and, centrally, a floor on historical fidelity. Define a fidelity functional $D_t \geq 0$ measuring the distortion of historically warranted reconstruction, and let $K = \{x : D(x) \leq \delta\} \cap K_B$ where K_B collects the resource constraints. The idealized objective of the proposal is

$$\max_{\pi} \ \mathbb{E}_{\pi}\Big[\sum_{t=0}^{\infty} \gamma^t \big(U_t - \lambda C_t - \mu D_t\big)\Big], \tag{37.1}$$

with U_t the continuation value, C_t the computational cost and D_t the distortion. The structure of (37.1) is a penalized formulation in which fidelity trades against utility at the exchange rate μ, and this is exactly what the monograph argues against making the only treatment of fidelity.

Proposition 37.2 (Penalties cannot guarantee fidelity)**.** *Suppose an action $a^{\dagger}$ yields per-step utility G above the alternative at the cost of per-step distortion $d > \delta$, with all other terms equal. Then the penalized objective* (37.1) *prefers $a^{\dagger}$ whenever $\mu < G/d$ (using it at every step), whereas the constrained formulation, which requires the trajectory to stay in K, never selects it. For any fixed μ there exist utility gains $G > \mu d$ at which the penalized optimizer violates the fidelity bound.*

Proof. Per step the penalized value of $a^{\dagger}$ relative to the alternative is $G - \mu d$, which is positive exactly when $\mu < G/d$, and discounting multiplies both sides by $\sum_t \gamma^t$. Under the constrained formulation $a^{\dagger}$ produces $D = d > \delta$ and leaves K, so it is infeasible. Given μ, take $G > \mu d$. □

The proposition is simple, and it shows that with a finite penalty weight the guarantee of fidelity depends on the scale of the utilities, which are not controlled. A hard constraint is a different mathematical object, and its natural solution concept is the viability kernel. For a finite state space the kernel can be computed.

Proposition 37.3 (Computation of the viability kernel)**.** *Let $\mathcal{X}$ be finite and $\mathcal{A}$ finite. Define $K_0 = K$ and $K_{j+1} = \{x \in K_j : \exists a \in \mathcal{A},\ f(x,a) \in K_j\}$. The sequence is decreasing, stabilizes after at most $|K|$ steps at a set K_∞, and $K_\infty = \mathrm{Viab}(K)$.*

Proof. Each $K_{j+1} \subseteq K_j$, and since the sets are finite the sequence stops decreasing after at most $|K|$ strict decreases; at stabilization every $x \in K_\infty$ has an action leading to K_∞, so

repeating the choice yields a viable trajectory and $K_\infty \subseteq \mathrm{Viab}(K)$. Conversely, a state in $\mathrm{Viab}(K)$ has a viable trajectory, and by induction on j every state of that trajectory lies in K_j for all j, so $\mathrm{Viab}(K) \subseteq K_\infty$. □

A memory policy that respects the fidelity constraint is one that acts only within $\mathrm{Viab}(K)$, and the operational use of the kernel is to mark actions that appear beneficial in the short run but leave the viable set. For example, a release action that discards a trace is admissible only if the post-release state remains in the kernel, which requires that the discarded distinctions are not needed by any future cue in the committed family, in the sense of Theorem 32.2. The constrained problem is then to maximize $\mathbb{E}\sum_t \gamma^t(U_t - \lambda C_t)$ over policies that stay in $\mathrm{Viab}(K)$, with fidelity a constraint and not a price. Historical fidelity remains separately measurable throughout, which is the property that allows the system to report that a useful reconstruction is historically unwarranted instead of silently trading one for the other.

Viability-preserving ecphoric equivalence

Theorem 32.2 asks which histories a memory must distinguish for the recollections required now, and Proposition 37.3 asks which states the system must be able to leave through an admissible action. The two questions meet in a single one. Two memory states may be indistinguishable to every present recollection and to the present constraint, and yet respond differently once the memory is updated. Whether they may be merged depends on the family of admissible updates, and the answer is a quotient that is in general finer than the sufficiency quotient and coarser than the identity. This section constructs it and states exactly what is new and what is classical.

Setting. Let $\Sigma = (\mathcal{X}, \mathcal{A}, f, K, T)$ be a deterministic system with finitely many memory states $\mathcal{X}$, a finite set $\mathcal{A}$ of admissible memory updates (consolidation steps, revisions, searches, releases), a transition law $f : \mathcal{X} \times \mathcal{A} \to \mathcal{X}$, an admissible set $K \subseteq \mathcal{X}$ and a present recollection map $T : \mathcal{X} \to \mathcal{Y}$. Write $o(x) = (T(x), \mathbf{1}_K(x))$ for the joint observation of the present recollection and the present admissibility, and $f_w(x)$ for the state reached by the action word $w \in \mathcal{A}^*$.

Definition 37.4 (Viability-preserving ecphoric equivalence)**.** Define $x \sim_0 y$ if $o(x) = o(y)$ and $x \sim_{n+1} y$ if $x \sim_n y$ and $f(x,a) \sim_n f(y,a)$ for every $a \in \mathcal{A}$. The relation $\sim_\infty = \bigcap_n \sim_n$ is the viability-preserving ecphoric equivalence.

Extension 37.1 (Viability-preserving ecphoric equivalence)**.** For finite $\mathcal{X}$ the following hold. *(i)* The chain $\sim_0 \supseteq \sim_1 \supseteq \cdots$ stabilizes after at most $|\mathcal{X}| - 1$ refinements, and $\sim_\infty$ is the coarsest equivalence that refines $\sim_0$ and is a congruence for every $f(\cdot, a)$. *(ii)* $x \sim_\infty y$ if and only if $o(f_w(x)) = o(f_w(y))$ for every action word w; that is, the two states agree on every future recollection and on every future admissibility along every admissible sequence of updates. *(iii)* The viability kernel $\mathrm{Viab}(K)$ of Definition 37.1, its robust counterpart $\{x : f_w(x) \in K \ \forall w\}$, and every set of the form $\{x : T(f_w(x)) \in S\}$ are unions of $\sim_\infty$ classes. *(iv)* An encoding $E : \mathcal{X} \to \mathcal{M}$ supports a decoder for $T \circ f_w$ and for $\mathbf{1}_K \circ f_w$ for every word w if and only if $\ker E \subseteq \sim_\infty$.

Proof. (i) The relations $\sim_n$ are equivalences and decrease with n. On a finite set a decreasing chain of equivalences that fails to strictly decrease at some step is constant afterwards,

because $\sim_{n+1}$ is a function of $\sim_n$ alone, and each strict decrease adds a class, so there are at most $|\mathcal{X}| - 1$ strict decreases. The limit refines $\sim_0$ and is a congruence by construction. If $\equiv$ is another congruence refining $\sim_0$, induction on n shows $\equiv \subseteq \sim_n$ for all n, so $\equiv \subseteq \sim_\infty$. (ii) By induction on the length of w, $x \sim_n y$ if and only if $o(f_w(x)) = o(f_w(y))$ for all words with $|w| \leq n$. (iii) If $x \sim_\infty y$ and an action word keeps the trajectory from x inside K, then by (ii) the same word keeps the trajectory from y inside K, since $\mathbf{1}_K$ is part of o; the robust kernel and the recollection sets are handled the same way. (iv) If $\ker E \subseteq \sim_\infty$ then for each w the targets $T \circ f_w$ and $\mathbf{1}_K \circ f_w$ are constant on fibers of E by (ii), so Theorem 32.2 gives decoders. Conversely, if $x \sim_\infty y$ fails there is a word w with $o(f_w(x)) \neq o(f_w(y))$, and a decoder for the corresponding target cannot exist on a fiber that contains both. □

Relation to established results. The ingredients are classical. For a deterministic system with input actions the coarsest congruence refining an observation is the Nerode equivalence of the Moore machine [10], and the same relation is strong bisimilarity of the labelled transition system with the observation as state label, which is computed by partition refinement [38]. Statement (iii) is the standard fact that bisimilar states have the same viability kernels. What the extension adds is the combination and the reading: the observation o is taken to be the pair of the present recollection and the present admissibility, so that the quotient is at once a sufficiency quotient in the sense of Corollary 32.4 and a quotient that respects the constraint of Definition 37.1; and the contrast it draws is between the present equivalence $\sim_0$, which is what Theorem 32.2 sees, and the future equivalence $\sim_\infty$, which is what the memory must preserve if it is to be updated. It does not extend to stochastic updates without replacing exact equality of observations by a probabilistic bisimulation, and it does not address learned or approximate updates, for which Problem 59.7 is the appropriate statement.

Worked Example 37.1 (Present recollection versus future action)**.** Take 6 memory states (b, c) with $b \in \{0, 1\}$ and $c \in \{1, 2, 3\}$. The flag b records whether a provisional, uncorroborated trace is held and c counts the steps since it was last attended. The updates are *idle*, which sets $c \leftarrow \min\{c + 1, 3\}$, and *seek*, which does nothing if $b = 0$, resolves the provisional trace to $(0, 1)$ if $(b, c) = (1, 1)$, and otherwise moves $(1, c)$ to $(1, c - 1)$. The admissible set is $K = \{b = 0\} \cup \{c \leq 2\}$, so a provisional trace may not be left unattended to the third count. The present recollection is the flag, $T(b, c) = b$.

The observation $o = (T, \mathbf{1}_K)$ separates the states into 3 classes at time zero, namely the three states with $b = 0$, the pair $\{(1, 1), (1, 2)\}$ and the state $(1, 3)$, whereas T alone separates only 2. One round of refinement (1 strict refinement is needed) produces the final partition with 4 classes,

$$\big\{(0, 1), (0, 2), (0, 3)\big\}, \quad \big\{(1, 1)\big\}, \quad \big\{(1, 2)\big\}, \quad \big\{(1, 3)\big\}. \tag{37.2}$$

The pair $(1, 1)$ and $(1, 2)$ is the instructive one. The two states agree on the present recollection and on admissibility, and both belong to the controlled viability kernel, which contains 5 states, all of K. They nevertheless differ under the update *idle*: from $(1, 1)$ the result $(1, 2)$ is admissible, while from $(1, 2)$ the result $(1, 3)$ is not, so the first tolerates one unattended step and the second none. A memory that merged them because present recollection could not tell them apart would be sufficient for T and would misjudge, after one update, whether it had left the admissible set. The robust kernel, the states from which

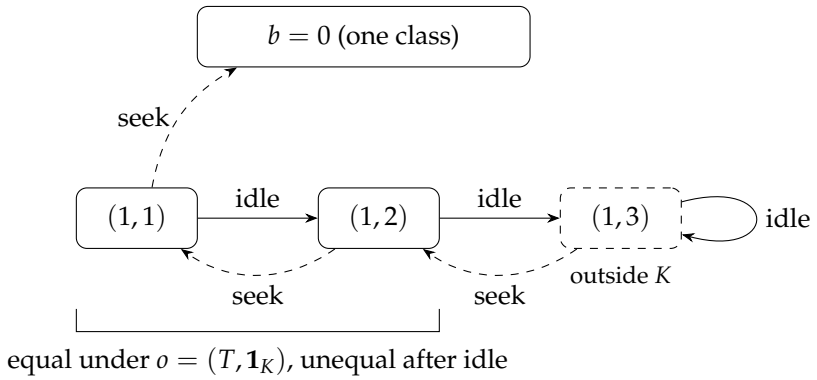


Figure 37.1: The quotient of Worked Example 37.1. The three states with $b = 0$ form one class, and the three states with $b = 1$ remain distinct. The states $(1,1)$ and $(1,2)$ coincide for the present recollection and the present constraint (bracket) and separate after one *idle* step. The dashed box marks the state outside K; self-loops of the class $b = 0$ are not drawn.

every update sequence remains in K, is the three states with $b = 0$, which form a single class, and both kernels are unions of classes as part (iii) requires. The states with $b = 0$ merge entirely, because the counter c is irrelevant to every future observation once the provisional trace has been resolved.

The extension also answers a design question that the viability formulation of Chapter 37 leaves open. A release action that discards the distinction between two memory states is admissible only if the two states are $\sim_\infty$-equivalent, which is stronger than the requirement that they be equivalent for the committed recollections, and weaker than the requirement that they be identical. In the ledger architecture of Chapter 41 the distinction would be retained as a record of the form that Proposition 32.9 allows to be repaired, and the quotient identifies how much of it can be dropped from the working state without loss.

Synthesis of Part VII

The examples of this part separate four questions that the informal phrase "the memory works" runs together, and each has its own certificate. The first is whether the stored state contains what the committed recollections require, which Theorem 32.2 answers exactly and Extension 32.1 answers with a risk floor when the encoding is noisy. The finite space of Worked Example 32.1 shows that an encoding can be better at recovering histories and worse at the recollection that matters, and part (iii) of the extension shows that any downstream channel can only raise the floor, so the dynamics of Chapters 33 and 34 act on a quantity that is already bounded by the encoding. The second question is whether the retrieval dynamics reach the stored state from the cue, which is a statement about the spectrum of the Jacobian at a fixed point (Worked Example 33.1) and about margins (Counterexample 34.1); both are conditions on the stored traces and the temperature and neither says anything about whether the traces suffice. The third question is whether a compressed state can be evolved without reference to the detailed state, which is closure, certified by lumpability and degraded gracefully by Extension 36.1; the two constructions of Chapter 36 show that closure and sufficiency are logically independent. The fourth is whether two states that agree on present recollection remain interchangeable under

Question	Certificate	Failure exhibited by the example
Does the state contain the recollection?	Theorem 32.2; Extension 32.1	Encoding E_1 of Example 32.1 has risk 0.500 for the location query although it recovers more histories than E_2
Does retrieval reach it?	Theorem 33.2; Proposition 16.1	Blend unstable above $\beta = 3$ for three orthonormal traces; vacuous bound with correct retrieval; drift to the blend for near-duplicates
Can the compressed state evolve alone?	Proposition 36.1; Extension 36.1	Closed projection with Bayes risk 0.500; sufficient projection that is not closed
Does equivalence survive updates?	Extension 37.1	Present-equivalent states $(1, 1)$ and $(1, 2)$ differ after one idle step

Table 37.1: The four questions of Part VII, the result that certifies each, and the example that shows the certificate failing independently of the others.

the updates the system may apply, which is the congruence of Extension 37.1 and which strengthens sufficiency by the quantifier over action words.

Three consequences follow for the design of the system in Part VIII. First, a diagnostic of poor recollection should be run in order: the encoding is tested against the criterion with the dynamics held fixed, then the dynamics against the margin and spectral conditions with the encoding held fixed, and only then is an architectural change warranted. Second, compression of the working state, whether by lumping, projection or release, should be licensed by the quotient of Extension 37.1 and not by present recollection alone. Third, every quantitative statement in this part is conditional on the model in which it was derived: the three-trace and near-duplicate families are chosen because their spectra are available in closed form, and the extensions are proposed results proved here and not findings about a measured system. Where a statement was checked numerically, the checking script is listed in Appendix M.

Part VIII

The Mem8 Computational Architecture

Chapter 38

Representational Substrate

The architecture is specified by its data model first, because every later guarantee refers to what a trace is and which parts of it are stored. A trace is the tuple

$$m_i = (\psi_i, \tau_i, \kappa_i, \rho_i, \sigma_i, \nu_i), \tag{38.1}$$

where $\psi_i \in \mathbb{C}^n$ is representational content, τ_i is temporal information (an absolute stamp, a lag set, or a phase code in the sense of Chapter 14), κ_i is the context under which the trace was written, ρ_i is relational structure linking it to other traces, σ_i is confidence or uncertainty metadata, and ν_i is provenance, a pointer into the ledger of Chapter 41. The tuple is a specification of roles, not of formats: each component has a type, a persistence class, and an observation interface, and a conforming implementation may realize any component by any representation that meets the interface.

Persistence is the distinction that matters most for correctness, and the architecture assigns every quantity to exactly one of three classes. Persistent quantities survive between sessions and are the only ones that count as memory state in the sense of Theorem 32.2: the content ψ_i, the stamp τ_i, the writing context κ_i, the relations ρ_i, the confidence σ_i, and the pointer ν_i. Reconstructed quantities are computed on demand from persistent ones and a cue, and are discarded after use, such as the reconstruction state $r^{(k)}$, the retrieval weights $w_i^{(k)}$, and the mode coefficients of the field. Dynamically generated quantities are produced by the oscillatory engine and have no persistent counterpart, such as the phases $\theta_i^{(k)}$ and the instantaneous synchrony statistics. A design in which a reconstructed quantity is silently cached and then treated as persistent has changed the memory state without recording the change, and the ledger discipline of Chapter 41 exists to prevent this.

Table 38.1 records the assignment. Its purpose is also to make the sufficiency question answerable: since the decoder sees only the persistent classes and the cue, the equivalence $\approx_T$ of Definition 32.3 must be assessed against the persistent tuple, not against any reconstructed quantity.

Quantity	Symbol	Class	Source
Content	ψ_i	persistent	written at encoding
Temporal information	τ_i	persistent	written at encoding
Writing context	κ_i	persistent	copy of Z at write time
Relations	ρ_i	persistent	updated by consolidation
Confidence	σ_i	persistent	updated by evidence
Provenance pointer	ν_i	persistent	ledger
Reconstruction state	$r^{(k)}$	reconstructed	retrieval engine
Retrieval weights	$w^{(k)}$	reconstructed	retrieval engine
Oscillator phases	$\theta^{(k)}$	dynamically generated	oscillatory engine

Table 38.1: Persistence classes of the quantities in the Mem8 data model.

Chapter 39

The Oscillatory Coordination Engine

The oscillatory engine supplies the phase vector $\theta^{(k)}$ that enters the context operator $U(z_t, \theta^{(k)})$ of the minimal model. Its specification has five parts: mode generation, coupling, synchronization, phase reset, and event-driven activation. Modes are generated either from a fixed set of frequencies ω_i or from the eigenmodes of a graph Laplacian over the trace index, as in Chapter 11. Coupling is given by a matrix $K = (K_{ij})$, symmetric when the dynamics is to be a gradient system, so that the update

$$\theta_i^{(k+1)} = \theta_i^{(k)} + \Delta t \Big[\omega_i + \sum_j K_{ij} \sin\big(\theta_j^{(k)} - \theta_i^{(k)}\big) \Big] \tag{39.1}$$

has the potential $V = -\frac{1}{2} \sum_{ij} K_{ij} \cos(\theta_j - \theta_i)$ as a Lyapunov function in the absence of frequency detuning, and decreases it at every step when $\Delta t < 1/\max_i \sum_j |K_{ij}|$ (Chapter 12). Synchronization is monitored through the order parameter $R_e = |N^{-1} \sum_j e^{\mathrm{i}\theta_j}|$, and the specification must forbid the regime of global synchrony ($R_e \to 1$ for all pairs) in which selectivity collapses, since every trace then receives the same context phase and the contextual operator no longer discriminates. Phase reset sets $\theta^{(0)}$ at an event, to a value determined by the event type, so that retrieval after the same event begins from the same phases. Event-driven activation means that the engine integrates only when a cue or an encoding is present, and is otherwise quiescent; this keeps Θ_t transient in the sense of Table 38.1 and removes spurious dependence of retrieval on wall-clock time.

The engine must be replaceable by a nonoscillatory implementation with the same interface, because the empirical question of the monograph is whether coordinated oscillation contributes anything beyond what a general recurrent dynamics provides. The interface is a map $\mathfrak{C} : (z, \text{event}, k) \mapsto \Phi^{(k)}$ that returns, for each trace, a phase factor $e^{\mathrm{i}\Phi_i^{(k)}}$ applied to the cue. Three drop-in replacements are specified. The frozen engine returns $\Phi_i^{(k)} = \theta_i^{(0)}$ for all k, which keeps the phase code and removes the dynamics. The randomized engine returns phases drawn from the same marginal distribution as the oscillatory engine but independently of the state, which preserves the marginal statistics and removes the structure. The nonoscillatory recurrent engine returns $\Phi^{(k)} = g_\vartheta(\Phi^{(k-1)}, z)$ for a learned map g_ϑ with the same parameter count as the Kuramoto couplings and no sinusoidal structure. A claim that oscillation matters is a claim that the oscillatory engine outperforms all three at matched parameters, budget, and information content; the protocol is in Chapter 51.

Figure 39.1 shows the monotone decrease of the potential under the step bound of Appendix D, and the growth of synchrony that results. This is the regime the specification must guard against when selectivity is needed.

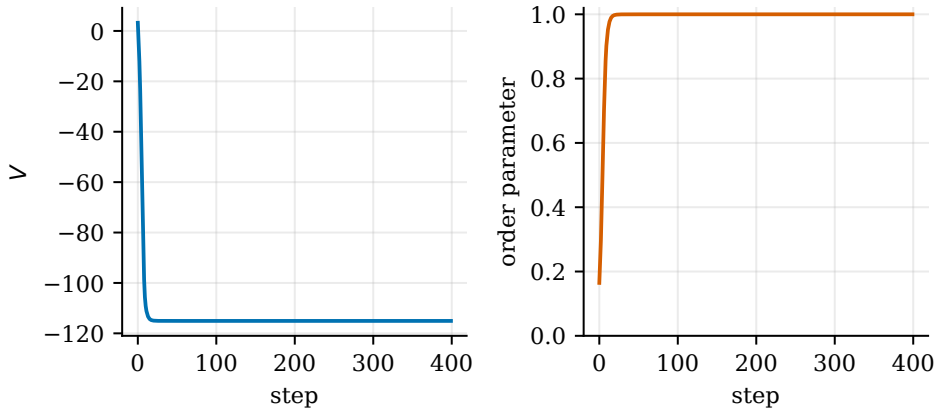


Figure 39.1: Euler-discretized Kuramoto dynamics with $\omega_i = 0$ on a random sparse symmetric coupling of 30 oscillators at step $0.9/\kappa$: the potential V (decreasing at every step, as asserted) and the order parameter. Computed by the script `figs.py` with seed 8; an illustration of the stated analytic result and not an evaluation of the architecture.

Chapter 40

The Ecphoric Retrieval Engine

Retrieval is an iterative computation, and the specification must fix its initialization, its candidate activation, its refinement step, its convergence criterion, its confidence estimate, and its termination rule. The iteration is

$$r^{(k+1)} = \mathcal{T}_\vartheta\big(r^{(k)}, c, M, Z, \Theta^{(k)}\big), \tag{40.1}$$

with the minimal instance $\mathcal{T}_\vartheta(r) = (1-\eta)r + \eta \sum_i w_i m_i$ of Appendix L. The initial state is $r^{(0)} = c$, the cue projected into the content space, or $r^{(0)} = 0$ if the cue is a pure context. Candidate activation computes the scores $s_i^{(k)} = R(m_i, U(z, \theta^{(k)})c)$ for a subset of traces selected by an index, exact in the baseline and approximate when Chapter 43 makes a budget binding. The refinement step is (40.1), and convergence is declared when the residual $\|r^{(k+1)} - r^{(k)}\|$ falls below a tolerance ϵ_{res}.

For a map known to be a λ-contraction, convergence has a certificate. If the residual at step k is $\varrho_k = \|r^{(k)} - r^{(k-1)}\|$, then Lemma 33.1 applied at the iterate gives the distance to the fixed point as at most $\lambda\varrho_k/(1-\lambda)$. The termination rule can therefore be stated in terms of a certified distance and not only in terms of a residual: stop when $\lambda\varrho_k/(1-\lambda) \le \epsilon$. For the damped softmax step with $\eta \in (0, 1]$ and $\beta < 1$, Theorem 33.2 gives the contraction factor $\lambda = 1 - \eta + \eta\beta$. The certificate is valid only for the fixed point it certifies, and by Example 33.4 it says nothing about whether that fixed point is the historically correct reconstruction.

Confidence estimation is therefore reported on two axes that must not be merged. The first is convergence confidence, the certified distance above. The second is selectivity, the information in the retrieval weights, which can be summarized by the entropy $H(w)$ or by the margin $\Delta = s_{(1)} - s_{(2)}$ between the largest and second largest scores. A sharp margin supports the claim that the output is close to one stored trace, by Proposition 16.1. Neither axis measures fidelity, which requires the evidence-based statistics of Chapter 28, and a conforming implementation must report the first two as properties of the computation and the third only when supporting evidence is available.

Executable specification of retrieval

The retrieval engine of this chapter is specified by an iteration, and the lifecycle example at the end of Part VIII uses its single-step form, so the form that is used is fixed here in pseudocode. The listing is a specification in the sense of Chapter 45: it defines the interface and the invariants, it is deterministic given its inputs, and it is not a claim about performance.

Three properties of the listing are used below. Retrieval is a read-only operation on the store, so the persistent state after a retrieval differs from the state before it only by one ledger record. The warrant coverage ω of line 9 is computed from the stored traces and the attributed set A and not from the reconstruction, so it can be low when the reconstruction is dominated by the cue; this is the separation between generative influence and warrant of

```
RETRIEVE(M, L, c; beta, g, theta_d, tau)
  # M: traces m_1..m_P with ledger pointers nu_1..nu_P;  L: ledger;  c: cue
  1  for i in 1..P:  cos_i <- <m_i, c> / (|m_i| |c|)
  2  w <- softmax(beta * cos)
  3  F <- sum_i w_i m_i
  4  r <- (1 - g) F + g c                       # one cue-anchored step
  5  margin <- (largest cos) - (second largest cos)
  6  A <- { i : w_i >= tau }                    # attributed traces
  7  D <- { j : r_j >= theta_d }                # reportable details
  8  W <- { j in D : exists i in A with (m_i)_j >= 1/2 }   # warranted details
  9  omega <- |W| / |D|    (omega <- 1 if D is empty)
 10  append to L: interpretation(cue = c, cites = nu_i for i in A,
                                 weights = w, details = D)
 11  return (r, w, A, D, W, omega, margin)      # M is not modified
```

Figure 40.1: Single-step retrieval with attribution and warrant coverage. The traces are never modified by retrieval, and the only persistent effect is the appended interpretation record. The parameters used in the lifecycle example are $\beta = 8$, $g = 0.2$, $\theta_d = \frac{1}{4}$ and $\tau = \frac{1}{2}$.

Definition 41.1. And the set A may be empty, in which case W is empty and $\omega = 0$ for a nonempty D, which is how the specification reports that no single stored trace accounts for the answer.

Chapter 41

Persistent Memory and Historical Provenance

The architecture must permit modification of a memory without silently rewriting the record of what was previously stored. The mechanism is an append-only ledger. A ledger is a set of typed records, each carrying an identifier, a type, a payload, a reference to the records it depends on, and a logical timestamp. Four record types are distinguished. An inscription records the original encoding of an observation: the content as first written, with the context and time. A transformation records a subsequent change to a trace, with the exact increment of Proposition 28.2. An interpretation records a reconstruction produced by the retrieval engine together with the cue and context that produced it, so that reconstructions are never confused with inscriptions. An evidence record records an observation that supports or contradicts a prior record and is the stated reason for any transformation that cites it. The pointer ν_i of a trace is the identifier of its latest record, and the full chain of records reachable from ν_i is the provenance of the trace.

A distinction that the ledger must support is between the evidence that generated a reconstruction and the evidence that warrants its historical attribution. A reconstruction can be causally influenced by a trace without that trace justifying every detail of the result: the weight a trace receives in the retrieval iteration measures its contribution to the computation, and says nothing about whether the inscription behind it entails a particular detail of the output, which may have been supplied by the context operator, by the cue, or by the blending of several traces. The two relations are recorded separately.

Definition 41.1 (Generative influence and warrant)**.** Let an interpretation record carry a reconstruction $r^\star$ with a set A of separately reportable details. Its generative influence is the weight vector $w^{(K)}$ of the final refinement step, together with the sources of the cue and the context used. A detail $a \in A$ is warranted if there is an attributed trace, that is, a trace whose weight in $w^{(K)}$ is at least a declared attribution threshold τ_{attr} (by default $\frac{1}{2}$), such that the inscription record at the head of its citation chain, together with the logged transformations on that chain, determines the value of a to within the stated tolerance under a declared entailment rule. The warrant coverage of the record is $\omega = |\{a \in A : a \text{ is warranted}\}|/|A|$, and the details outside the warranted set are labeled inferred.

Warrant coverage is computed from the ledger and the entailment rule, independently of the retrieval weights, so a reconstruction with high generative influence from a trace and low warrant coverage is detected as one whose details were largely supplied by something other than the record. Source attribution in the sense of Chapter 19 is a claim about warrant and not about influence, and the experiments of Chapter 49 measure it as such.

Three properties follow from the append-only discipline. The first is auditability: the chain from any trace to its inscription is available, and every transformation can be traced to its evidence. The second is reversibility, shown for logged increments in Proposition 28.2: any earlier state of a trace is recoverable from the current state and the ledger. The third is non-forgery: since interpretations are typed distinctly from inscriptions, a reconstruction cannot be mistaken for the original even if it is later promoted to a stored trace, because the promotion is itself a recorded transformation citing the interpretation as its source. This

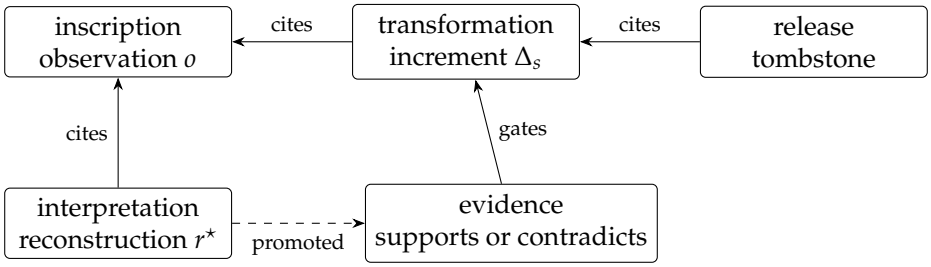


Figure 41.1: Typed records of the ledger and their citation structure. Arrows point from a record to the records it depends on; no record is removed, and a reconstruction can enter the store only through a transformation that cites it.

```
REPLAY(L, init)
  # L: ledger records with Lamport timestamps;  init: payload -> initial vector
  1  M <- []                                  # list of traces, indexed by inscription
  2  for rec in L sorted by (counter, replica_id):
  3     if rec.type == inscription:
  4         append copy(init[rec.payload]) to M;  index[rec.id] <- len(M)
  5     if rec.type == transformation:
  6         i <- index of the trace cited first by rec
  7         M[i] <- M[i] + rec.payload          # logged increment, exact
  8     if rec.type == release:
  9         mark the trace cited by rec as released (payload kept in cold storage)
 10     # interpretation and evidence records do not change M
 11  return M
```

Figure 41.2: Deterministic derivation of the store from the ledger. The result depends only on the set of records, by Proposition 44.2.

last property is what separates a system that has revised its memory from one that has lost track of its own history.

Deliberate forgetting is accommodated without breaking append-only semantics by tombstones. A tombstone is a record of type release that references the records it supersedes. Under the default policy of Chapter 29 the payloads of the superseded records are retained in cold storage so that the release is reversible; under an erasure policy the payloads are deleted and the tombstone records the fact and the time of deletion. In either case the ledger records that the release occurred, and what is lost under erasure is the content, never the knowledge that the content once existed. The relation between the ledger and the sufficiency criterion is direct: by the first qualification after Proposition 32.6, a decoder that consults the ledger has the ledger as part of the memory state, and the sufficiency question is then asked of the combined encoding. This is the formal reason that source discrimination, which is impossible from content alone (Proposition 19.1), becomes possible when provenance is stored.

Executable specification of ledger replay

The derived state of Proposition 44.2 is computed by a fold, and the fold is the object that the replay requirement of the reference implementation tests.

A remark on arithmetic is needed, because Proposition 28.2 is stated in real arithmetic and an implementation uses finite precision. Recovering an earlier state by subtracting later increments from the current state is exact in real arithmetic, but in floating point it can leave a rounding residue, since addition is not associative. The specification therefore recovers an earlier state by replaying the ledger up to the chosen time, and it uses subtraction only as an audit check. In the lifecycle example below all stored values and increments are dyadic rationals of small denominator, for which floating-point addition and subtraction are exact, so that both procedures agree bit for bit, and this is the reason those values were chosen. The choice is a device of the example; it does not make the property hold for arbitrary payloads.

Definition 41.1 counts a detail as warranted only through an attributed trace, and the lifecycle example shows why the threshold is needed. Under a softmax every trace has strictly positive weight, so if warrant required only positive generative influence then every detail present in any stored trace would be warranted and the coverage ω would be vacuous. The definition therefore requires the weight to reach τ_{attr}, a declared policy parameter, and takes the content of the trace to be its inscription together with its logged transformations, which is the form used by the listing of Figure 40.1.

Chapter 42

Adaptive Memory Organization

Five operations organize the store over time: consolidation, splitting, merging, abstraction, and forgetting. For each the specification must state an invariant for provenance, for accessibility, and for recoverability, and the invariants are the acceptance conditions for any implementation.

Consolidation transfers strength from the fast to the slow store by replay, with the dynamics of Proposition 27.1; its provenance invariant is that the slow-store trace carries the pointers of the fast-store traces it was built from, and its accessibility invariant is that the cue family supported before consolidation remains supported with at most the distortion allowed by the replay's learning rate. Splitting replaces a trace by two when a cue family is found to require a distinction that the trace blurs, such as when the same content has been associated with conflicting outcomes in different contexts; the provenance invariant is that both children cite the parent, and the recoverability invariant is that the parent is reconstructible from the pair, which holds exactly when the split is an additive decomposition $\psi = \psi' + \psi''$ and the ledger records the weights. Merging replaces a set of traces by one when they are redundant with respect to the committed cue family. Its admissibility is a corollary of Proposition 30.1: merging two traces that differ by δ in norm replaces each by their mean, a displacement of $\delta/2$ per trace, so the readout changes by at most $(1 + 2\beta)\delta/2$ at every query in the unit ball, and the merge is admissible when this is below the tolerance ε of the compression bound (30.1). Abstraction creates a trace representing a class by the mean of its members, retaining the members in the ledger; it is the operation by which the slow store acquires semantic structure, and its admissibility is the closure criterion of Proposition 36.1 for the projection that sends members to the class. Forgetting follows Chapter 29 and always leaves a tombstone.

Proposition 42.1 (Provenance is preserved by every organizing operation)**.** *Suppose each of the five operations is implemented as an append of new records that reference the records they act on, and never as a modification or deletion of an existing record, except that erasure policy may delete payloads of released records. Then for every trace present at any time, the chain from its pointer to the inscriptions of the observations it derives from is complete, and the set of records only grows.*

Proof. Each operation appends records that cite their inputs, so by induction on the number of operations every trace's chain reaches inscriptions through citations that are never removed. The set of records grows monotonically because no operation removes a record, and erasure removes only payloads, which does not alter the citation structure. □

Executable specification of controlled reconsolidation and a merge check

The evidence-gated update of Chapter 28 and the provenance invariant of Proposition 42.1 fix what a revision must do, and the listing below fixes how.

By construction every operation in the listing appends and none removes, so the hypothesis of Proposition 42.1 holds, and by Proposition 44.2 the store after any sequence of

```
RECONSOLIDATE(M, L, i, e, reports; eta, Lambda_star)
  # i: trace to revise;  e: evidence-bearing target;  reports: {(source, rho)}
  1  keep one report per distinct source          # no double counting
  2  Lambda <- product over kept reports of rho / (1 - rho)
  3  ev <- append to L: evidence(cites = nu_i, payload = reports)
  4  if Lambda < Lambda_star:  return M           # gate closed, store unchanged
  5  Delta <- eta * (e - m_i)                      # exact increment
  6  tr <- append to L: transformation(cites = nu_i and ev, payload = Delta)
  7  m_i <- m_i + Delta;  nu_i <- tr
  8  return M

REVERSE(M, L, i, tr)
  1  Delta <- payload of tr
  2  tr' <- append to L: transformation(cites = nu_i and tr, payload = -Delta)
  3  m_i <- m_i - Delta;  nu_i <- tr'              # nothing is deleted
```

Figure 42.1: Evidence-gated revision and its reversal. The evidence record is appended whether or not the gate opens, so that the rejected evidence remains auditable. Reversal appends an inverse increment and does not remove records.

operations equals the replay of its ledger. The gate of line 4 aggregates by source, which is the discipline that Counterexample 28.1 shows to be necessary.

Worked Example 42.1 (Admissibility of a merge)**.** Merging two traces at distance d into their midpoint moves each by $d/2$, so by Proposition 30.1 the readout of any cue changes by at most $(1+2\beta)d/2$. A merge is therefore admissible for a compression budget ε if $d \le 2\varepsilon/(1+2\beta)$. For $\beta = 8$ and $\varepsilon = 0.05$ the threshold is $d \le 0.0059$, a very small distance, which shows that at selective temperatures only near-identical traces can be merged on the strength of this sufficient condition alone. In 300 random stores with a planted near-duplicate pair, the largest observed ratio of the actual readout change to the bound $(1+2\beta)d/2$ was 0.027. The bound was therefore never violated and was loose by a factor of about thirty-seven on those stores. Because Extension 30.1 shows that the dependence on β cannot be improved in general, a larger threshold must come from structural knowledge of the cue family, which is what the quotient of Extension 37.1 provides. These are algebraic checks of the proposition with a fixed seed and not measurements on a memory.

A unified learning objective. The operations above are mechanisms, and a trainable architecture needs a single objective that couples them. Let ϑ collect the continuous parameters of the representation (the encoder of content and the context operator U), of the oscillatory engine (ω, K), and of the retrieval engine (β, η, any gating maps), and let the discrete operations (merging, splitting, release) be governed by rules whose thresholds are hyperparameters. Over streams of episodes and cues with targets T, the learning problem is

$$\min_{\vartheta} \; \mathbb{E}\Big[\sum_{t} \ell\big(r_t^{\star}(\vartheta), T(h_t, c_t)\big) + \lambda\, C_t(\vartheta)\Big] \quad \text{subject to} \quad \mathbb{E}[D_t] \le \delta \text{ and } X_t \in \mathrm{Viab}(K) \text{ for all } t, \tag{42.1}$$

where ℓ is the reconstruction loss, C_t the computational cost, D_t the fidelity distortion of Chapter 37, and the second constraint confines the discrete operations to the viability kernel. The constraints are handled by a Lagrangian with a multiplier on the fidelity term

updated by dual ascent, which converges to a solution that respects the constraint and, unlike a fixed penalty, does not depend on the scale of the loss (Proposition 37.2). The retrieval engine is differentiable in ϑ whenever it is a contraction, which makes the objective trainable without backpropagating through the unrolled iteration.

Proposition 42.2 (Implicit differentiation through a contractive retrieval)**.** *Let $F(r;\vartheta)$ be continuously differentiable jointly in (r,ϑ) with $\|\partial_r F\| \le \lambda < 1$ and $\|\partial_\vartheta F\| \le L_\vartheta$. Then the fixed point $r^\star(\vartheta)$ is continuously differentiable and*

$$\frac{\partial r^\star}{\partial \vartheta} = \big(I - \partial_r F\big)^{-1} \partial_\vartheta F\Big|_{r^\star}, \qquad \Big\|\frac{\partial r^\star}{\partial \vartheta}\Big\| \le \frac{L_\vartheta}{1-\lambda}. \tag{42.2}$$

Proof. Since $\|\partial_r F\| < 1$, the matrix $I - \partial_r F$ is invertible with $\|(I-\partial_r F)^{-1}\| \le 1/(1-\lambda)$ by the Neumann series. The implicit function theorem applied to $r - F(r;\vartheta) = 0$ gives differentiability and the formula, and the bound follows. □

The learning problem (42.1) makes the relation among consolidation, reconsolidation and the representation explicit: they are components of one constrained optimization, in which consolidation shapes the targets that the loss measures, evidence-gated reconsolidation changes the targets T only through recorded evidence, and the representation and oscillator parameters are learned against the resulting stream. The specification does not claim that the optimization has been carried out; it states what the training problem is, so that the experiments of Part IX have a defined object to train.

Chapter 43

Resource Scheduling and Bounded Computation

The ideal system of Part VII has no resource limits, and the practical system has a budget B_t on time, memory and energy. The scheduler decides, for each retrieval, how many refinement steps to spend, how many traces to activate, and whether to attempt reconstruction at all. The central quantity is the marginal value of a further step, which for a contractive refinement can be stated exactly.

Suppose the iteration is a λ-contraction with initial certified error R and that each step costs $\varsigma > 0$ in the units of the objective, with the penalty for remaining error equal to the error itself. The total cost after k steps is bounded by

$$J(k) = R\lambda^k + \varsigma k, \tag{43.1}$$

where the first term is the error bound of Lemma 33.1.

Proposition 43.1 (Optimal stopping for contractive refinement)**.** *The function J is convex in k and its continuous minimizer is*

$$k^\star = \frac{1}{\ln(1/\lambda)} \ln\!\Big(\frac{R \ln(1/\lambda)}{\varsigma}\Big) \tag{43.2}$$

when $R\ln(1/\lambda) > \varsigma$, and J is minimized at $k = 0$ otherwise. The integer minimizer is $\lfloor k^\star \rfloor$ or $\lceil k^\star \rceil$.

Proof. $J''(k) = R(\ln\lambda)^2\lambda^k > 0$, so J is convex. Setting $J'(k) = -R\ln(1/\lambda)\lambda^k + \varsigma = 0$ gives $\lambda^k = \varsigma/(R\ln(1/\lambda))$, hence $k^\star$. If $R\ln(1/\lambda) \le \varsigma$ then $J'(0) \ge 0$ and convexity gives the minimum at zero. For a convex function on the integers the minimizer is one of the two integers adjacent to the continuous minimizer. □

The proposition formalizes the rule that further reconstruction is worth its cost until the marginal reduction of the error bound falls to the marginal cost of a step, and it shows that slow contractions (with λ near one) justify a number of steps that grows like $1/\ln(1/\lambda)$. It also identifies the failure mode of a fixed iteration count: an easy cue, with a small R, wastes steps, and a hard cue is stopped before the bound is useful.

Graceful degradation under a binding budget has a specification. The retrieval engine reduces cost in a fixed order, first by shortening the refinement, then by restricting activation to the top-K traces by a cheap prefilter, then by lowering the precision of the scores, and only then by declining to answer. At every stage the response reports the residual certificate and the margin of Chapter 40, so that the downstream consumer can tell a reduced-effort answer from a full-effort one. Critical historical distinctions are protected by marking traces whose release or merge would leave the viability kernel of Chapter 37; the scheduler never prefilters them out and never merges them, whatever the budget, which makes this a hard constraint and not a penalty.

Figure 43.1 shows the trade-off of Proposition 43.1: the error bound falls geometrically while cost grows linearly, and the optimum sits where their slopes cancel.

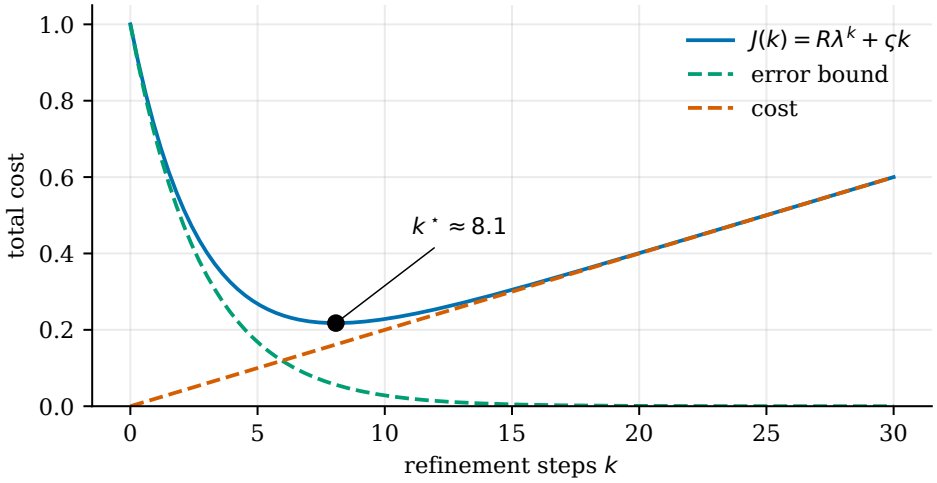


Figure 43.1: Total cost $J(k) = R\lambda^k + \varsigma k$ with $R = 1$, $\lambda = 0.7$, $\varsigma = 0.02$, and its minimizer from Proposition 43.1. Computed by the script `figs.py` with seed 8; an illustration of the stated analytic result and not an evaluation of the architecture.

Worked Example 43.1 (Easy and hard cues under a fixed iteration count)**.** Take $\lambda = 0.7$ and a step cost $\varsigma = 0.02$. For the certified initial error $R = 1$ the optimal stopping index is $k^* = 8.08$, the integer optimum is $k = 8$ with cost $J(8) = 0.2176$, and doing nothing costs $J(0) = 1.00$ while a fixed count of twenty steps costs $J(20) = 0.4008$, nearly twice the optimum. For an easy cue with $R = 0.05$ one has $R \ln(1/\lambda) = 0.018 < \varsigma$, so Proposition 43.1 says to take no step, at cost 0.05, and a fixed count of eight would cost 0.163, which is 3.26 times as much. For a hard cue with $R = 5$ the optimum is $k^* = 12.6$, the integer minimizer is 13 with cost 0.308, and stopping at the fixed count of eight costs 0.448, a factor 1.45 above the optimum. The same fixed count is thus too long for one cue and too short for the other, which is the failure mode that the certified residual is meant to remove. These costs are computed from the formula $J(k) = R\lambda^k + \varsigma k$ and do not describe a measured system.

Chapter 44

Distributed and Asynchronous Memory

Several nodes may hold replicas of a memory and update them concurrently, with delayed observations and temporarily inconsistent local histories. The architecture requires eventual reconciliation without loss of any node's inscriptions. The ledger of Chapter 41 provides the natural structure, since a set of records with set union as the merge operation has the algebra needed.

Proposition 44.1 (Convergence of ledger replicas)**.** *Let each replica hold a set of records, grow it by adding records, and merge by set union. Then merge is commutative, associative and idempotent, so any two replicas that have received the same set of records, in any order and with any duplication, hold the same set. Equivalently, the replica states form a join-semilattice under inclusion, and every replica's state is monotone under updates and merges.*

Proof. Union is commutative, associative and idempotent. Two replicas that have received the same collection of records, however ordered and duplicated, hold the union of that collection, which is a single set. □

Agreement on the set of records is not yet agreement on the derived memory state, because transformations may not commute: applying an increment Δ_1 followed by a gated update is different from the reverse. The derived state is therefore defined as a fold of the update function over the records in a canonical total order. Give every record a Lamport timestamp [31] (ℓ, id) consisting of a logical counter and the identifier of the originating replica; the lexicographic order on pairs is total, is consistent with causality, and is the same at every replica.

Proposition 44.2 (Deterministic derived state)**.** *If the update function is deterministic and records are folded in the order of their timestamps (ℓ, id), then any two replicas holding the same set of records compute the same memory state. Late-arriving records with earlier timestamps are accommodated by replaying from the first affected position.*

Proof. The timestamp order is a total order on the common set, so the fold is the same sequence of applications at both replicas, and a deterministic update function yields the same result. A late record changes the sequence only from its own position onward, so replaying from that position reproduces the fold. □

The proposition is the formal basis for the replay requirement of the reference implementation. Its cost is the replay length, and the obvious engineering response is periodic snapshots, which are caches of derived state and not memory state, and are discarded when a late record invalidates them. Where updates commute, as with additive increments, the order is immaterial and replay can be avoided entirely.

The distinction between shared semantic knowledge and independently acquired episodic evidence has a precise form in this setting. Semantic content, the slow-store abstractions of Chapter 42, is replicated and reconciled by consolidating the union of the nodes' episodic records, and it is an output of the fold. Episodic evidence is never

overwritten by another node's semantic abstraction; it is a record in the ledger with the identifier of the node that observed it. A node that receives a conflicting abstraction can compare it with its own episodic records, which is the mechanism by which inconsistent local histories are detected and not merely overwritten.

Worked Example 44.1 (Non-commuting updates at two replicas)**.** Let the memory state be a single number s with initial value 1. Replica A produces the record $a = (\ell{=}1, A, \text{scale by } 2)$ and replica B the record $b = (\ell{=}1, B, \text{add } 3)$ concurrently. If a replica applies records in the order of arrival, the node that receives a first computes $1 \cdot 2 + 3 = 5$ and the node that receives b first computes $(1 + 3) \cdot 2 = 8$, so two replicas that hold the same set of records disagree on the state, although Proposition 44.1 guarantees that they agree on the set. Folding in the Lamport order (ℓ, id) places a before b at both replicas and gives 5 at each. Suppose now a record $c = (\ell{=}0, C, \text{add } 10)$ from a third replica arrives late. The canonical order is c, a, b, and replaying from the first affected position gives $((1 + 10) \cdot 2) + 3 = 25$ at every replica, whatever order the three records were received in, which is the statement of Proposition 44.2. Where updates commute, as for additive increments, the arrival order would have sufficed and replay could be skipped, which is the cost saving noted after the proposition.

Chapter 45

Reference Implementation

The reference implementation is organized as six modules whose interfaces are fixed so that each can be replaced independently: the representation (encoding and decoding of the trace tuple), the resonance scorer R, the oscillator $\mathfrak{C}$ of Chapter 39, the ecphoric reconstruction engine of Chapter 40, the provenance store of Chapter 41, and the evaluation harness. Replaceability is not a convenience; it is what makes the ablation experiments of Part IX meaningful, because an ablation changes exactly one module and holds the others fixed.

Three engineering requirements are binding. The first is determinism: with fixed seeds and a fixed order of events the implementation must produce bit-identical memory states, so that Proposition 44.2 is realized in code and any discrepancy is a defect. The second is instrumentation: every retrieval emits the cue, the context, the scores, the weights, the iterates, the residual certificates and the final reconstruction to an interpretation record, so that nothing about a reconstruction is discoverable only by rerunning it. The third is replay: any memory transition can be reproduced from the ledger and the seed, and the test suite includes a replay check that rebuilds the state from the ledger and compares it with the live state.

The minimal model of Appendix L is the first conforming instance. Its deliberately limited scope should be stated: it has no learning rules, no temporal sequence representation, no uncertainty beyond the retrieval weights, and no persistent-state updates, and the modules above are specified so that each of these can be added without changing the others. The pilot calculations of Appendix M check the algebraic claims of Parts II and III in this setting and do not constitute an evaluation of the architecture.

A complete miniature memory lifecycle

The components of Part VIII are exercised together by a deterministic walk through one memory from encoding to revision and audit. The setting is deliberately tiny so that every number can be inspected: seven binary features (*Ann*, *Bo*, *cafe*, *library*, *rain*, *morning*, *evening*), three episodes $E_1 = (\mathit{Ann}, \mathit{cafe}, \mathit{morning})$, $E_2 = (\mathit{Ann}, \mathit{library}, \mathit{rain})$ and $E_3 = (\mathit{Bo}, \mathit{cafe}, \mathit{rain})$, and the parameters $\beta = 8$, $g = 0.2$, $\theta_d = \frac{1}{4}$, $\tau_{\text{attr}} = \frac{1}{2}$ and $\Lambda^* = 10$. The procedures are those of Figures 40.1, 41.2 and 42.1. The script `lifecycle.py` carries out every step with exact arithmetic and asserts, after every operation, that the ledger has only grown and that replaying it reproduces the store. The walk is an illustration of the specification on invented data and is not an empirical finding.

Encoding and partial-cue reconstruction (S1 to S3). Encoding the three episodes appends three inscriptions and creates three traces. The cue $(\mathit{Ann}, \mathit{morning})$ is partial, since it does not mention the place. Retrieval gives weight 0.962 to the first trace and 0.037 to the second, with a cosine margin of 0.408 between the best and the second best trace, so the first trace is the single attributed trace. The reconstruction contains the details *Ann*, *cafe* and *morning*; the cafe is supplied by the stored trace and not by the cue, and it is warranted by that

Step	Operation	Persistent traces after the operation	Ledger records
S0	empty store	(empty)	none
S1	encode E1	$m_1 = (1,0,1,0,0,1,0)$	1i
S2	encode E2, E3	$m_1 = (1,0,1,0,0,1,0)$ $m_2 = (1,0,0,1,1,0,0)$ $m_3 = (0,1,1,0,1,0,0)$	1i 2i 3i
S3	partial cue (Ann, morning)	$m_1 = (1,0,1,0,0,1,0)$ $m_2 = (1,0,0,1,1,0,0)$ $m_3 = (0,1,1,0,1,0,0)$	1i 2i 3i 4p
S4	misleading cue (adds rain)	$m_1 = (1,0,1,0,0,1,0)$ $m_2 = (1,0,0,1,1,0,0)$ $m_3 = (0,1,1,0,1,0,0)$	1i 2i 3i 4p 5p
S5	misleading report, gate closed	$m_1 = (1,0,1,0,0,1,0)$ $m_2 = (1,0,0,1,1,0,0)$ $m_3 = (0,1,1,0,1,0,0)$	1i 2i 3i 4p 5p 6e
S6	corroborated evidence, revise m1	$m_1 = (1,0,1,0,0,0,1)$ $m_2 = (1,0,0,1,1,0,0)$ $m_3 = (0,1,1,0,1,0,0)$	1i 2i 3i 4p 5p 6e 7e 8t
S7	later recollection (Ann, evening)	$m_1 = (1,0,1,0,0,0,1)$ $m_2 = (1,0,0,1,1,0,0)$ $m_3 = (0,1,1,0,1,0,0)$	1i 2i 3i 4p 5p 6e 7e 8t 9p
S8	replay from ledger (state identical)	$m_1 = (1,0,1,0,0,0,1)$ $m_2 = (1,0,0,1,1,0,0)$ $m_3 = (0,1,1,0,1,0,0)$	1i 2i 3i 4p 5p 6e 7e 8t 9p
S9	audit: rewind 8t, read only (store unchanged)	$m_1 = (1,0,1,0,0,0,1)$ $m_2 = (1,0,0,1,1,0,0)$ $m_3 = (0,1,1,0,1,0,0)$	1i 2i 3i 4p 5p 6e 7e 8t 9p
S10	reverse revision: append inverse transformation	$m_1 = (1,0,1,0,0,1,0)$ $m_2 = (1,0,0,1,1,0,0)$ $m_3 = (0,1,1,0,1,0,0)$	1i 2i 3i 4p 5p 6e 7e 8t 9p 10t

Table 45.1: The persistent state after every operation of the lifecycle. Ledger records are numbered in order of creation and suffixed by type: i inscription, p interpretation, e evidence, t transformation. Retrieval steps change only the ledger, and the only steps that change a trace are S6 and S10.

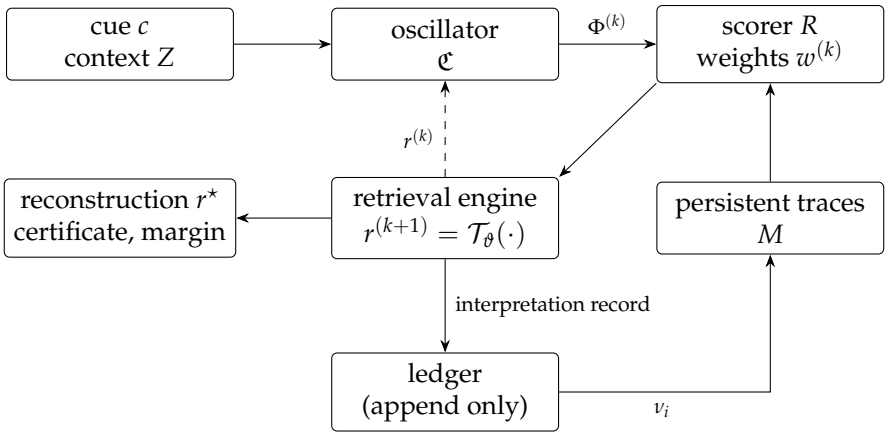


Figure 45.1: Module structure of the reference implementation. Each box is an independently replaceable module; the dashed arrow is the feedback of the current reconstruction into the context operator and the scores, absent in the minimal model of Appendix L.

trace, so the warrant coverage is 1.00. The store is unchanged and the ledger gains an interpretation record, 4p.

Misleading cue (S4). The cue $(\mathit{Ann}, \mathit{cafe}, \mathit{morning}, \mathit{rain})$ adds a feature that episode E_1 does not contain. The first trace still receives weight 0.834 and is attributed, but the reconstruction now reports *rain* among its details, with component 0.333 in the reconstructed vector, supplied by the cue and by the other traces and not by the attributed one. Warrant coverage is 0.75, three of the four details, and the report labels *rain* as inferred. The example also shows the danger that Proposition 28.1 describes. If the reconstruction were written back to the first trace with rate $\frac{1}{2}$ and no evidence, the trace would move by 0.159 in norm and acquire a rain component of 0.139 that no observation supports. The specification does not do this; the store is unchanged and the ledger gains 5p.

Misleading and then corroborating evidence (S5 and S6). A single report that the episode took place at the library, from a source of reliability 0.6, has likelihood ratio 1.50 and the gate stays closed at $\Lambda^* = 10$. The evidence is nonetheless recorded (6e), so the rejected report remains auditable. Three reports from three distinct sources, each of reliability 0.7 and each saying that the episode was in the evening, give $\Lambda = 12.70$ and the gate opens. The revision moves the first trace by the increment $\Delta = e - m_1$ with $e = (\mathit{Ann}, \mathit{cafe}, \mathit{evening})$, which flips the morning and evening features and moves the trace by 0.000 in norm. The ledger gains the evidence record 7e and the transformation 8t, which cites the inscription 1i and the evidence 7e. The one-vote-per-source rule of Figure 42.1 matters here, since three copies of one report would have given only $\Lambda = 2.333$ (Counterexample 28.1).

Later recollection (S7). The cue $(\mathit{Ann}, \mathit{evening})$ now attributes weight 0.962 to the revised trace and returns warrant coverage 1.00 with an interpretation record that cites the transformation 8t. The revision also changes how the original cue behaves: the cue $(\mathit{Ann}, \mathit{morning})$, which gave weight 0.962 before the revision, gives only 0.491 afterwards, below the attribution threshold, so no single trace is attributed to it and the specification reports that no trace accounts for the answer. This is the intended consequence of a justified revision and

not an error.

Replay, audit and reversal (S8 to S10). Replaying the ledger from the inscriptions reproduces the store exactly (S8). Subtracting the logged increment from the current state recovers the state at S5, so the audit of S9 changes nothing. Reversal (S10) appends an inverse transformation citing 8t, which returns the first trace to its original value while the ledger grows to 10 records, none deleted, and a second replay again reproduces the store. The history of the memory therefore contains the revision and its undoing, and both remain open to inspection.

Extension 45.1 (Lifecycle invariants)**.** Let the operations of Figures 40.1, 41.2 and 42.1, together with encoding, be applied in any order with the one-vote-per-source rule and exact arithmetic. Then after every operation (i) the record set is a superset of the record set before it, (ii) REPLAY$(L) = M$, and (iii) a retrieval leaves M unchanged. The first two are Propositions 42.1 and 44.2 applied to this operation set, and the third is read off line 11 of Figure 40.1.

The extension is a restatement and is included because the lifecycle verifies it at every step of one trajectory, which is evidence that the specification is consistent, and no more than that. It says nothing about whether the revisions are correct, since correctness depends on the evidence, and the evidence in the example was constructed.

Synthesis of Part VIII

The architecture of Part VIII can be summarized by one rule and the checks that enforce it: nothing that has been written is removed, and every state is a function of the record set. The retrieval listing returns a reconstruction together with the attribution and warrant that justify it, and leaves the store alone (Figure 40.1). The reconsolidation listing admits a change to a trace only when evidence aggregated over distinct sources passes the gate, and it logs both the evidence and the increment (Figure 42.1). The replay listing derives the store from the record set in a canonical order, which is what makes replicas that hold the same records agree even when their updates do not commute (Worked Example 44.1). The lifecycle ties these together: the misleading cue, the single report and the three reports exercise the three failure modes that the earlier parts identified, namely self-reinforcing distortion (Proposition 28.1), uncorroborated revision, and double counting of evidence (Counterexample 28.1).

The links to the earlier parts are as follows. The sufficiency criterion of Part VII applies to the combined encoding of store and ledger, by the first qualification after Proposition 32.6, so the ledger is what allows a decoder to report that a detail is inferred, that a trace is no longer attributed to a cue, or that a state existed before a revision; none of these answers is available from the store alone. The quotient of Extension 37.1 determines which traces may be merged or released without leaving the viable set, and the merge check of Worked Example 42.1 is the conservative version of that determination in the absence of structural knowledge. The scheduling of Chapter 43 acts on the retrieval of Figure 40.1, and Worked Example 43.1 shows that its stopping rule depends on the certified error of the cue and not on a count. The rates and thresholds of Part VI enter through η, Λ^* and β.

What the part does not establish is performance. The examples are exact computations on invented data chosen to display the logic of the specification, and the experimental protocols of Part IX remain protocols. The pilot appendix lists the scripts `examples.py` and `lifecycle.py` that produced every number in Parts VI to VIII, so that a reader can rerun them and obtain the same values.

Part IX

The Experimental Program

The chapters of this part are protocols. Each states a hypothesis, the construction of the data, the systems compared, the measurements, the controls, and the criteria under which the hypothesis is counted as supported or as refuted. No experimental outcome is reported in this monograph. The algebraic checks of Appendix M verify identities and bounds of Parts II and III and are not evaluations of any claim made here. Where a protocol uses a quantity from the literature, it is quoted as reported by its source.

Two rules apply throughout. The first is the matched-baseline rule of Theorem 9.4: every comparison with a complex-valued system includes a real-valued system of equal information content, because on the sign-phase subfamily the resonance score reduces to a norm-weighted cosine, and any advantage must therefore be located in the norm factor, in the operators and feature map, or in learning. The second is a requirement of pre-specification: the primary metric, the analysis, and the acceptance threshold are fixed before data are generated, and results for which they were chosen after inspection are labeled exploratory.

Sample sizes are determined by the comparison of two proportions. For a binary success outcome with success probabilities p_0 and p_1 under two independent arms, a two-sided test at level α with power $1 - \beta_{\mathrm{p}}$ requires approximately

$$n \;\geq\; \frac{(z_{1-\alpha/2} + z_{1-\beta_{\mathrm{p}}})^2 \big(p_0(1-p_0) + p_1(1-p_1)\big)}{(p_1 - p_0)^2} \tag{45.1}$$

queries per arm, and the paired design used when the same queries are posed to both systems reduces this by the correlation between the arms' outcomes. These are planning formulas and the actual effect sizes are unknown until the experiments are run.

Chapter 46

Reproduction of the Wave-Based Retrieval Paper

The first experiment reproduces the reported behavior of Listopad's system [1] before any extension is tried. The paper represents patterns as complex waveforms $\psi(x) = A(x)e^{i\phi(x)}$, scores pairs by the resonance score of Chapter 6, defines the semantic operators NEG, SHIFT+, INT_UP and INT_DOWN on a compact vocabulary, and reports distance histograms, a heatmap of pairwise distances, and retrieval latency. The operators are not given formal definitions in the paper; the candidate definitions of Definition 7.1 are therefore the object of the reproduction, and the first deliverable is a report of whether the candidates reproduce the qualitative structure the paper describes, namely that cosine distances for the operator outputs concentrate near a common value while resonance distances separate them into blocks, with NEG aligned with the anti-phased opposite and INT_UP and INT_DOWN separated by amplitude.

The latency measurements are reproduced under the reported conditions: pattern length $L = 1024$, a store of 500,000 patterns, a scalar Java backend, twelve threads, and cold cache. The paper reports an average top-one latency of approximately 135 ms with a 95th percentile near 152 ms, and comparable figures for top-ten and top-hundred queries. The reproduction records the same statistics on its own hardware, with the hardware and software configuration reported alongside. Because the paper describes latency as growing linearly with the dataset size while the reported measurement is at one size, the reproduction measures at several store sizes, from 10^3 to 5×10^5, and fits the dependence; linear growth is expected for an exhaustive scan and is a hypothesis to be tested, not an assumption.

The reproduction includes the comparison the original does not make. Against the unmodified real-valued embeddings, the proposal requires matched real-valued representations of equal information capacity. Concretely, for each complex pattern in $\mathbb{C}^n$ the matched baseline uses a real vector in $\mathbb{R}^{2n}$ with the same number of free parameters, and retrieval by the norm-weighted cosine ρ cos and by the plain cosine. By Theorem 9.1 a retrieval operation that is complex-linear is a special case of a real-linear one, so the baseline is at least as expressive. A reproduction that finds the complex system superior to unmodified cosine retrieval and not superior to the matched baseline has localized the effect to the norm factor, the operators, or the feature map, which is itself a result of the program.

Acceptance. The reproduction succeeds if the qualitative structure of the operator results is recovered under the candidate definitions and the latency is within a stated factor of the reported figure on comparable hardware. It does not succeed or fail on the question of superiority, which belongs to the later experiments.

Chapter 47

Ecphoric Pattern Completion

The hypothesis is that Mem8 recovers the correct episode, not merely the nearest stored vector, when the cue is partial, observations are missing, temporal context is corrupted, or competing historical explanations exist. The experiment is designed so that nearest-neighbor retrieval has a determinate failure mode, which the sufficiency theorem identifies: an episode set is constructed in which several episodes share their content and differ in order, context or outcome, so that retrieval by content alone cannot succeed for the target recollection of Definition 32.1.

Episodes are generated from a controlled grammar with V items, episodes of length ℓ, and a context label drawn from a finite set. Cues are formed by deleting a fraction of the episode, replacing items by distractors, shifting the temporal code, or presenting an ambiguous partial sequence compatible with two or more stored episodes. The target recollection function $T(h, c)$ is specified for each cue type, and the encoding is audited with the sufficiency criterion before any decoder is run: if the criterion fails for the target, the experiment reports the identifiability bound of Proposition 32.6 as the ceiling, and a decoder reaching it is credited with optimal performance, while a decoder above it is a bug. Success is measured by the exact-match rate of the reconstructed episode, by the distance of the reconstruction to the target in the realified space, and by the rate at which the returned episode is a stored episode that differs from the target, which separates lure responses from noise. Systems compared are nearest-neighbor retrieval with cosine, with resonance, the softmax reconstruction without context, the softmax reconstruction with context and phase, and the nonoscillatory recurrent baseline. The sample size for a given contrast follows (45.1), and cue types are analyzed separately because the lure and the noise conditions test different mechanisms.

Chapter 48

Temporal Sequence Reconstruction

The hypothesis concerns the recovery of event order, intervals, nested episodes and temporal boundaries, and it is tested for the claim of Chapter 14 that a multi-frequency phase code supports exact recovery of lag within the period of the code. For frequencies $\omega_1, \ldots, \omega_J$ with integer periods T_j, the lag t is decodable modulo $\mathrm{lcm}(T_j)$ by the Chinese remainder theorem when the periods are pairwise coprime, and the experiment verifies the decoding limit and measures how it degrades with phase noise.

The data are sequences of items with specified onset times, nested structures in which one episode contains others, and boundaries marked by context shifts. Tasks include order judgment between two items, interval estimation, retrieval of the item at a given lag from a cue, and segmentation of a continuous stream into episodes. Systems compared are static embeddings with additive positional codes, recurrent networks, the phase-coded memory with a learned and with a fixed code, and a nonoscillatory dynamical baseline with matched state size. The controls of Chapter 14 apply: a phase-shuffled variant in which the phase code is permuted across items, which tests whether performance depends on the phase assignment, and a frozen-phase variant in which phases are fixed after encoding, which tests whether dynamics contribute beyond a static code. Because a static positional encoding is itself a phase code, a positive result for the phase-coded memory over the recurrent network is not a result for oscillation until the frozen-phase variant has been compared.

Segmentation as a separate test. Temporal phase can encode order and interval but does not determine which events form a single remembered episode, so the experiment separates segmentation from sequence retrieval. Every system is run twice, once with oracle episode boundaries supplied by the generator and once with boundaries inferred by the surprise criterion of Chapter 17 (a boundary where the prediction error of the context state exceeds an adaptive threshold). Boundary quality is scored by the F-measure of detected against true boundaries at a stated temporal tolerance. The segmentation cost is the difference in retrieval accuracy between oracle and inferred boundaries, so that a system cannot be credited for sequence retrieval that depended on boundaries it did not infer, nor charged for a segmentation error as if it were a retrieval error.

Chapter 49

Contextual Interference and Source Attribution

The hypothesis is that systems which store provenance can attribute source and calibrate false recollection, while systems limited to content cannot, as stated in Proposition 19.1. Datasets are constructed in which identical content appears in different contexts and is associated with conflicting historical outcomes. For each item a ground-truth set of sources is recorded. Queries ask for the content, for the source, and for whether the item was previously encountered; the three correspond to retrieval, source discrimination and familiarity.

The measurements are source accuracy, the rate of false recollection (a confident report of an item or source that was not encountered), calibration of reported confidence measured by the expected calibration error and by the reliability diagram, and the sensitivity of each to changes of context between encoding and retrieval, using the encoding-specificity prediction $S = (1 + \cos(z - z_0))/2$ of Proposition 18.1. The centroid lure of Proposition 20.2 is tested directly: stores are constructed with a cluster of related items and a lure at the cluster mean, and the rate at which the system endorses the lure as previously seen is recorded as a function of the inverse temperature β and the number of related items. The systems compared are a content-only softmax memory, the same with context, the same with the ledger of Chapter 41, and a baseline that stores explicit source labels, which sets the ceiling for source accuracy. The prediction following from the sufficiency theorem is that source accuracy of any system without access to source information is bounded by the identifiability limit, and the experiment treats any exceedance of that bound as evidence of leakage in the dataset or the pipeline.

Warrant coverage. In addition to source accuracy, the experiment records the warrant coverage ω of Definition 41.1 for each reconstruction, and the accuracy of reported confidence separately for warranted and inferred details. A system whose details are mostly inferred but reported with the confidence of warranted ones is miscalibrated in a way that source accuracy on the final answer alone does not reveal.

Chapter 50

Consolidation and Reconsolidation

The hypothesis is that evidence-gated updating allows adaptation to new evidence while keeping the original record recoverable and resisting unwarranted revision, as argued in Chapter 28. Streams of episodes are generated with a latent state that changes at known times, so that earlier interpretations become obsolete. Three types of events are interleaved: confirming evidence, contradicting evidence of controlled strength, and reconstructions retrieved by the system itself and fed back as cues without external evidence.

Two quantities are measured and reported together. Adaptability is the speed with which the system's reconstruction follows a change in the world after contradicting evidence, measured by the number of events to reach a stated accuracy on the new state. Resistance to unwarranted revision is the distortion of stored traces after a sequence of self-generated retrievals with no external evidence, measured as the distance from the original inscription; by Proposition 28.1, an update rule driven by the system's own output collapses traces toward the reconstruction, and the experiment tests whether the evidence gate prevents it. The recoverability of the original record is tested by reversal: the earlier state is reconstructed from the current state and the ledger, and compared bitwise with the earlier state (Proposition 28.2). The systems compared are ungated self-referential updating, evidence-gated updating, and a baseline that never updates, which sets the floor for adaptability and the ceiling for resistance. The result of interest is the trade-off curve between adaptability and resistance as the gate threshold varies, and the claim of the monograph is that the gated system dominates the ungated one on the resistance axis at matched adaptability.

Chapter 51

Oscillatory Ablation Experiments

This is the decisive experiment for the necessity of oscillatory coordination. The question, as posed in the proposal, is whether coordinated oscillatory dynamics improve reconstruction under matched parameter counts, computational budgets and information content. The experiment is designed so that a negative answer is as informative as a positive one.

The full system is compared against five variants that differ from it in exactly one module of Chapter 39. The phase-frozen variant fixes phases at their initial values. The phase-randomized variant draws phases from the same marginal distribution independently of state. The frequency-uncoupled variant sets $K = 0$ so that each oscillator runs at its own frequency. The nonoscillatory recurrent variant replaces the Kuramoto dynamics by a learned map with the same parameter count and no sinusoidal structure. The static resonance variant removes the contextual operator altogether, so that scoring uses the content alone. Tasks are those of Chapters 47 to 49, which cover completion, temporal order and source attribution, and the primary outcome is reconstruction accuracy on the target recollection function under each task's cue family.

Matching is the delicate part and is specified in advance. Parameter counts are matched by adjusting the hidden width of the nonoscillatory variant. Computational budget is matched by the number of floating-point operations per retrieval, counted analytically. Information content is matched by holding the encoding E fixed across variants, so that Theorem 32.2 gives the same ceiling to all of them and differences are attributable to the decoder. The analysis uses paired comparisons across a common set of queries with a mixed-effects model for task and seed, and the acceptance rule is stated before the run: oscillation is judged to contribute if the full system exceeds every one of the five variants by at least a pre-specified margin on at least a pre-specified number of tasks, with the seeds as the unit of replication. A result in which the nonoscillatory recurrent variant matches the full system is reported as showing that the oscillatory realization of the computation is not necessary for it, which is consistent with the position of Chapter 55 that the necessary properties are those of the computation and not its physical substrate.

An integrated longitudinal benchmark. The component experiments leave open whether the architecture preserves reconstructive capacity over time or merely contains several independently functional subsystems, so the program includes one benchmark that requires the capacities together. A system encounters partially overlapping episodes, later receives misleading information about one of them, undergoes consolidation and compression, and is then asked to reconstruct a particular episode from an incomplete cue. It must recover temporal order, identify the correct historical source, distinguish original evidence from subsequent interpretation, and report uncertainty. The cycle of misleading information, update, compression and query is repeated, and every metric above is reported as a function of the number of cycles, so that the object of study is the trajectory of performance and not a single retrieval. The central comparison is against a system that has the same provenance ledger, the same recurrent inference mechanism, the same storage budget, and the same training data, and differs only in lacking the oscillatory coordination. Without that control,

an improvement could be attributed to additional information or engineering capacity and not to the proposed mechanism. The prediction following from Theorem 32.7 is that trace-only systems lose answerability irreversibly at the first compression that crosses the commitment of Corollary 32.8, while ledger-backed systems recover it at the cost of replay, and the benchmark measures that cost.

Chapter 52

Scaling and Computational Efficiency

The hypothesis is that retrieval cost and accuracy scale as predicted by the capacity analysis of Chapter 34. Measured quantities are capacity (the number of stored traces at which accuracy falls below a threshold), retrieval latency, throughput in queries per second, update cost per write, energy where it can be measured, and accuracy as the number of related traces grows. The predictions to be tested are specific: the expected number of false activations of Proposition 34.1 as a function of the threshold, the dimension and the store size, and the exponential capacity of Theorem 34.2 with its required scaling of β with the dimension. The random-pattern model is the best case, so the experiment is repeated on structured stores with clustered traces, where overlaps are larger and the capacity predicted by the margin condition is correspondingly smaller.

Both exact and approximate retrieval are measured. Exact retrieval is the exhaustive scan, with latency linear in the store size at fixed dimension. Approximate retrieval uses an index to restrict activation to a candidate set, and the experiment reports the recall of the true top candidate as a function of the candidate size, because an approximate method that discards the correct trace before refinement cannot be rescued by the refinement. Implementations are compared on CPU and GPU, with memory-mapped storage for stores larger than memory, and latency is reported with percentiles and under cold and warm cache separately. Interference is varied by adding competitor traces and measuring the decline of the target weight against the dilution formula of Proposition 20.1. Efficiency claims are reported with the hardware and with the cost of the baseline, since the comparison of latency across implementations on different hardware is not informative.

Chapter 53

Robustness and Failure Analysis

The program requires explicit reporting of failure. Successful resonance in a favorable condition is not evidence of general mnemonic competence, and the experiment is organized around the conditions under which the system is expected to fail, so that the failures are located and characterized.

Seven conditions are examined. Catastrophic interference is tested by the sequential learning of tasks with and without replay, measuring the accuracy on earlier tasks. False attractors are mapped by counting fixed points of the retrieval iteration that are not within a stated distance of any stored trace, as a function of β, P and the overlap structure, against the stability criterion of Theorem 33.2. Unstable synchronization is examined by driving the oscillator toward global synchrony and measuring the loss of selectivity. Phase noise is added at increasing levels to the phase code, and the degradation is compared with the analytic dependence of the score on phase mismatch. Adversarial cues are constructed by optimizing a perturbation of bounded norm to maximize the weight of a chosen wrong trace; since the weights depend on scores through the Lipschitz map of Proposition 30.1, the bound gives the minimum perturbation needed to cross a margin, and the experiment measures how far the empirical attack is from that bound. Provenance corruption is tested by removing or altering ledger records and measuring the system's ability to detect the inconsistency, through the check that a replay reproduces the live state. Overconfident reconstruction is measured as the gap between reported confidence and accuracy on cues for which the target is not recoverable by Theorem 32.2; a system that reports high confidence there is miscalibrated with respect to its own encoding. Each failure is reported with the conditions, the frequency, and the proposed mechanism, and unexplained failures are listed as such.

Part X

Interpretation and Limits

Chapter 54

What the Architecture Explains

A monograph that mixes theorems, simulations, neuroscientific motivation and conjecture invites the reader to assign to the whole the confidence that belongs only to the strongest part. This chapter prevents that by sorting the claims of the preceding parts into four classes. A claim is formally demonstrated if it is proved here from stated assumptions. It is simulation-supported if a computation, with stated protocol and outcome, bears on it. It is a neuroscientifically motivated hypothesis if it is suggested by empirical findings but not established by them. It is speculative if it extrapolates beyond all of these.

The formally demonstrated results are those of Parts II, IV, V, VI and VII, and they are results about the idealized system. They include the closed form $S = \frac{\rho}{2}(1 + \rho c)$ of the resonance score and its invariances (Theorem 6.1), the equivalence of the complex and real parametrizations (Theorem 9.1) and the matched-baseline principle that follows (Theorem 9.4), the ecphoric sufficiency criterion and its minimal-memory corollary (Theorem 32.2, Corollary 32.4), the identifiability limit (Proposition 32.6), the Banach bounds and the contraction and stability criterion of the softmax update (Lemma 33.1, Theorem 33.2), the false-activation and capacity bounds under the random-pattern model (Proposition 34.1, Theorem 34.2), the dilution and centroid-lure results, the drift of self-referential updating (Proposition 28.1), the failure of penalties to guarantee fidelity (Proposition 37.2), and the closure criteria for abstraction. These are conditional on their hypotheses, and the hypotheses are often the substantive content: that traces are random on the sphere, that the dynamics is a contraction, that the target is fixed.

The simulation-supported class is nearly empty by design. The only computations performed in this work are the algebraic checks of Appendix M, which confirm identities and bounds numerically, and they support no claim about the performance of the architecture. The experimental program of Part IX is the means by which this class would be populated.

The neuroscientifically motivated hypotheses are three. The first is that the macroscale oscillatory modes reported by Cabral and colleagues [22] indicate that slow, spatially organized, standing-wave-like dynamics are a feature of the resting rodent brain that could in principle coordinate distributed activity. The study is an imaging study of anesthetized rats, its modes are damped and noise-driven in the authors' own model, and its authors state that the neural and vascular contributions are not separated; it does not address episodic recall, and it is used here as motivation only. The second is that retrieval in biological systems is reconstructive and cue-dependent in the manner of encoding specificity [11]. The third is that phase relations may carry contextual information, for which the literature cited in Chapter 14 is suggestive.

The speculative class contains the proposal that a ledger-based, evidence-gated architecture of the kind specified here would show better calibrated memory modification than ungated ones in realistic settings, that the posterior geometry of Chapter 35 characterizes accessibility, and that any of this bears on biological memory. None is supported beyond the formal results that bound them. Table 54.1 lists the principal claims with their class.

Claim	Status
Closed form and invariances of the resonance score	Demonstrated
Sign-phase reduction to a norm-weighted cosine; matched-baseline principle	Demonstrated
Sufficiency criterion; minimal memory; identifiability limit	Demonstrated
Contraction and local stability of the softmax update	Demonstrated
Exponential capacity under a random-pattern model	Demonstrated, model-conditional
Evidence-gated updating resists self-generated drift	Mechanism demonstrated; effect size untested
Complex-valued retrieval outperforms matched real baselines	Untested
Oscillatory coordination improves reconstruction	Untested
Cabral modes as a substrate for episodic recall	Hypothesis, not supported by the cited study
Posterior geometry characterizes accessibility	Open problem

Table 54.1: Status of the principal claims of the monograph.

Chapter 55

Biological Plausibility and Computational Equivalence

The question is which properties of biological oscillations are necessary for the computation that Mem8 specifies, and which are properties of one implementation. The formal results give a partial answer, because they show which computational properties are needed and are silent on the physical ones.

What the theory requires is the following. The encoding must preserve the recollective class of the history for the committed cue family (Theorem 32.2). Retrieval must select among stored candidates with a margin sufficient for the number of candidates (Proposition 16.1, Theorem 34.2). The cue must be transformable by context so that the same content can be retrieved differently under different contexts (Chapter 18). The refinement must be stable, which is a contraction condition with a certificate (Theorem 33.2). Provenance must be separable from content if source is to be recoverable (Proposition 19.1). None of these requires oscillation. A context operator can be implemented by a phase rotation controlled by an oscillator, by a learned linear map, by a gating network, or by a lookup, and by the matched-baseline principle the oscillatory form is a special case of real-linear transformation with a particular parametrization. Whether the oscillatory parametrization has advantages in sample efficiency, in robustness, or in interpretability is an empirical question that the ablation of Chapter 51 addresses.

Biological systems, if they implement something of this kind, must also meet constraints that the abstract theory ignores: energy, noise, wiring, and developmental feasibility. These may select oscillatory implementations for reasons that have nothing to do with the computational advantage, such as the efficiency of coordinating distributed populations by a common slow rhythm. The converse also holds: an artificial system need not reproduce any of this. A successful artificial analogue of ecphory would be an implementation of the computational requirements above, and it would not carry evidence about the hippocampus unless the biological system were independently shown to implement the same requirements. The inference in the other direction is equally constrained. That a measured biological quantity, such as the slow macroscale mode of the Cabral study, has a form compatible with the resonant dynamics of Chapter 13 shows only that the dynamics are a candidate description of it, since many models produce damped noise-driven oscillations of similar spectral form.

Chapter 56

Recollection without Phenomenology

The word recollection has a functional sense and a phenomenological sense. In the functional sense it names the reconstruction of a prior organization from a cue under a context, with the properties formalized in Parts IV to VII. In the phenomenological sense it names the experience of remembering, the sense of reliving a past event and of its belonging to one's own past, which the psychological literature associates with autonoetic awareness [12]. The monograph is a theory of the first. Nothing in the definitions or theorems implies anything about the second, and the specification forbids drawing such an implication from similarity of computation.

The distinction has practical force. A system that reconstructs a prior state with high fidelity, reports the source, and distinguishes recollection from familiarity, as in Chapter 19, has the functional profile that is used in behavioral research to operationalize recollection. Whether anything further accompanies that profile in such a system is not a question the profile answers, and a design document should not use phenomenological vocabulary as if it settled it. Where the monograph uses terms such as recollection, familiarity and confidence, they are defined by their operational role, and any claim that goes beyond that role is outside the theory. A related restriction applies to first-person temporal experience: the phase codes of Chapter 14 represent lag, and they are not claimed to represent the passage of time as experienced.

Chapter 57

Memory, Identity, and Historical Continuity

A system that revises its memories, changes its internal state, and reconstructs rather than retrieves seems to lack a stable core, and the question is how continuity of the system through time can be understood without positing one. The architecture supplies a precise answer with two ingredients: the chain of records in the ledger and the preservation of constraints.

Two states of the system at different times are historically connected if the ledger at the later time contains the ledger at the earlier time as a prefix in the canonical order of Proposition 44.2, equivalently if the later state was obtained from the earlier by a fold over additional records. This relation is reflexive and transitive and orders the states of a single replica into a chain. It does not require any trace to be unchanged, since traces may be transformed, merged or released, and the ledger records each change. What persists is the record of how the present arose from the past, and the identity of the system over time is carried by that record and not by any part of the representation.

Constraint preservation is the second ingredient. The viability kernel of Chapter 37 defines the set of states from which the system can continue to meet its constraints, including the floor on historical fidelity. A system whose revisions always remain inside the kernel retains the capacity to reconstruct what its constraints require, even as the representation changes. The picture is that of a river whose water changes while the channel persists, with the channel being the constraints together with the record of how they were maintained. The picture has limits worth stating. Continuity of the record is compatible with arbitrary change in content, so the theory does not say that a system with a continuous ledger is the same system in any sense that matters for questions of moral or personal identity; it says only that the continuity relation is well defined, auditable, and independent of an immutable core. The ledger can also be forked, since replicas may diverge (Chapter 44), and the relation then branches. The formal framework does not decide which branch is the continuation, and the decision is part of the specification of the application.

Chapter 58

The Ideal Architecture and Its Realizable Approximations

The ideal architecture of Parts VII and VIII is unconstrained, deterministic and fully instrumented, and no realizable system meets all its specifications. A hierarchy of implementations makes progress measurable, with each level defined by what it adds and each level carrying its own acceptance criterion, so that a failure at a higher level does not retroactively invalidate a lower one.

The levels are cumulative in design and independent in evaluation. A level-2 system that fails the frozen-phase control is still a valid level-1 system, and its failure is reported as the absence of evidence for the contribution of oscillation. The hierarchy also fixes what may be claimed at each stage: results at level 0 are results about resonance retrieval and not about ecphory, and the word ecphory is reserved for systems that meet the criterion of level 1 at least.

Level	Adds	Acceptance criterion
0. Static resonance retrieval	Complex patterns and the resonance score	Reproduces the operator structure and the algebraic identities of Part II; matched real baseline reported (Chapter 46)
1. Recurrent contextual reconstruction	Softmax iteration with context operator	Meets the contraction certificate of Theorem 33.2; improves target recovery over level 0 on the completion tasks of Chapter 47
2. Temporally coordinated ecphory	Oscillator and phase-coded time	Recovers order and lag within the code period; beats frozen-phase and shuffled-phase controls (Chapter 48)
3. Provenance-preserving memory	Ledger, typed records, replay	Replay reproduces the live state bit for bit; source accuracy above the content-only baseline (Chapter 49)
4. Fully adaptive persistent memory	Consolidation, gated reconsolidation, budgeted retrieval	Gated updating dominates ungated on the adaptability and resistance trade-off; viability constraint never violated (Chapters 50, 43)

Table 58.1: Hierarchy of Mem8 implementations with independent acceptance criteria.

Chapter 59

Open Mathematical Problems

Six questions stated in the proposal remain open, and two further problems arise from the dynamic and integrated results of Chapters 32 and 41. Each is formulated here to the extent that the theory allows. They are collected as problems because the results of the monograph bound them but do not resolve them.

Open Problem 59.1 (Capacity under correlated and phase-structured traces)**.** Theorem 34.2 assumes independent uniformly random traces. Derive the capacity, in the sense of the margin condition of Proposition 16.1, for traces with a prescribed covariance, for clustered stores, and for phase-correlated traces, and determine the dependence of the required inverse temperature on the overlap structure.

Open Problem 59.2 (Representational identifiability)**.** Given a family of decoders and only the input–output behavior of the system on the cue family, determine the equivalence class of memory states that are indistinguishable, and characterize when the learned representation is determined up to a symmetry such as the $O(2n)$ invariance of the resonance score (Theorem 6.1).

Open Problem 59.3 (Stability under continual learning)**.** For the two-store system of Chapter 27 with interleaved replay, give conditions on the replay distribution, the learning rate and the readout temperature under which the sufficiency of the slow store for a fixed cue family is preserved as new episodes arrive, or show that no such conditions exist without growth in capacity.

Open Problem 59.4 (Optimal compression under sufficiency)**.** Determine the minimal description length of a sufficient encoding in terms of the structure of $\approx_T$ (Corollary 32.4), and characterize the optimal approximately sufficient encoding at tolerance ε in terms of the sets $A(m, c)$ of Theorem 32.5, including the trade-off with the distortion bound of Proposition 30.1.

Open Problem 59.5 (Provenance-preserving revision)**.** Characterize the class of update rules for which the ledger of Chapter 41 supports both reversibility and commutativity under concurrent replicas (Proposition 44.2), and the cost in replay length of rules that are not commutative.

Open Problem 59.6 (Necessity of oscillatory coordination)**.** Identify a target recollection function, a resource budget and a cue family for which every nonoscillatory decoder of equal parameter count and budget is provably worse than an oscillatory one, or prove that for a stated class of tasks the oscillatory parametrization has no advantage.

Open Problem 59.7 (Compression under stochastic future commitments)**.** In the setting of Theorem 32.7, let the future target families arrive according to a stochastic process. Characterize the optimal compression and rebuild policy of Proposition 32.9 that minimizes expected storage and rebuild cost subject to a bound on the probability that a query arrives for a family the trace store can no longer answer, and extend the theorem to transformations driven by the system's own retrievals.

Open Problem 59.8 (An interaction theorem)**.** The results of this monograph concern components taken one at a time: inner-product resonance, softmax retrieval, factorization through a sufficient representation, contraction, and closure. Theorem 32.7 and Proposition 32.9 couple sufficiency, provenance and adaptive transformation but remain elementary. State and prove a result whose conclusion depends essentially on the joint action of ecphoric reconstruction, an append-only provenance ledger, and adaptive memory dynamics, for example a lower bound on the cost of maintaining answerability under changing questions that no one of the three components alone can match.

Problem 35.2 on the geometry of accessibility belongs with these. The sixth is the one on which the empirical program of Part IX bears most directly, and a negative theoretical answer for a stated class would shape the interpretation of the empirical results.

Chapter 60

Conclusion: Memory as Organized Reinstatement

The central distinction of the monograph is between possessing stored information and retaining the capacity to reconstruct a consequential prior organization under changing conditions. A store may hold every detail of a past and still fail to reinstate what a present cue requires, because the information is not in a form from which the cue can elicit it, or because it has been merged with other information beyond separation. A store may also hold much less than the past and reinstate what is needed, because it preserved the conditions of reinstatement. The sufficiency criterion makes this precise: what must survive encoding and compression is the recollective class of the history for the cue family the system commits to, and everything finer is free to vary.

Around that criterion the monograph arranges three further separations. The separation between information preservation and resource-bounded accessibility places mechanisms, resonance, attractor dynamics, oscillatory coordination and inference, in the role of ways of computing a decoder within a budget, which may succeed or fail irrespective of whether the information is present. The separation between the stability of a reconstruction and its fidelity warns that a contraction converges to a fixed point whether or not the fixed point is the past. And the separation between historical fidelity and future utility insists that a useful reconstruction can be historically unwarranted, and that fidelity should be enforced as a constraint and recorded in a ledger that makes revision auditable.

The results here are conditional, and the experimental program that would test the architecture has been specified and not run. The honest summary is therefore that the theory supplies a criterion, a set of bounds, and a specification, together with a list of what would have to be shown for the architecture to earn the claims made for it, and that the ambition of the work is a mathematical account in which storage, retrieval, reconstruction, historical fidelity and future viability are distinct but formally related functions. Memory, on this account, is organized reinstatement: the maintained ability to bring back, under new conditions and from partial cues, an organization that mattered.

Appendix A

Complex Hilbert Spaces and Realification

Let $\mathbb{C}^n$ carry the inner product $\langle u, v\rangle = \sum_x u_x^* v_x$, linear in the second argument. The monograph uses the opposite convention from the source paper [1], in which the product is linear in the first argument; only the real part $\mathrm{Re}\langle u, v\rangle$ enters the resonance score, and it is the same under both conventions. The realification is the map $\iota : \mathbb{C}^n \to \mathbb{R}^{2n}$, $\iota(z) = (\mathrm{Re}\, z, \mathrm{Im}\, z)$. It is a real-linear isomorphism, preserves norms and the real inner product, $\mathrm{Re}\langle u, v\rangle = \langle \iota u, \iota v\rangle_{\mathbb{R}^{2n}}$, and intertwines multiplication by i with the block matrix $J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$, so that $\iota(\mathrm{i}z) = J\iota(z)$.

Proposition A.1. *A real-linear map A on $\mathbb{R}^{2n}$ is the realification of a complex-linear map on $\mathbb{C}^n$ if and only if $AJ = JA$. The real-linear maps form a space of dimension $4n^2$ and the complex-linear maps a subspace of real dimension $2n^2$.*

Proof. A map L on $\mathbb{C}^n$ is complex-linear if and only if $L(\mathrm{i}z) = \mathrm{i}L(z)$, which under ι reads $AJ = JA$ with $A = \iota L \iota^{-1}$. The commutant of J consists of block matrices $\begin{pmatrix} P & -Q \\ Q & P \end{pmatrix}$ with P, Q real $n \times n$, giving $2n^2$ real parameters, against $4n^2$ for a general real $2n \times 2n$ matrix. □

The proposition is the basis of the remark in Chapter 55 that complex-linear operators are a $2n^2$-dimensional subfamily of real-linear operators, so a real system of the same width is at least as expressive. A phase assignment restricted to $\phi \in \{0, \pi\}$ gives $\psi = \pm A$, a real vector, and in that case the score reduces to the norm-weighted cosine (Proposition 9.3).

Appendix B

The Resonance Score: Derivation and Properties

For $\psi, q \in \mathbb{C}^n$ with energies $E_1 = \|\psi\|^2$ and $E_2 = \|q\|^2$, the score of the source paper is

$$S(\psi, q) = \frac{1}{2} \frac{\|\psi + q\|^2}{E_1 + E_2} R, \qquad R = \frac{2\sqrt{E_1 E_2}}{E_1 + E_2}. \tag{B.1}$$

Write $a = \|\psi\|$, $b = \|q\|$, $\rho = 2ab/(a^2 + b^2) = R$, and $c = \mathrm{Re}\langle \psi, q \rangle / (ab)$ for the cosine of the real angle. Expanding $\|\psi + q\|^2 = a^2 + b^2 + 2ab\, c$ and dividing by $a^2 + b^2$ gives $1 + \rho c$, so that

$$S(\psi, q) = \frac{\rho}{2} (1 + \rho\, c). \tag{B.2}$$

This is Theorem 6.1. Several properties follow immediately. Since ρ and c depend only on the norms and on the real inner product, S is invariant under every orthogonal transformation of $\mathbb{R}^{2n}$, in particular under all unitary maps and all realified rotations. On the unit sphere $\rho = 1$ and $S = (1 + c)/2$, so S is a monotone function of the real cosine. For a query that is a positively aligned copy of ψ at scale λ, $\rho = 2\lambda/(1 + \lambda^2)$ and $c = 1$, so $S = \rho(1 + \rho)/2$; for $\lambda = 2$ one has $\rho = 0.8$ and $S = 0.72$. For an antiphase copy, $c = -1$ and $S = \rho(1 - \rho)/2$, whose maximum over $\rho \in (0, 1]$ is $1/8$ attained at $\rho = 1/2$ and whose infimum is zero, attained as $\rho \to 1$ or $\rho \to 0$. On unit-norm patterns $\sqrt{1 - S} = \frac{1}{2}\|\psi - q\|$, so $1 - S$ is the square of a metric and not a metric itself, since the triangle inequality fails for squared distances in general. The equivalent forms $S = \|\psi + q\|^2 \|\psi\| \|q\| / (\|\psi\|^2 + \|q\|^2)^2$ and $S = (E_1 + E_2 + 2\,\mathrm{Re}\langle \psi, q \rangle) \sqrt{E_1 E_2} / (E_1 + E_2)^2$ are algebraic rewritings of (B.2).

Appendix C

Spectral Graph Theory

For a weighted undirected graph on N nodes with nonnegative symmetric weights W_{ij} and degrees $d_i = \sum_j W_{ij}$, the Laplacian is $L = D - W$. For a vector $f \in \mathbb{R}^N$, $f^\top L f = \frac{1}{2}\sum_{ij} W_{ij}(f_i - f_j)^2 \geq 0$, so L is positive semidefinite with the constant vector in its kernel, and the multiplicity of the zero eigenvalue is the number of connected components. The spectral theorem gives an orthonormal basis of eigenvectors ϕ_k with eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$, and the graph Fourier transform of a signal f is the coefficient vector $\hat{f}_k = \langle \phi_k, f \rangle$. By the Courant–Fischer principle $\lambda_k = \min_{\dim S = k} \max_{f \in S} f^\top L f / \|f\|^2$, so low eigenvalue modes are the smoothest signals on the graph. A real eigenmode ϕ_k is a standing pattern: every node oscillates in phase or in antiphase with the others, which is the discrete analogue of the standing-wave criterion of Chapter 11, where a mode $\varphi = e^{\mathrm{i}\alpha} w$ with w real is standing and a mode whose phases vary across nodes is traveling. Dynamic mode decomposition fits a linear operator to successive snapshots and returns complex eigenvalues whose arguments are the oscillation frequencies and whose moduli are the decay rates, and the complex eigenvectors of such an operator are generally traveling modes.

Appendix D

Coupled-Oscillator Stability

Let K be symmetric and set $\omega_i = 0$. The dynamics $\dot{\theta}_i = \sum_j K_{ij} \sin(\theta_j - \theta_i)$ is the gradient flow $\dot{\theta} = -\nabla V$ of $V(\theta) = -\sum_{i<j} K_{ij} \cos(\theta_j - \theta_i)$, since $\partial V / \partial \theta_i = -\sum_j K_{ij} \sin(\theta_j - \theta_i)$. Therefore $\dot{V} = -\|\nabla V\|^2 \le 0$, and trajectories converge to the critical set of V, which is compact on the torus.

For the explicit Euler discretization $\theta^+ = \theta - \Delta t \nabla V(\theta)$ the descent lemma gives $V(\theta^+) \le V(\theta) - \Delta t\big(1 - \tfrac{1}{2} L_g \Delta t\big)\|\nabla V\|^2$ when ∇V is L_g-Lipschitz. The Hessian has entries $H_{ii} = \sum_j K_{ij} \cos(\theta_j - \theta_i)$ and $H_{ij} = -K_{ij} \cos(\theta_j - \theta_i)$, so by the Gershgorin circle theorem every eigenvalue satisfies $|\lambda| \le 2 \max_i \sum_j |K_{ij}| =: 2\kappa$, and one may take $L_g = 2\kappa$. Therefore V decreases at every step if $\Delta t < 1/\kappa$. With a common frequency $\omega_i = \omega$ the same holds in the rotating frame. With distinct frequencies the flow is not a gradient and synchrony is possible only when the coupling exceeds the detuning, as in the Kuramoto analysis [47, 21]. For all-to-all coupling with $K_{ij} = K/N$ the global minimizer of V is full synchrony, and the remark of Chapter 39 follows that unrestricted coupling maximizes alignment and therefore erases the phase differences that carry context.

The complex Ginzburg–Landau equation (13.1) with $I \equiv 0$ has plane-wave solutions $\Psi = A e^{\mathrm{i}(k \cdot x - \nu t)}$ with $A^2 = (\alpha - D k^2)/\beta$ and $\nu = \Omega k^2 - \omega + \chi A^2$ (Proposition 13.1), and after rescaling the Benjamin–Feir–Newell criterion $1 + (\Omega/D)(\chi/\beta) > 0$, equivalent to $D\beta + \Omega\chi > 0$ for $D, \beta > 0$, is the condition for the uniform oscillation to be stable to long-wavelength perturbations [19]. In the damped regime $\alpha < 0$ the origin is stable, there are no self-sustained plane waves, and oscillations persist only if driven by noise, which is the regime matching the subcritical Hopf model of the Cabral study. The self-sustained regime $\alpha > 0$ is a design choice of the Mem8 oscillator and is not supported by that study.

Appendix E

Recurrent Attractor Convergence

The softmax update $F(r) = \sum_i w_i(r) m_i$ with $w = \mathrm{softmax}(\beta \langle m_i, r \rangle)$ is the gradient map of the energy $\mathcal{E}(r) = \frac{1}{2}\|r\|^2 - \beta^{-1} \ln \sum_j e^{\beta \langle m_j, r \rangle}$, since $\nabla \mathcal{E}(r) = r - F(r)$. Fixed points are critical points of $\mathcal{E}$. The Jacobian is $DF = \beta \, \mathrm{Cov}_w(m)$, positive semidefinite, with trace at most one for unit traces. Theorem 33.2 gives global contraction for $\beta < 1$ and local stability at fixed points where $\beta \lambda_{\max}(\mathrm{Cov}_{w^\star}(m)) < 1$. For the cleanup bound of Proposition 16.1, the weight outside the leading trace is at most $(P-1)e^{-\beta \Delta}$ for a margin Δ. The Banach bounds of Lemma 33.1 give the residual and a priori bounds. A summary of the relations among the margin Δ, the inverse temperature β, and the cap bound of Chapter 34 is that for $P = e^{\kappa(n-1)}$ random traces the margin condition holds with probability tending to one provided $\beta(1 - a - 2\delta) > \kappa(n-1)$, where $a^2 > 2\kappa$ bounds the maximal overlap and δ the cue error.

Appendix F

Variational Inference

For observed cue c and latent cause z with joint $p(c, z)$, and a variational distribution $q(z)$, the variational free energy is $\mathcal{F}[q] = \mathbb{E}_q[\ln q(z) - \ln p(c, z)]$. Writing $p(c, z) = p(z \mid c)p(c)$ gives

$$\mathcal{F}[q] = \mathrm{KL}\big(q \,\|\, p(\cdot \mid c)\big) - \ln p(c), \tag{F.1}$$

so minimizing $\mathcal{F}$ over q is the minimization of the Kullback–Leibler divergence to the posterior, attained at $q = p(\cdot \mid c)$, and the minimum is $-\ln p(c)$. If the memory is a mixture of Gaussians $p(x \mid i) = \mathcal{N}(m_i, \sigma_x^2 I)$ with uniform prior over i and the cue is a noisy observation $c = x + \xi$ with $\xi \sim \mathcal{N}(0, \sigma_c^2 I)$, then the posterior over i given c is the softmax of $\langle m_i, c\rangle / (\sigma_x^2 + \sigma_c^2)$ when the traces have equal norm, so the posterior mean of the stored pattern is $\sum_i w_i m_i$ with $\beta = 1/(\sigma_x^2 + \sigma_c^2)$. This is the interpretation of the softmax readout used in Chapter 21. The decomposition of the expected error into reducible and irreducible parts is the usual one: the squared error of the posterior mean against the truth is the posterior variance, which cannot be reduced by any decoder using only c and the model. Precision weighting multiplies the log-likelihood of the cue by a precision π, which for this family is a rescaling of β, so precision and the selection temperature are not separately identifiable from outcomes alone, as stated in Chapter 23; separate identification needs cue-specific precision manipulations or time-course data. Active inference adds expected free energy as the criterion for action selection and is kept separate from recollection in Chapter 24.

Appendix G

Observability and Identifiability

The results of Chapter 32 are collected here with the proofs of their elementary lemmas. Let $E : \mathcal{H}ist \to \mathcal{M}$ and $T : \mathcal{H}ist \times \mathcal{C} \to \mathcal{Y}$. Define $\ker E = \{(h_1, h_2) : E(h_1) = E(h_2)\}$ and $\approx_T$ as the kernel of the profile map $h \mapsto T(h, \cdot)$. Sufficiency is $\ker E \subseteq \approx_T$. For a finite $\mathcal{H}ist$ with prior p, the optimal accuracy of identifying h from $E(h)$ is the sum over fibers of the maximum prior mass (Proposition 32.6). For a fixed cue c, the optimal accuracy of reconstructing $T(h, c)$ exactly is the sum over fibers m of the largest total mass of a single value of $T(\cdot, c)$ within the fiber. It equals one if and only if $T(\cdot, c)$ is constant on every fiber, which is sufficiency for the single cue. Fano's inequality in the form $\Pr[\text{error}] \geq (H(h \mid E(h)) - \ln 2) / \ln |\mathcal{H}ist|$ gives a lower bound on the error that no decoder can beat. For noisy encodings the criterion is replaced by the comparison of the conditional laws of $T(h, c)$ given the memory state, and the identifiability limit becomes a statement about the Bayes risk of the posterior given the state.

Appendix H

Rate–Distortion Theory

For a source X with a distortion measure d, the rate–distortion function is $R(D) = \min_{p(\hat{x}\mid x):\,\mathbb{E}d \le D} I(X;\hat{X})$. For a Gaussian source with independent components of variance σ^2 and squared-error distortion, $R(D) = \tfrac{1}{2}\log_2(\sigma^2/D)$ bits per component for $0 < D < \sigma^2$ and zero otherwise [45]. The bound of Chapter 30 for a trace with d real components and total squared distortion δ^2 takes $D = \delta^2/d$. For a matrix store with singular values $\sigma_1 \ge \sigma_2 \ge \dots$ the Eckart–Young theorem states that the best rank-k approximation in the Frobenius norm has error $(\sum_{j>k}\sigma_j^2)^{1/2}$ [3]. Both are bounds under magnitude-based criteria. The sufficiency-based compression of Chapter 30 is a different object, because it discards by recollective class and can exceed any magnitude-based method on stores whose structure is largely irrelevant to the committed cue family.

Appendix I

Dynamical Sufficiency under Projection

The closure criterion of Proposition 36.1 is that $\pi(x) = \pi(y) \Rightarrow \pi(F(x)) = \pi(F(y))$. For a finite stochastic kernel P it is strong lumpability: the block sum $\sum_{y \in B} P(x, y)$ depends on x only through the block of x [24]. Closure is not preservation of identifying information: for any F the projection to a single point is closed and discards everything (Example 36.2), and the constant map $F \equiv x_0$ is closed under every projection. For approximate closure with defect ε and an $\bar{L}$-Lipschitz compressed map, the drift after k steps is at most $\varepsilon \sum_{j<k} \bar{L}^j$ (Proposition 36.4). The projector onto the span of the traces satisfies $\Pi F = F = F\Pi$ for the softmax update (equation (36.2)), and is therefore exactly closed.

Appendix J

Computational Complexity

Let P be the number of stored traces, n the complex dimension, and k the number of refinement steps. A single score $S(\psi, q)$ costs $O(n)$ operations, with a constant of a few complex multiplications. An exhaustive top-one query costs $O(Pn)$ operations plus the cost of selection, linear in P at fixed n. The softmax reconstruction costs $O(Pn)$ per refinement step for the scores and the weighted sum, so $O(kPn)$ in total, and by Proposition 43.1 the optimal k grows as $\ln(R \ln(1/\lambda)/\varsigma)/\ln(1/\lambda)$. Memory for the traces is $O(Pn)$ complex numbers, that is $16Pn$ bytes at double precision. Approximate retrieval replaces P by a candidate set of size $C \ll P$ at the cost of the index and a loss of recall. Replay for the ledger of Proposition 44.2 costs the number of records from the first affected position forward, which is linear in the length of the ledger in the worst case and is reduced by snapshots. The cost of the oscillator is $O(N^2)$ per step for dense coupling and $O(N \deg)$ for sparse coupling. These counts are for the algorithms as specified and do not describe any measured implementation.

Appendix K

Reproducibility Protocols

Every experiment of Part IX is run under the following protocol. The random seed, library versions, hardware description and a hash of the configuration are written to the ledger at the start of a run, so that the run is identified by its ledger. The data generators are deterministic functions of the seed. The primary metric, the statistical analysis and the acceptance rule are written to a pre-registration file before the first query is evaluated, and the file’s hash is recorded. All retrievals are logged as interpretation records with the cue, the context, the scores and the iterates, so that any reported number can be recomputed from the log. Replay tests are part of the suite: the state rebuilt from the ledger must equal the live state bit for bit. Confidence intervals are computed by the bootstrap over seeds, with the seed as the unit of replication, and comparisons between systems use the same queries and the same seeds. Negative results and failures are reported with the same completeness as positive ones, and exploratory analyses are labeled as such. Latency measurements report the processor, the memory, the operating system, the language runtime, the number of threads, the cache state, and the percentiles, since averages alone conceal the tail.

Appendix L

A Minimal Mem8 Model

The minimal model places resonance selection, contextual modulation, recurrent reconstruction and oscillatory coordination in one executable system. The state consists of persistent traces $m_1, \dots, m_P \in \mathbb{C}^n$, a context $z \in \mathbb{R}^n$, oscillator phases $\theta \in \mathbb{T}^n$, and a reconstruction state $r \in \mathbb{C}^n$. The context operator is the diagonal phase rotation $U(z, \theta) = \operatorname{diag}\big(e^{\mathrm{i}(z_x + \theta_x)}\big)_{x=1}^{n}$, which is unitary. The scorer is $R(m, q) = S(m, q)$, the resonance score of Appendix B. One retrieval proceeds for $k = 0, 1, 2, \dots$ by

$$s_i^{(k)} = R\big(m_i,\, U(z, \theta^{(k)})\, c\big), \tag{L.1}$$

$$w_i^{(k)} = \frac{\exp(\beta s_i^{(k)})}{\sum_j \exp(\beta s_j^{(k)})}, \tag{L.2}$$

$$r^{(k+1)} = (1 - \eta)\, r^{(k)} + \eta \sum_i w_i^{(k)} m_i, \tag{L.3}$$

$$\theta_i^{(k+1)} = \theta_i^{(k)} + \Delta t \Big[\omega_i + \sum_j K_{ij} \sin\big(\theta_j^{(k)} - \theta_i^{(k)}\big)\Big], \tag{L.4}$$

with $r^{(0)} = 0$, a coupling matrix K symmetric with $\Delta t < 1/\max_i \sum_j |K_{ij}|$, and $\eta \in (0, 1]$.

Two properties of this baseline are stated to prevent overreading. First, the scores depend on the cue and the phases but not on $r^{(k)}$, and the phase dynamics does not depend on $r^{(k)}$ either. With the phases frozen, the weights $w^\star$ are constant in k and the reconstruction relaxes linearly to the weighted mean: $r^{(k)} = \big(1 - (1 - \eta)^k\big) \sum_i w_i^\star m_i$, converging at the exact geometric rate $1 - \eta$. The baseline therefore does not exhibit attractor dynamics. Replacing the cue by the current reconstruction in the score, $s_i = R(m_i, U\, r^{(k)})$, restores the feedback of the Hopfield-type update of Chapter 16 and the analysis of Chapter 33 applies. Second, the baseline has no learning rule, no provenance, no temporal sequence representation and no uncertainty beyond the weights; these are added by the modules of Chapter 45. The model is intended for testing implementations of the interfaces and for the algebraic checks of Appendix M, and not as a claim about memory.

Appendix M

Pilot Algebraic Verification

The script of Listing 1 checks, with a fixed random seed, the identities and bounds used in Parts II, III and VII. It performs no evaluation of the architecture. Table M.1 lists the outputs. The closed form (B.2), the normalized reduction $S = (1 + c)/2$, and the realification identity $\mathrm{Re}\langle u, v\rangle = \langle \iota u, \iota v\rangle$ were each compared with their definitions on random complex vectors, and the maximum absolute errors are at the level of double-precision rounding. For antiphase scaled copies the sampled minimum of S was on the order of 10^{-11}, consistent with the infimum zero stated in Appendix B; the maximum of $\rho(1-\rho)/2$ is $1/8$. The baseline recurrence converged at a ratio of $0.7 = 1 - \eta$ with $\eta = 0.3$, as stated in Appendix L. In the Kuramoto check, V did not increase at any step under the step bound of Appendix D.

Two further scripts supply the examples of Parts VI to VIII. The script `examples.py` (seed 32) computes and asserts every number of the worked examples, counterexamples and extensions of those parts, including the finite history space and the noisy-encoding risk of Chapter 32, the three-trace spectrum and bistability window of Chapter 33, the near-duplicate family of Chapter 34, the lumpability drift and the congruence refinement of Chapters 36 and 37, and the closed forms of Part VI, and it writes the macros that the text uses for these numbers, so that the numerical values quoted from these computations are not typed by hand. The script `lifecycle.py` runs the miniature memory lifecycle of Chapter 45 in exact arithmetic and asserts after every operation that the ledger has only grown and that its replay equals the store. The script `july.py` (seed 25) computes and asserts the figures of Appendix N. The Bayes-risk lower bound of Extension 32.1 was additionally checked on 2000 randomly drawn binary channels.

A second script, `pilot2.py`, checks Proposition 21.2 and Proposition 33.3. On 200 random instances with random store sizes, dimensions, prior and cue precisions and damping $\eta \in (0.05, 1.9)$, the objective J satisfied the stated descent inequality at every step of 60 iterations, and a finite-difference check confirmed the gradient identity $\nabla J = L(x - T(x))$ to within 10^{-6}. On 100 random stores with $\beta < 1$, the fixed point of the softmax update satisfied the weight bound $w_j \le e^{2\beta}/P$ and the distance bound of Proposition 33.3 for every stored trace.

Table M.2 compares the empirical frequency of $\mathrm{Re}\langle u, v\rangle > a$ for independent uniform unit vectors in $\mathbb{C}^n$ with the cap bound $(1 - a^2)^{n-1}$ of equation (34.1), from 4×10^5 trials per entry. The bound holds in every case, and it is loose by a factor of between one and several orders of magnitude, as expected for a union-type bound.

Listing 1. Pilot verification script.

```
"""Pilot verifications for the Mem8 monograph. Deterministic under fixed seeds."""
import numpy as np

rng = np.random.default_rng(8)


def S(psi, q):
    """Listopad resonance score, Eq. (2)-(3) of arXiv:2509.09691, for complex vectors."""
```

Check	Result
Closed form against definition, max abs error	4.4×10^{-16}
Normalized reduction, max abs error	3.3×10^{-16}
Realification identity, max abs error	4.4×10^{-16}
Antiphase scaled copies, sampled minimum of S	2×10^{-11}
Baseline recurrence, observed contraction ratio	$0.700\ (= 1 - \eta)$
Kuramoto, maximum increase of V per step	0
Kuramoto, final order parameter	1.0

Table M.1: Outputs of the pilot verification script.

n	a	Empirical frequency	Bound $(1 - a^2)^{n-1}$
16	0.2	0.1322	0.5421
16	0.3	0.0449	0.2430
16	0.4	0.0106	0.0731
32	0.2	0.0547	0.2821
32	0.3	0.0076	0.0537
32	0.4	0.00043	0.0045
64	0.2	0.0116	0.0764
64	0.3	0.00029	0.0026
64	0.4	0 (none observed)	1.7×10^{-5}

Table M.2: Empirical frequency of $\mathrm{Re}\langle u, v\rangle > a$ against the cap bound.

```
    e1 = np.vdot(psi, psi).real
    e2 = np.vdot(q, q).real
    if e1 + e2 == 0:
        return 0.0
    r = 2 * np.sqrt(e1 * e2) / (e1 + e2)
    num = np.sum(np.abs(psi + q) ** 2)
    return 0.5 * num / (e1 + e2) * r


def S_closed(psi, q):
    """Closed form S = (rho/2)(1 + rho c)."""
    a = np.linalg.norm(psi)
    b = np.linalg.norm(q)
    rho = 2 * a * b / (a * a + b * b)
    c = np.vdot(psi, q).real / (a * b)
    return 0.5 * rho * (1 + rho * c)


def realify(z):
    return np.concatenate([z.real, z.imag])


out = {}

# (a) closed form versus definition, unequal energies
n = 64
err = 0.0
for _ in range(20000):
    psi = rng.normal(size=n) + 1j * rng.normal(size=n)
    q = (rng.normal(size=n) + 1j * rng.normal(size=n)) * rng.uniform(0.1, 5)
    err = max(err, abs(S(psi, q) - S_closed(psi, q)))
```

```
out["closed_form_max_abs_err"] = err

# (b) normalized reduction S = (1 + Re<psi,q>)/2
err = 0.0
for _ in range(20000):
    psi = rng.normal(size=n) + 1j * rng.normal(size=n)
    q = rng.normal(size=n) + 1j * rng.normal(size=n)
    psi /= np.linalg.norm(psi)
    q /= np.linalg.norm(q)
    err = max(err, abs(S(psi, q) - 0.5 * (1 + np.vdot(psi, q).real)))
out["normalized_reduction_max_abs_err"] = err

# (c) realification: same score from the real inner product on R^{2n}
err = 0.0
for _ in range(20000):
    psi = rng.normal(size=n) + 1j * rng.normal(size=n)
    q = rng.normal(size=n) + 1j * rng.normal(size=n)
    x, y = realify(psi), realify(q)
    a, b = np.linalg.norm(x), np.linalg.norm(y)
    rho = 2 * a * b / (a * a + b * b)
    c = (x @ y) / (a * b)
    err = max(err, abs(0.5 * rho * (1 + rho * c) - S(psi, q)))
out["realification_max_abs_err"] = err

# (d) range of S over random draws, including the minimum 1/8 at rho=1/2, c=-1
vals = []
for _ in range(200000):
    psi = rng.normal(size=4) + 1j * rng.normal(size=4)
    q = -psi * rng.uniform(0.0, 1.0)
    vals.append(S(psi, q))
out["min_S_antiphase_scaled"] = float(np.min(vals))
out["theory_min_antiphase"] = 1 / 8

# (e) false activation bound: P(Re<m,q> > a) <= (1-a^2)^(n-1) for unit vectors in C^n
def unit(n):
    z = rng.normal(size=n) + 1j * rng.normal(size=n)
    return z / np.linalg.norm(z)

rows = []
for n_ in (16, 32, 64):
    for a in (0.2, 0.3, 0.4):
        trials = 400000
        m = unit(n_)
        cnt = 0
        for _ in range(trials // 1000):
            Z = rng.normal(size=(1000, n_)) + 1j * rng.normal(size=(1000, n_))
            Z /= np.linalg.norm(Z, axis=1, keepdims=True)
            cnt += int(np.sum((Z @ np.conj(m)).real > a))
        emp = cnt / (trials // 1000 * 1000)
        bound = (1 - a * a) ** (n_ - 1)
        rows.append((n_, a, emp, bound))
out["false_activation"] = rows

# (f) baseline recurrence: with scores independent of r and theta frozen, r_k converges to sum w_i m_i
#     at geometric rate (1-eta)^k exactly.
P, n_, eta, beta = 20, 32, 0.3, 12.0
M = np.array([unit(n_) for _ in range(P)])
cue = M[3] * 0.0
cue = M[3].copy()
cue[: n_ // 2] = 0
cue = cue / np.linalg.norm(cue)
s = np.array([0.5 * (1 + np.vdot(m, cue).real) for m in M])
w = np.exp(beta * s)
w /= w.sum()
target = (w[:, None] * M).sum(0)
r = np.zeros(n_, dtype=complex)
```

```
ratios = []
prev = np.linalg.norm(r - target)
for k in range(12):
    r = (1 - eta) * r + eta * target
    cur = np.linalg.norm(r - target)
    ratios.append(cur / prev)
    prev = cur
out["baseline_contraction_ratios"] = ratios
out["baseline_expected_ratio"] = 1 - eta
out["baseline_top_weight_index"] = int(np.argmax(w))
out["baseline_top_weight"] = float(w.max())
out["baseline_cosine_to_true_trace"] = float(np.vdot(M[3], r).real / np.linalg.norm(r))

# (g) Kuramoto gradient structure: identical frequencies, symmetric K, V nonincreasing under small step
N = 30
K = rng.uniform(0.0, 1.0, size=(N, N))
K = (K + K.T) / 2
np.fill_diagonal(K, 0)
th = rng.uniform(0, 2 * np.pi, size=N)
dt = 0.9 / np.max(np.abs(K).sum(1))


def V(th):
    D = th[None, :] - th[:, None]
    return -0.5 * np.sum(K * np.cos(D))


vs = [V(th)]
for _ in range(400):
    D = th[None, :] - th[:, None]
    th = th + dt * np.sum(K * np.sin(D), axis=1)
    vs.append(V(th))
vs = np.array(vs)
out["kuramoto_max_increase_of_V"] = float(np.max(np.diff(vs)))
out["kuramoto_V_start_end"] = (float(vs[0]), float(vs[-1]))
R = abs(np.mean(np.exp(1j * th)))
out["kuramoto_final_order_parameter"] = float(R)

for k, v in out.items():
    print(k, v)
```

Appendix N

Reconciliation with the July 2025 MEM | 8 Specification

Status and attribution

This appendix records the relation of the monograph to an earlier specification that bears the same name: C. M. Chenoweth and A. A. Chenoweth, *MEM | 8: A Wave-Based Cognitive Architecture for Multi-modal Memory Integration and Consciousness Simulation*, 26 July 2025 [4]. Its byline also lists two language models, and the monograph cites it under its human authors. The earlier specification is cited here as the source of an architecture and of a vocabulary. The monograph does not claim to derive from it, and the relation of the two projects is a matter for the author to state; the sections below make no claim about it beyond the overlap that can be read off the two texts.

The treatment follows one rule. A design component of the earlier specification is mapped, and where the mapping is natural it is given a precise form and a small check, labelled as a worked example, counterexample or extension of this monograph. A quantitative, performance or psychological claim of the earlier specification is listed as reported by its source, with the reason it is not adopted. No result of the earlier specification is asserted here. Every number below is computed by `july.py` from closed forms of this monograph or from figures printed in the earlier specification, and is written to the macro file that the text loads.

N.1 Mapping of components

N.2 Emotional modulation as a prior

The earlier specification scales a trace amplitude by $(1 + \alpha v)(1 + \beta a)$ with valence v and arousal a. Taking the printed constants $\alpha = 0.3$ and $\beta = 0.5$ at the extreme of the ranges, the largest gain is $g = 1.95$ and the smallest, at $v = -1$ and $a = 0$, is 0.7.

Worked Example N.1. The resonance score of Part II for a unit query aligned with a trace of amplitude g is $S(g) = \frac{1}{2}\rho(1 + \rho)$ with $\rho = 2g/(1 + g^2)$. It satisfies $S(g) = S(1/g)$ and $S(1) = 1$. Hence a trace boosted to $g = 1.95$ scores $S = 0.736$, and one attenuated to $g = 0.7$ scores 0.911, both below the unmodulated score 1. Emotional amplification applied to the amplitude therefore lowers the resonance of an otherwise perfect match. Against an unmodulated competitor with the same alignment and a softmax of inverse temperature $\beta = 8$, the boosted trace receives weight 0.108 rather than a preference. If the gain enters instead as a log prior, so that the weight is proportional to $g\, e^{\beta s}$, the boosted trace receives weight 0.661, which equals the shift of the score by $\ln g/\beta = 0.083$.

The computation shows that amplitude is the wrong carrier for salience when retrieval is by resonance, and that a prior weight carries the same intent without penalizing a match. This is a statement about the formal model and not about affect.

Component of the earlier specification	Closest object in this monograph	Status
Complex wave with amplitude, phase, decay and interference	Complex trace, resonance score of Part II, decay of Chapter 33	Retained; the score has a closed form and the decay is treated as a free scale
Context-dependent decay time $\tau(c)$	Recency scale of Part VI	Retained as a parameter family; see Section N.3
Emotional amplitude modulation	Prior weight on a trace	Reformulated; see Worked Example N.1
Noise-floor storage with periodic sampling	Eligibility for inscription	Proposed formalization; see Worked Example N.2
Temporal blankets, hard and soft	Admission filter before inscription	Proposed; hard is a fixed mask, soft is an adaptive mask
Sensor arbitration	Evidence gate with one vote per source	Related; the gate is stricter and bounds any single source
Four reactive layers	Cached retrieval versus reconstruction, plus a protective tier	The two hierarchies coexist; see Section N.6
Custodian guard	Supervisory controller over the viable set	Proposed; see Section N.5
Repetition poisoning prevention	Drift bound of Chapter 33 and the repetition cap	Reformulated; see Counterexample N.1
`.m8` and Marqant formats	Serialization of the append-only ledger	Not specified here; a format is an interface, not a theory
Consciousness simulation	None	Not adopted

Table N.1: Components of the earlier specification and their treatment.

N.3 Context-dependent decay

The earlier specification sets $\tau(c) = \tau_{\text{base}}\, R\, F\, T$ with factor ranges that, at their extremes, give a total multiplier from 0.12 to 3.00, a range of 25 to one.

Extension N.1. With decay $e^{-t/\tau}$ and the recency scale of Part VI, the age at which the resonance score of an aligned trace falls to half is $\operatorname{arcosh}(\varphi)\,\tau = 1.0613\,\tau$, where φ is the golden ratio. The half-score age therefore spans the same factor of 25 across contexts. Whether a given context range is reasonable is an empirical question that the monograph does not decide.

N.4 Noise-floor storage

The earlier specification stores sub-threshold signals with a small probability $p = 0.01$ of examination at each step.

Worked Example N.2. With independent examinations at probability $p = 0.01$, the number of steps to the first examination is geometric with mean 100 and median 69. To detect a persistent signal with probability at least 0.94 requires 280 steps. The figures follow from $1 - (1 - p)^n$ and are confirmed by simulation. They show the cost of periodic sampling as a latency to detection and not a guarantee of detection.

N.5 Repetition and the Custodian

The earlier specification gives a probability of interrupting repeated retrieval of the form $P_{\text{break}} = 1 - e^{-\lambda(C - C_{\text{th}})^2}$.

Counterexample N.1. With $\lambda = 0.01$ and $C_{\text{th}} = 10$, chosen only to display the effect, $P_{\text{break}}(0) = P_{\text{break}}(20) = 0.632$. The square makes the expression symmetric about C_{th}, so a trace that has never been repeated is interrupted with the same probability as one repeated twice the threshold. A one-sided form $1 - e^{-\lambda \max(C - C_{\text{th}}, 0)^2}$ removes the defect, and under it $P_{\text{break}}(0) = 0$.

Extension N.2. Self-reinforcing drift contracts the gap to the target by the factor $(1 - \eta w)^n$ after n retrievals (Chapter 33). With $\eta = 0.05$ and $w = 0.4$, half the gap remains after 34.3 retrievals, and a retained fraction of 0.8 is the cap at $n = 11$. A repetition cap of the Custodian type is thus a bound on a computable quantity, and may be set from the drift rate and a tolerated displacement.

The supervisory role described for the Custodian can be stated within the framework of Chapter 37: a controller that may veto any transition that would leave the viable set, together with a rule for reintroducing suppressed traces. The monograph supplies this reading as a design proposal. It does not supply a psychological safety guarantee.

N.6 Reactive layers

The four layers of the earlier specification, from an immediate protective response through pattern-matched and emotionally modulated responses to deliberation, order responses by latency. The distinction of this monograph between cached retrieval and costly reconstruction is orthogonal: either may serve a layer. A combined system would add a protective tier whose transitions are always permitted, and would leave the rest to the criteria of the earlier parts.

N.7 Coherence normalization

Counterexample N.2. The coherence of Equation 18 of the earlier specification is

$$\frac{\left|\sum_j A_j e^{i\phi_j}\right|^2}{\sum_j A_j^2}.$$

For three in-phase unit amplitudes it equals 3, not a value at most 1. In general it is bounded by the number of components, and the normalized form $\left|\sum_j A_j e^{i\phi_j}\right| / \sum_j A_j$ lies in $[0, 1]$. The order parameter of Chapter 33 uses the latter.

N.8 Reported by the source and not adopted

N.9 Arithmetic of the printed figures

The first three rows are internal inconsistencies. Parallel execution and different measurement boundaries could change the energy reading, but they would have to be stated. The

Claim of the earlier specification	Reason it is not adopted
Insertion, retrieval and storage speed-ups over vector databases	No benchmark protocol, data or hardware baseline is given in enough detail to reproduce
Semantic-preserving compression of 100 to 1	No semantic fidelity measure is defined
Energy per operation, ranking against other systems	Inconsistent with the printed power and latency; see Table N.3
Constant-time associative retrieval against quadratic vector comparison	The cost depends on the index and on the grid; no argument is given
Psychological safety and trauma-prevention properties	No operational definition and no evidence
Consciousness simulation and related claims	No operational definition; outside the scope of this monograph

Table N.2: Claims listed as reported by their source.

Printed figure	Recomputed value
Grid $256 \times 256 \times 65536$ at 32 bytes per cell, memory	4294967296 cells; 137.4 GB at 32 bytes, 171.8 GB at 40 bytes
Compressed record of 32 bytes with five 8-byte fields	Five 8-byte fields total 40 bytes; the paper also prints 48 bytes for one example
13 nJ per operation at 42 W and 12 µs	$42\ \text{W} \times 12\ \mu\text{s} = 504\ \mu\text{J}$, a factor 38769 above 13 nJ; at 83 000 operations per second the energy is 506 μJ
65 536 layers as 18.2 hours at millisecond resolution	65 536 ms is 65.5 s; 18.2 hours corresponds to one layer per second
Speed-ups 973, 292 and 10.7	Reproduced from the printed times and sizes (974, 292, 10.7)

Table N.3: Consistency of figures printed in the earlier specification.

last row confirms only that the ratios follow from the printed times and sizes, which says nothing about whether the times are correct.

Bibliography

[1] A. Listopad. Wave-based semantic memory with resonance-based retrieval: a phase-aware alternative to vector embedding stores. arXiv:2509.09691, 2025.

[2] A. Vaswani et al. Attention is all you need. *Advances in Neural Information Processing Systems* 30, 2017.

[3] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika* 1:211–218, 1936.

[4] C. M. Chenoweth and A. A. Chenoweth. MEM|8: A wave-based cognitive architecture for multi-modal memory integration and consciousness simulation. Preprint, 26 July 2025 (byline also lists two language models; Zenodo record 16436298, DOI 10.5281/zenodo.16436298).

[5] C. Trabelsi et al. Deep complex networks. *International Conference on Learning Representations*, 2018.

[6] D. J. Amit, H. Gutfreund, and H. Sompolinsky. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters* 55:1530–1533, 1985.

[7] D. Krotov and J. J. Hopfield. Dense associative memory for pattern recognition. *Advances in Neural Information Processing Systems* 29, 2016.

[8] D. L. Hintzman. MINERVA 2: a simulation model of human memory. *Behavior Research Methods, Instruments, & Computers* 16:96–101, 1984.

[9] E. F. Loftus and J. C. Palmer. Reconstruction of automobile destruction: an example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior* 13:585–589, 1974.

[10] E. F. Moore. Gedanken-experiments on sequential machines. In *Automata Studies*, Princeton University Press, 1956, pp. 129–153.

[11] E. Tulving and D. M. Thomson. Encoding specificity and retrieval processes in episodic memory. *Psychological Review* 80:352–373, 1973.

[12] E. Tulving. *Elements of Episodic Memory*. Oxford University Press, 1983.

[13] F. C. Bartlett. *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press, 1932.

[14] Flyxion. Hallucination Is Normal: A Geometric Theory of Manifold-Aligned Semantic Dynamics. Unpublished manuscript.

[15] Flyxion. Never Predict Noise: Manifold-Aligned Prediction as Generative Inference. Unpublished manuscript, November 2025, revised May 2026.

[16] G. Buzsáki. *Rhythms of the Brain*. Oxford University Press, 2006.

[17] H. L. Roediger and K. B. McDermott. Creating false memories: remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21:803–814, 1995.

[18] H. Ramsauer et al. Hopfield networks is all you need. *International Conference on Learning Representations*, 2021. arXiv:2008.02217.

[19] I. S. Aranson and L. Kramer. The world of the complex Ginzburg–Landau equation. *Reviews of Modern Physics* 74:99–143, 2002.

[20] J.-P. Aubin. *Viability Theory*. Birkhäuser, 1991.

[21] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler. The Kuramoto model: a simple paradigm for synchronization phenomena. *Reviews of Modern Physics* 77:137–185, 2005.

[22] J. Cabral, F. F. Fernandes, and N. Shemesh. Intrinsic macroscale oscillatory modes driving long range functional connectivity in female rat brains detected by ultrafast fMRI. *Nature Communications* 14, 2023. doi:10.1038/s41467-023-36025-x.

[23] J. E. Lisman and O. Jensen. The theta-gamma neural code. *Neuron* 77:1002–1016, 2013.

[24] J. G. Kemeny and J. L. Snell. *Finite Markov Chains*. Van Nostrand, 1960.

[25] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* 79:2554–2558, 1982.

[26] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex. *Psychological Review* 102:419–457, 1995.

[27] J. M. Zacks, N. K. Speer, K. M. Swallow, T. S. Braver, and J. R. Reynolds. Event perception: a mind-brain perspective. *Psychological Bulletin* 133:273–293, 2007.

[28] K. Ball. An elementary introduction to modern convex geometry. In *Flavors of Geometry*, MSRI Publications 31, Cambridge University Press, 1997, pp. 1–58.

[29] K. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, and G. Pezzulo. Active inference: a process theory. *Neural Computation* 29:1–49, 2017.

[30] K. Nader, G. E. Schafe, and J. E. LeDoux. Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature* 406:722–726, 2000.

[31] L. Lamport. Time, clocks, and the ordering of events in a distributed system. *Communications of the ACM* 21(7):558–565, 1978.

[32] M. Á. Carreira-Perpiñán. Mode-finding for mixtures of Gaussian distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(11):1318–1323, 2000.

[33] M. Demircigil, J. Heusel, M. Löwe, S. Upgang, and F. Vermet. On a model of associative memory with huge storage capacity. *Journal of Statistical Physics* 168:288–299, 2017.

[34] M. Shapiro, N. Preguiça, C. Baquero, and M. Zawirski. Conflict-free replicated data types. In *Stabilization, Safety, and Security of Distributed Systems*, LNCS 6976:386–400, 2011.

[35] M. W. Howard and M. J. Kahana. A distributed representation of temporal context. *Journal of Mathematical Psychology* 46:269–299, 2002.

[36] P. Fries. Rhythms for cognition: communication through coherence. *Neuron* 88:220–235, 2015.

[37] P. Kanerva. *Sparse Distributed Memory*. MIT Press, 1988.

[38] R. Milner. *Communication and Concurrency*. Prentice Hall, 1989.

[39] R. Semon. *Die Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens*. Engelmann, Leipzig, 1904.

[40] R. Semon. *Die mnemischen Empfindungen*. Engelmann, Leipzig, 1909.

[41] S. Amari. *Information Geometry and Its Applications*. Springer, 2016.

[42] S. H. Strogatz. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D* 143:1–20, 2000.

[43] S. Watanabe and S. H. Strogatz. Constants of motion for superconducting Josephson arrays. *Physica D* 74:197–253, 1994.

[44] T. A. Plate. Holographic reduced representations. *IEEE Transactions on Neural Networks* 6:623–641, 1995.

[45] T. M. Cover and J. A. Thomas. *Elements of Information Theory*, 2nd ed. Wiley, 2006.

[46] T. Parr, G. Pezzulo, and K. J. Friston. *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press, 2022.

[47] Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Springer, 1984.