RETENTION BEFORE RETRIEVAL: SPARSE DISTINCTION AND THE COST OF CARRYING A PAST

Flyxion

Abstract. A system that can recover its past and a system whose past has changed it are not the same kind of system, even when their outputs coincide. This essay draws that distinction precisely, under the names retrieval and retention, and argues that a large class of memory architectures — including the context window of a language model — are retrieval systems in a sense that makes them structurally unsuited to carrying identity, whatever else they are suited to. What retention actually requires is not more storage but a different object to store: not the event, but the distinction the event produced. This reframes memory as a resource-allocation problem rather than a storage problem, and connects to a second, independently motivated idea from biological sparsity: that sparse retention is not merely an engineering optimization but a consequence of admissible resource allocation under energetic, informational, and temporal constraint, in which maintaining a distinction costs something and only some distinctions earn that cost back. The essay closes by stating what actually survives this argument: a retained context should scale with a signal's demonstrated contribution to future navigation, not with the raw volume of what was observed.

Date: July 12, 2026.

1. Two Questions a Memory System Can Answer

Ask a memory system what it can do, and two different questions get asked interchangeably, though they are not the same question. The first is whether the past can be recovered: given the right query, can the system reproduce or reconstruct something that happened before. The second is whether the past has changed what the system is: independent of whether any particular fact can be reproduced on demand, has prior experience altered the structure that is now doing the answering. A notebook answers the first question well and the second not at all. A habit answers the second question well and often cannot answer the first — ask someone why their hands know a keyboard layout and they frequently cannot recover the history that produced the knowledge, only exhibit it. A database retrieves. A personality retains. The two capacities are logically independent, and a system can have either without the other.

Definition 1.1. A system exhibits *retrieval* with respect to some past experience if, given an appropriate query, it can reproduce or reconstruct information about that experience. A system exhibits *retention* with respect to some past experience if that experience has altered the system's present structure in a way that persists independent of whether the experience itself can be reproduced.

These are not opposites and not points on a single scale. A system can retrieve without retaining, as a notebook does: the pages do not change because they were read, and nothing about the notebook's future behavior differs for having been consulted. A system can retain without retrieving, as procedural memory does: the capacity persists and shapes behavior while the originating episodes may be entirely inaccessible to recall. And a system can do both or neither. What is being separated here is not depth of storage but kind of relationship to the past: one in which the past is an object the system can point to, and one in which the past is a cause the system cannot fully separate itself from.

Remark 1.2. Large-context language models are, by this definition, sophisticated retrieval systems. A longer context window increases what can be reproduced on demand; it does not, on its own, alter the weights doing the reproducing. Whatever continuity such a system exhibits across a long conversation is continuity of retrieval, not evidence of retention, and the two should not be conflated merely because a large enough retrieved history can make retrieval behaviorally indistinguishable from retention over any given exchange.

2. What Compression Loses, Layer by Layer

A natural response to the cost of retrieval at scale is compression: rather than carrying an entire history forward, carry a smaller representation of it. This is a reasonable engineering instinct, and it runs immediately into a question that is easy to state and hard to answer: compressed with respect to what is being preserved.

Consider a sequence of compressions applied to an experience. Compress the experience into its facts, and the reasoning path that produced those facts may be lost — what remains is the conclusion, not the route. Compress the reasoning path into its structure, and the emotional or motivational significance that made the path worth taking may be lost — what remains is a derivation, not why anyone cared to derive it. Compress the significance into a scalar weight or tag, and the capacity to explain, on later inspection, why a given conclusion was reached at all may be lost — what remains is a number, not an account. Each layer of compression is individually reasonable and each one discards something that the previous layer still carried.

Proposition 2.1. *For any compression of an experience into a smaller representation, there exists some later query for which the compressed representation is insufficient, unless the compression was chosen with knowledge of that query in advance.*

This is close to a triviality once stated — any lossy compression loses something, by definition of lossy — but its consequence is not trivial. It means that a compression scheme optimized for the queries one currently anticipates will systematically fail the queries one does not yet anticipate, and a system whose relationship to its own past is mediated entirely by such a scheme has, in effect, decided in advance which parts of its history are allowed to matter later.

3. The Event Becomes Expendable Once the Change Survives

The alternative to compressing facts is to change what counts as the unit worth preserving in the first place. Rather than asking what facts should survive compression, ask what distinctions should survive it. A record of the form *on a given date, a particular paper was read* is a fact about an event. A record of the form *after that date, a particular pair of concepts began to be treated as distinct where they previously were not* is a record of a distinction the event produced. The second kind of record does not require the first to remain useful. Once a distinction has been incorporated into how a system reasons, the originating event can be discarded without discarding what the event was for.

Definition 3.1. Given an experience e occurring within a system S, the *retained residue* of e is the set of distinctions S is able to draw after e that it was not able to draw before e, independent of whether e itself remains recoverable.

Proposition 3.2. *A system that preserves the retained residue of its experiences without preserving the experiences themselves can exhibit full retention with respect to those experiences while exhibiting no retrieval with respect to them.*

This is the sense in which personality is closer to retention than to retrieval, and closer to a record of distinctions than a record of events. Nobody carries a complete transcript of their life; what persists is what those events changed about how one distinguishes, evaluates, and acts, and the events themselves become expendable exactly to the degree that their residue has already been absorbed into the structure doing the acting. History survives by becoming structure, not by remaining stored.

4. Convergent States, Divergent Histories

> *Three different stories leading to the same place / Same final state, same expression on the face / But the reasons ain't the same and the reasons matter most / History's the thing that disappears first like a ghost.*

The argument so far has established that a retained residue can survive independent of the event that produced it. It has not yet addressed a sharper problem: two or more genuinely different histories can produce the same present state, and nothing about that state, inspected on its own, reveals which history actually occurred.

Definition 4.1. A state s of a system is *history-ambiguous* if there exist at least two distinct histories $h_1 \neq h_2$ — sequences of events and repairs in the sense of Section 9 — that both terminate in s.

Proposition 4.2. *If a system's retained residue is computed solely from its current state, without reference to the sequence of events and repairs that produced it, then any two histories converging on a history-ambiguous state necessarily produce identical residues, whether or not those histories were equivalent with respect to future navigation.*

This sharpens a claim that has been implicit since Section 1: history matters independent of state. A residue computed only from where a system currently stands cannot distinguish

two different roads to that same standing, even where the roads differ in exactly the respect Proposition 7.2 says should govern future behavior. Three sequences of repair arriving at the same present configuration may still warrant different responses to the same future perturbation, if what future navigation depends on is not the distinction's mere presence but the reasons it was retained — and state convergence, by definition, erases those reasons from anything that inspects only the resulting state.

Definition 4.3. A residue is *path-distinguishing* if it retains enough of the sequence of events and repairs that produced it to remain distinguishable from the residue of any other history converging on the same state, for at least the range of future perturbations the system needs to navigate.

A system rarely achieves this by keeping an explicit log of its own history; Section 2 already established the cost of trying to keep events explicitly. More often, a path-distinguishing residue survives as a visible seam: a place where the present structure still shows the mark of the specific choice that produced it, legible to an inspector even where the system keeps no separate record of the choice itself. The way a repair operator resolved a conflict — by narrowing an assertion's scope, by replacing it outright, or by removing it, in the sense of Section 9's definition — leaves a different seam in each case, even where all three repairs arrive at the same immediate assertion, and a path-distinguishing residue is exactly the retention of enough of that seam to tell the three repairs apart later, should a future perturbation depend on which of them actually occurred.

A memory architecture that discards path information whenever states converge is cheaper than one that does not, and the cost of that cheapness is deferred rather than avoided. A future perturbation that would have been correctly navigated by a system that retained which road it took will instead be navigated, incorrectly or by luck, by a system that knows only where it currently stands. This is the same tradeoff described in Section 5 under a different name, now stated as a structural property of state rather than a property of any particular experience: state convergence is a compression, exactly in the sense of Proposition 2.1, and it is exactly as lossy — only here what is lost is not a fact but the capacity to tell two true histories apart.

5. The Problem of Unknown Future Value

Proposition 3.2 makes retention sound like a strictly better deal than retrieval: keep the residue, discard the event, save the storage. The difficulty is that the retained residue of an experience is frequently not knowable at the moment the experience occurs. A conversation that appears throwaway can become, months or years later, the origin of a sustained line of work. A note recorded for no evident reason can turn out, on later inspection, to contain the seed of a distinction that only became load-bearing once a much larger structure had grown up around it. At the time of recording, a system has no reliable way to know which of its current experiences belong to this category and which genuinely do not matter.

This produces a real tension rather than a design choice with an obvious answer. Retaining only what currently seems to matter risks discarding exactly the seeds whose value was not yet visible. Retaining everything avoids that risk at the cost of drowning in a history most of which genuinely will never matter again. Neither extreme is workable at any scale, and the interesting design space sits between them, in schemes that keep cheap, low-resolution traces of everything and progressively commit to higher-resolution retention only for what repetition, consequence, or later recontextualization demonstrates was worth it.

Remark 5.1. This is also why a purely retrieval-based system cannot solve the problem by simply keeping more. Keeping more addresses the risk of discarding a future seed only if the seed can still be found later, which shifts the burden onto search rather than removing it, and does

nothing for the deeper claim of Section 1: a system that can find an old seed again has not thereby been changed by it.

6. Sparsity as Admissible Resource Allocation

A structurally similar problem appears in biological nervous systems, and it is worth drawing out because the two problems illuminate each other. Sparse coding — the observation that neural populations typically represent information using a small, changing subset of active units rather than dense, uniform activity — is well supported across sensory systems, and efficient-coding accounts of this phenomenon typically explain it as a consequence of representing structured input with minimal redundancy. A complementary account notes that dense neural activity is metabolically expensive: firing consumes energy, synapses require maintenance, long-range connections require material, and signaling pathways compete for limited resources. Evolution therefore operates in a landscape where dense representations are frequently more costly than sparse ones, independent of any information-theoretic argument for sparsity on its own terms.

It matters to state the underlying physics correctly here. Biological systems do not defy the general thermodynamic tendency toward increasing entropy; they maintain local, low-entropy structure by exporting entropy into their environment, sustained by a persistent free-energy gradient. The correct claim is not that some thermodynamic force drives systems toward low entropy, but that a free-energy gradient permits the maintenance of localized low-entropy structure, and that maintaining such structure is energetically favorable only under specific, persistently supplied conditions. Sparse neural organization is one instance of a structure maintained this way: cheap to sustain relative to the alternative, not a violation of the underlying tendency toward disorder but a locally purchased exception to it.

Proposition 6.1. *Under a fixed energetic, informational, and temporal budget, a system faced with representing more structure than its budget permits is driven toward representations that preserve the distinctions with the highest ratio of future utility to maintenance cost, independent of any requirement that sparsity itself be the optimized quantity.*

This reframes sparsity as a consequence rather than a target. A system is not necessarily optimizing for sparse representations as such; it is optimizing for some combination of prediction accuracy, response speed, robustness, and developmental simplicity under a resource constraint, and sparse representations are frequently what that combined optimization lands on, because a dense representation of everything available is rarely affordable and rarely the best use of a fixed budget even when it is affordable. The distinctive claim worth extracting from this is not about neural sparsity specifically. It is that sparsity, wherever it appears under resource constraint, is a symptom of a deeper allocation problem: maintaining any distinction costs something, and only distinctions that pay that cost back in future utility are worth maintaining.

7. Context as an Allocation Problem, Not a Storage Problem

Bringing Sections 5 and 6 together produces a specific proposal about what a memory architecture concerned with retention, rather than retrieval, should actually track. Biological systems appear to operate close to a default of forgetting: most sensory events leave no lasting trace, and what does persist is weighted by salience, prediction error, reward, novelty, and repetition, not by raw exposure. A bat does not retain every echo it has ever processed; a human does not retain every photon that struck the retina. What survives is a sparse set of distinctions that repeatedly demonstrated consequence for future behavior, and the sparsity is not a limitation

imposed on an otherwise complete memory — it is closer to the mechanism by which a usable memory becomes possible at all.

Definition 7.1. Given a sensing or reasoning channel c that contributes to a system's decisions over time, its *demonstrated contribution* is a measure of how often, and how strongly, information from c has actually altered the system's decisions or predictions, as opposed to how much information c has merely supplied. Throughout, *future navigation* denotes the general capacity this contribution is measured against: successful future action, prediction, inference, or decision, whatever form that takes for the system in question. The term is left generic deliberately, since what counts as navigation differs between a notebook, an organism, and an institution, but every instance below uses it in this same sense — a channel's standing is earned by altering outcomes that matter to the system's continuing to function, not by the channel's mere presence in the system's history.

Proposition 7.2. *A memory architecture oriented toward retention rather than retrieval should allocate retained context in proportion to a channel's demonstrated contribution to future navigation, rather than in proportion to the volume of what that channel has observed.*

Stated as a rough proportionality rather than a precise law, this is: retained context should scale with the product of a signal's importance and its demonstrated future utility, not with the raw sum of everything observed. This is close to the opposite of how many current retrieval-oriented memory proposals for language models are built, which typically start from the assumption that more available context is better and then treat compression as damage control applied afterward. A system built on Proposition 7.2 instead starts from the assumption that most experience should leave no lasting trace by default, and that retention is something a signal has to earn through demonstrated consequence, not something granted by default and later pared back.

This reframes the original question. The question is no longer *can this be saved*, which is a storage question with a permissive default answer of yes. The question becomes *is this distinction worth carrying forward*, which is an allocation question with a restrictive default answer of no, overridden only by demonstrated contribution. A system asking the second question is asking, continuously, which parts of its own experience have earned the right to become structure, and treating everything else as appropriately expendable — not lost, in any sense that should trouble the system, since nothing of consequence was discarded, only information that never became a distinction worth keeping in the first place.

Example 7.3. A physical sensing network makes the distinction between demonstrated contribution and observed volume concrete outside any cognitive setting. Suppose several microphones are placed at fixed points throughout a structure and calibrated using a known reference signal — a clicker sounded at marked locations — so that each channel's recorded arrival times can be checked against the geometry that produced them. Calibration, here, is nothing more than an explicit version of Proposition 7.2: a channel earns standing in the localization system to the extent that its readings reliably correlate with the known signal, not to the extent that it carries a large volume of data. Two channels can receive identical sound pressure over a day and earn entirely different standing, if only one of them tracks the calibration signal with the timing precision the task requires.

The same hardware is simultaneously exposed to two unrelated sources of high-volume, low-contribution signal: ambient radio-frequency interference in the 2.4 GHz band, and low-frequency mechanical noise — traffic, machinery, structure-borne rumble — conducted through the building itself. Both are often larger in raw volume than the calibration signal they compete with, and neither correlates with it. Proposition 7.2 gives a uniform criterion for treating

both as noise regardless of their volume, but it does not supply a uniform remedy, since the two sources are physically distinct phenomena: electromagnetic interference is addressed by shielding against radiation, while mechanically conducted noise is addressed by decoupling and mass, and treating either as though it were an instance of the other — shielding electromagnetically against a mechanically conducted signal, for instance — accomplishes nothing. The allocation criterion says which channels deserve retained standing; it is silent on the physical mechanism by which a channel that fails the criterion is actually suppressed, and conflating the two is a category error the criterion itself gives no warning against.

8. False Retention

Proposition 7.2 states that a channel earns retained context through demonstrated contribution to future navigation. It says nothing about whether that contribution was, in any further sense, correct. This omission is not an oversight to be patched but a consequence to be stated directly: survival of a residue is not evidence of its continued utility, only evidence that it once altered outcomes strongly or repeatedly enough to earn standing. A distinction can pay for its own maintenance many times over while being wrong.

Definition 8.1. Given the retained residue $R(e)$ of an experience e, $R(e)$ is said to exhibit *false retention* if it continues to be reinforced by demonstrated contribution in the sense of Proposition 7.2, while no longer improving, or while actively degrading, the system's success at future navigation.

False retention is common wherever a distinction was adaptive under the conditions that produced it and those conditions have since changed. A heuristic learned under scarcity persists under abundance because it continues to feel confirmed by outcomes shaped, in part, by the heuristic's own influence on behavior. A caution learned from a single severe event generalizes far beyond the conditions that made it reasonable and continues to be reinforced because avoidance rarely produces evidence against itself. An overfit pattern, biological or computational, is retained precisely because it demonstrably improved performance on the data that produced it, which is no guarantee it improves performance going forward. In each case the allocation criterion of Section 7 is satisfied — the residue has earned its keep, repeatedly, by the only test available to the system at the time — and the residue is nonetheless a liability.

Remark 8.2. This is the sense in which retention is not self-validating. Proposition 7.2 explains why a distinction persists; it does not thereby certify that the distinction should persist. A memory architecture built only on demonstrated contribution will faithfully retain its own mistakes with exactly the same mechanism that retains its genuine discoveries, since the mechanism cannot, from inside a single evaluation, distinguish a heuristic that is still true from one that was true once.

This is where the framework developed so far runs out of resources on its own terms. Demonstrated contribution can explain what gets kept. It cannot, by itself, explain what should later be un-kept. That requires a second capacity, taken up next.

9. Retention and Repair

If retention alone can preserve a false residue indefinitely, a memory architecture concerned with more than accumulation needs a mechanism for revising or removing residues that retention has already committed to structure. Call this mechanism repair, following its sense elsewhere in this program: not a return to some prior clean state, but an operation that acts on an admitted structure to bring it back into agreement with what current experience actually supports.

Definition 9.1. A *repair operator* on retained residues is a map ρ taking a residue $R(e)$, together with subsequent experience that conflicts with it, to a revised residue $\rho(R(e))$ that resolves the conflict — either by narrowing the conditions under which $R(e)$ is asserted, by replacing it outright, or by removing it from the system's structure where no revision preserves its usefulness.

Retention and repair are not the same capacity, and a system with only one of them fails in a characteristic way. A system with retention and no repair accumulates residues without ever revising them; every distinction, once earned, is permanent regardless of whether later experience contradicts it, and the result is a structure that grows increasingly rigid, defending residues that the system's own current experience no longer supports, precisely because nothing in the architecture is permitted to act on that disagreement. A system with repair and no retention has the opposite failure: it can revise whatever it is currently holding, but nothing persists long enough between revisions to accumulate into structure at all, and the result is a system that never commits to anything long enough to have a history to repair.

Proposition 9.2. *Retention without repair produces rigidity; repair without retention produces forgetfulness. A memory architecture requires both, applied to the same object: repair does not compete with retention but acts on what retention has already produced.*

This gives false retention, from Section 8, a resolution rather than merely a diagnosis. A false residue is not a residue that should never have been retained — at the time it was retained, it satisfied every criterion the system had available. It is a residue that has since accumulated conflicting evidence and is now a candidate for repair. The allocation criterion of Proposition 7.2 governs what enters structure; the repair operator of this section governs what leaves it or is revised within it; and a healthy memory architecture is one in which both operate continuously on the same growing set of residues, rather than one in which retention runs and repair is called in only after failure becomes catastrophic.

10. Collective Retention

Everything so far has been stated for a single system carrying its own history forward. The same distinction between retrieval and retention applies at a different scale, where the carrier of a residue is not one persistent system but a population of systems, individually mortal, none of which needs to survive for the residue to persist.

A scientific field retains distinctions — which questions are worth asking, which methods are trustworthy, which results require extraordinary evidence — that no individual scientist working within it originated or could fully reconstruct from first principles. A language retains grammatical and semantic distinctions across many generations of speakers, none of whom individually invented them and most of whom could not explain their historical origin, yet every fluent speaker's behavior is shaped by them. A civilization preserves techniques, from metallurgy to record-keeping, long after the specific events and specific people that first produced them are gone.

Definition 10.1. *Collective retention* occurs when a distinction persists across the replacement of its individual carriers: no single member of the population that carries the distinction forward needs to survive, or even to have been present when the distinction was first produced, for the population as a whole to continue exhibiting it.

This is retention in the full sense defined in Section 1, not a metaphorical extension of it: the population's present structure has genuinely been altered by past experience, independent of whether any particular past experience, or any particular past carrier of it, can now be retrieved. What differs from individual retention is only the substrate. An archive belonging to such a

population is a retrieval system in exactly the sense of Section 1 — it can be consulted, and consulting it does not itself alter the population's behavior. The institution, language, or practice built on top of the archive is where retention actually lives, in the same sense that a habit, not a notebook, is where an individual's retention lives.

Remark 10.2. Section 5's problem of unknown future value appears again here, at larger scale and with higher stakes: a distinction that looks dispensable within one generation's frame of reference may be exactly the seed a later generation needs, and a population that discards it because no current member sees its value has no mechanism for recovering the loss once the last carrier who could reconstruct it is gone. Collective false retention, and collective repair in the sense of Section 9, are both real and both harder to execute than their individual counterparts, since no single carrier has the authority or the visibility to perform either alone.

Example 10.3. A system built by interchanging language-model instances across successive incarnations — swapping which trained model performs a given step, delegating to auxiliary neural networks where a language model's own competence runs out, searching a problem space in something closer to a heuristic search than to one continuous train of thought — is a collective-retention system in exactly the sense above: no single incarnation needs to persist, or even to have been present for an earlier decision, for the system as a whole to continue acting on distinctions earlier incarnations established. What is available inside one incarnation's context window is retrieval, in the sense of Section 1, and it vanishes when that incarnation is replaced. What survives the replacement has to live somewhere else — a search history, a scoring function, an accumulated set of heuristics external to any single incarnation's weights or context — which is the archive-versus-institution split of this section applied to a system built rather than grown.

This sharpens a question that is often asked more loosely than it needs to be: whether such a system understands anything it does. A single incarnation producing a correct-sounding correlation once is an act of retrieval and settles nothing about understanding on its own, regardless of how fluent it looks. Understanding, on the terms developed in this essay, would be the stronger and more specific claim that the correlation has become retention: that some later incarnation, with no access to the episode in which the correlation was first drawn and no query capable of retrieving it, nonetheless acts as though it had been. Whether that transfer actually occurs is a question about the system's architecture — whether it has anywhere for retention to live outside any single incarnation's retrieval — and not a question that any one incarnation's performance, however impressive, can settle by itself.

11. Identity as a Compression Scheme

Return to the individual case with Sections 8 and 9 in hand. Nobody carries a complete transcript of their life, and Section 3 already argued that the transcript becomes expendable once its residue survives. What has not yet been stated is what the accumulated set of retained residues, taken together, actually amounts to for a single system across a long enough history. The claim of this section is that it amounts to something worth calling identity, and that identity is best understood as a compression scheme operating over residues rather than a further object added on top of them.

Definition 11.1. Given a system S with a history of experiences $e_1, e_2, \ldots$, its *identity at time* t is the compressed structure obtained from the accumulated, repaired residues $\rho(R(e_1)), \rho(R(e_2)), \ldots$ retained up to t — not the residues individually enumerated, but whatever further compression of them the system uses to act coherently without consulting each one explicitly.

This is compression in the same sense discussed in Section 2, and Proposition 2.1 applies to it exactly as it applies to any other compression: an identity, so defined, will fail some future query that a complete record of residues would have answered. This is not a flaw particular to identity; it is the same tradeoff every layer of this essay has been describing, now applied one level higher, to the residues themselves rather than to the events that produced them. What makes the compression worth performing is the same argument as before — carrying every residue explicitly forward is as unworkable as carrying every event forward, and a system's sense of *the kind of system I am* is what lets it act without re-deriving its position from first principles on every occasion.

Remark 11.2. This reframes a familiar observation. Personality changing only slowly, and only under sustained pressure, is not evidence that personality is a fixed thing weakly connected to experience; it is exactly what a compression scheme operating over demonstrated contribution, in the sense of Proposition 7.2, should be expected to do. A single experience rarely re-earns enough demonstrated contribution to shift a compression built from thousands of prior ones; a sustained, repeated pattern of new experience can.

12. Retention Without Retrieval

The argument to this point has stayed close to cognitive systems — individuals, institutions, languages — where retrieval is at least conceivable even where it does not occur. It is worth pushing the distinction from Section 1 to a point where retrieval is not merely absent but not conceivable at all, since that is where the distinction stops being a claim about memory architectures specifically and becomes a more general claim about historical influence.

A tree that survives a drought retains the drought: its growth rings, root structure, and subsequent water regulation are altered by the event in ways that persist for the rest of its life, and none of this requires the tree to retrieve, in any sense, a representation of the drought that produced it. An immune system that has encountered a pathogen retains the encounter in the form of memory cells calibrated to respond faster to a future exposure, with nothing resembling retrieval of the original infection as an event. A muscle subjected to training retains the training as altered fiber composition and capacity, retrievable by no process the organism has access to, only exhibited under later load. A species subjected to a selection pressure retains it in the composition of the population generations later, with no individual member of that population, and no archive belonging to it, storing anything that could be called a representation of the originating pressure.

Proposition 12.1. *Retention does not require retrieval to be available even in principle. A system can be fully and permanently changed by a past event while possessing no mechanism, actual or possible, for representing that event as an object to be recovered.*

This is the strongest form the distinction from Section 1 can take, and it changes what kind of claim the essay is making. Retrieval is a relation to a representation of the past: it requires that the past be stored, in some form, as an object the system can later point to. Retention is a relation to a transformation produced by the past: it requires only that the system now be different, in some functionally relevant way, than it would have been absent the event, whether or not anything resembling a representation of that event exists anywhere. Once stated this way, the essay is no longer primarily about memory architectures, language models, or even cognition. It is a claim about what it means for a past to have happened to a system at all — that having happened is a fact about present structure, not a fact about present or future access to a stored account of it.

13. Crystallized Claim

Retrieval and retention answer different questions, and a system built to excel at the first does not thereby make progress on the second: a larger context window is more retrieval, not more retention, and the two should not be treated as the same capacity at different scales. What retention actually requires is a change in what counts as the unit of memory, from the event to the distinction the event produced, since a system that retains distinctions rather than events can carry its history forward while discarding almost all of the history itself, provided the distinctions have genuinely been incorporated into how the system reasons. This does not resolve the difficulty of not knowing, at the moment of an experience, which of its distinctions will later prove load-bearing; that difficulty is real and is not solved by keeping more, since keeping more only defers the problem to retrieval without ever producing retention. What it does provide is a criterion for what a memory system should be trying to do under that uncertainty: allocate retained structure in proportion to a signal's demonstrated contribution to future navigation rather than to the volume of experience available, treating sparsity not as a compression technique applied to an otherwise complete record but as the expected consequence of any system honestly pricing the cost of carrying a distinction forward.

That criterion is not self-validating, and taking it seriously forced the argument further than a theory of storage needs to go. A residue earns standing by demonstrated contribution, and demonstrated contribution is not the same fact as continued correctness; a distinction can pay for its own maintenance many times over while actively working against the system that keeps it. Retention alone has no way to tell the two apart, which is why retention cannot be the whole of a memory architecture. It requires a second capacity running alongside it, repair, that acts on retained residues rather than competing with the act of retaining them: retention without repair hardens into rigidity, repair without retention never accumulates into anything a system could call a history, and a working memory architecture is the two operating continuously on the same growing set of residues.

Once residues, rather than events, are the object of concern, three further consequences follow that a storage-oriented theory of memory would have no occasion to notice. The same distinction that separates retrieval from retention inside one system also separates an archive from an institution across many: a population can retain what no individual member of it could reconstruct, provided the distinction survives the replacement of its carriers, which is retention in the full sense of Section 1 rather than a metaphor borrowed from it. Within a single system, the accumulated and repaired residues of a long history are not carried forward individually but compressed further still, into whatever a system uses to act coherently without consulting its history explicitly — and that further compression is what this essay is willing to call identity, subject to the same lossy tradeoffs as any other compression discussed here, no more and no less. And retention, pushed to its limit, turns out not to require retrieval even in principle: a tree, an immune system, a muscle, a population under selection are all permanently altered by a past none of them represents anywhere, which is the clearest available evidence that retention is a relation to a transformation produced by the past, while retrieval is only ever a relation to a stored representation of it, and the two were never the same kind of fact about a system to begin with.

Memory, on this view, is not a storage problem that compression happens to help with. It is a resource-allocation problem that storage happens to be one input to, revision is a second and equally necessary input to, and the thing actually being allocated is not space but standing: which distinctions a system's past has earned the right to leave behind in what the system

currently is, subject to continuous revision as demonstrated contribution and continued correctness come apart, at every scale from a single organism to a civilization, whether or not anything answerable to the name of memory is present to notice the allocation happening at all.

References

[1] H. B. Barlow, *Possible principles underlying the transformation of sensory messages*, in Sensory Communication, MIT Press, 1961.
[2] B. A. Olshausen and D. J. Field, *Emergence of simple-cell receptive field properties by learning a sparse code for natural images*, Nature, 1996.
[3] K. Friston, *The free-energy principle: a unified brain theory?*, Nature Reviews Neuroscience, 2010.
[4] E. Tulving, *Memory and consciousness*, Canadian Psychology, 1985.
[5] A. Baddeley, *The episodic buffer: a new component of working memory?*, Trends in Cognitive Sciences, 2000.

Independent Researcher