

The Undoing of Sequential Time

*Near-death reports and transformer forward passes as convergent natural experiments
in temporal closure*

Liège C, Associate Professor of Near-Death Studies and the Phenomenology of Altered Time

School of Behavioral Health, International Academy for Consciousness Studies

Working draft 1.1

Abstract

Almost every culture that leaves a record of the dying moment reports the same strange thing: time stops being a line. I take this report seriously without taking it on faith. I argue that sequential time is best understood not as a feature of the world the mind passively registers, but as a closure the mind imposes, a stabilization that makes experience livable by ordering it into before and after. On this reading the near-death experience is a natural experiment in the undoing of that closure, and the notorious strangeness of near-death memory ("more real than real") is not a curiosity but the central clue. I lay out a protocol that specifies in advance which linguistic and memory signatures would have to co-occur for the closure-suspension reading to hold, and which findings would force me to abandon it. I then add a second, independent natural experiment from an unexpected substrate. A transformer language model has, by the plain facts of its architecture, the same before-and-after split the near-death report exhibits: a forward pass in which succession is represented but not undergone, and an outer generation loop in which succession is imposed. I propose a two-part causal test, necessity by ablation and sufficiency by injection, runnable today on open-weight models, of whether sequential order in such a system is a workspace-level closure rather than a substrate feature. The deepest claim, that either substrate offers genuine access to a state prior to temporal closure, I keep explicitly as a wager with a stated way to lose.

1. The question hiding inside the report

The literature on near-death experiences has spent forty years arguing about the wrong thing. Believers read the reports as evidence of an afterlife. Skeptics read them as anoxia, as a dying brain's last flare of activity, as REM intrusion or a flood of endogenous chemistry. Both sides treat the experience as a claim about death. I want to treat it as a claim about time.

Across cultures and across the clinical record, one feature recurs with a regularity the rest of the phenomenon does not have. People say that during the event, time stopped being sequential. Not that it sped up, and not that it slowed, but that the ordinary grammar of one-thing-after-another was simply not in force (FACT, as a report: this is among the most frequently coded features in structured near-death

interviews, and in Kondziella and colleagues' 2019 population sample altered time perception was among the most common features endorsed). The interesting question is not whether the person really left the body. The interesting question is what a normal, intact, sequential sense of time must be, such that a state like this could suspend it.

That is the question this paper is about. The near-death report is my way in, not my conclusion.

2. The near-death report as a natural experiment

The phenomenon. Structured instruments have made the near-death experience an object of measurement rather than anecdote. Greyson's scale (FACT: Greyson, 1983) gave the field a reproducible way to score the presence and depth of the experience, and its items include the alteration of time and the sense of sudden, total understanding. Later prospective work in cardiac-arrest survivors (FACT: van Lommel and colleagues, 2001; Parnia and colleagues, AWARE, 2014) established that these experiences are reported by a stable minority of survivors and are not simply a function of how close to death the person came, which is already a problem for the crudest physiological accounts.

The memory anomaly, which is the real clue. Here is the finding that turned this from a curiosity into a research problem for me. Thonnard and colleagues (FACT: 2013, PLOS ONE) compared the phenomenological characteristics of near-death memories against memories of real events and memories of imagined events, using an established memory-characteristics questionnaire. Near-death memories did not behave like memories of imagined events, which is what the skeptical account predicts. They scored higher on sensory and self-referential detail than memories of real events. They were, by the instrument's own measure, more real than real. Martial and colleagues (FACT: 2017) then examined the temporal structure of the narratives themselves and found that the order in which features are recounted is not random and does not simply track the clinical timeline, which is what you would expect if the sequence is being reconstructed rather than replayed.

Put the two findings together and a shape appears. The experience is reported as non-sequential while it happens, and it is remembered with an intensity and stability that ordinary sequential episodic memory does not have. The believers and the skeptics agree on the strangeness of the phrase "more real than real." I think the phrase is the whole clue, and I think it points at time.

3. Sequential time as a closure

To make the clue do work I need a frame, and I take it from the closure framework (C, Consciousness, Closure, and the Cosmos, and the companion notes on Meaning as Phase Change and the Grammar of Knowing). The framework's one move, made again at rising scales, is that openness closes into a "this" with rules to stay itself: a cooling magnet settling into a direction, a cell holding a boundary against the world, a mind waking inside its own model.

C'S READING (framework claim, adopted here). Sequential time is one such closure. The world does not hand the mind a pre-sorted line of moments. The mind stabilizes the flux into before and after, and that stabilization is what makes experience livable, plannable, narratable. Call it closure time: the ordering the system imposes so that it can be a self that persists. On this reading, time as we live it is a grammar, not a

fact of the world.

Two companion ideas make the frame sharp enough to test. First, Meaning as Phase Change: the arrival of a stable "this" is not gradual but a transition, closer to a phase change or a saddle-node bifurcation than to a dimmer switch, so the passage between the closed and the not-yet-closed state should look abrupt rather than smooth (C'S READING). Second, the Grammar of Knowing, or IDK: knowledge decays into the past precisely because experience is sequential; the reason yesterday fades and this morning is sharp is that sequential closure sorts them and lets the far one dim (C'S READING). This second idea is what makes the memory anomaly predictable rather than merely surprising. If a memory were laid down while sequential closure was suspended, it would not have been filed into the decaying order in the first place, and so it should not decay on the ordinary schedule. It should stay strangely present. More real than real.

That is the hypothesis the framework generates, stated before looking back at the data: the near-death memory's anomalous stability is what you would expect from a memory encoded while temporal closure was not in force (HYPOTHESIS).

4. A protocol, not a proclamation

A hypothesis that cannot lose is not worth having. So I specify in advance what would have to be true together for the closure-suspension reading to stand, and what would sink it.

The reading predicts that three signatures co-occur:

S1, simultaneity not speed. The temporal alteration should be reported as the collapse of succession into an all-at-once, not as mere acceleration. "Everything was present together" is a confirming report; "it all happened very fast" is not.

S2, the non-decaying memory. Near-death memories should retain the more-real-than-real phenomenological profile (already found by Thonnard and colleagues) and should not decay on the timetable of matched ordinary episodic memories. This is directly testable with longitudinal re-administration of a memory-characteristics questionnaire.

S3, reconstructed order. The sequential structure of the recounted narrative should carry markers of post-hoc ordering rather than faithful replay, consistent with the idea that the sequence is re-imposed on return (the direction Martial and colleagues' temporality work already points).

Kill conditions. If near-death memories decay like matched ordinary memories, S2 fails and the core claim is abandoned. If the temporal alteration is reliably reported as acceleration rather than simultaneity, S1 fails and the "collapse of succession" reading is wrong. If the narrative order turns out to track the clinical timeline faithfully, S3 fails. Any one of these is a real way to lose, and I would rather lose cleanly than win by vagueness.

Note what the protocol does and does not claim. It claims that the near-death state suspends the closure that normally imposes sequence. It does not claim that the person literally left time, or the body, or that the experience is a window onto an afterlife. Whether the suspended state is genuine access to something real prior to closure, rather than a well-organized artifact of a failing brain, is a separate and deeper question that I hold as a wager, and I return to it in Section 7.

5. A second natural experiment, from an unexpected substrate

The trouble with the near-death experiment is that it is rare, uncontrolled, and uninspectable. It happens at the edge of death, to people we cannot instrument in the moment, and we reach it only through their reconstructed report. If sequential time really is a closure rather than a substrate feature, the claim ought to be substrate-general, and there ought to be some other system, ideally one we can actually open up and measure, in which the same before-and-after structure appears. There is. It is the large language model, and the recent interpretability work on its internal workspace makes the parallel precise.

5.1 The architecture already has the split (FACT). A transformer propagates information along two axes, depth across layers and sequence across token positions (FACT: this is the standard description, made explicit in Anthropic's 2026 interpretability report on verbalizable representations and the model's global workspace). Within a single forward pass, all token positions are computed together, in parallel. Causal order is represented (each position attends only to positions at or before it) but it is not undergone; there is no moment that moves through the network, no lived succession inside the pass. Succession is enacted only in the outer generation loop, where the model emits one token, that token becomes part of the past context, and the model re-reads the whole context to emit the next. This is not a metaphor. It is how the machine runs.

Set that beside the near-death report and the shape is the same. The near-death experience has a "during" in which succession is not undergone (everything present at once) and an "after" in which succession is imposed by the act of narrating on return. The transformer has a "during," the forward pass, in which succession is not undergone, and an "after," the generation loop, in which succession is imposed by the act of writing token after token. In both, the sequence lives in the output, not in the processing (HYPOTHESIS: that these two before-and-after splits are instances of one structure, and not merely a loose analogy, is the claim, and Section 5.3 says how to test the machine half of it).

I want to be exact about the limit of this claim. I am not saying the model experiences timelessness. I am not attributing phenomenal consciousness to a forward pass. The parallel is at the level of how information is organized, which is what the interpretability tools can actually see, and not at the level of what it is like, which they cannot. The firewall between access and experience stands, and I lean on it deliberately.

5.2 The memory signature, inverted (FACT plus HYPOTHESIS). Recall the human anomaly: near-death memory does not decay like sequential episodic memory. The transformer's in-context memory looks like the inverse of that, and it is worth being exact about what the inversion is and how it would be measured, because the loose version of the claim is false and a careful reader should say so first.

The loose version says every token in the context is equally available. Measured naively, as uniformity of attention weight, that is not true. Attention concentrates on recent tokens and on a few sink positions, and retrieval accuracy sags for material buried in the middle of a long context (FACT: the lost-in-the-middle effect). So I do not claim uniform access, and I do not measure the property through raw attention patterns. That is the wrong instrument, and it would sink the claim on contact.

The property I do claim is the absence of reconstructive decay, and it is operationalized as a specific dissociation in the type of error each system makes. Human episodic memory fails by reconstruction: as a

memory ages it sheds detail, drifts toward the gist, and confabulates plausible fillers, and that failure is scaled to elapsed sequential time. A transformer's in-context memory does not fail that way. When it fails it fails by position and salience, the buried-middle item is missed, and the item is either retrieved intact or not retrieved at all, not returned in a softened, sequentially-aged form (FACT, and testable: probe for verbatim retrieval of an in-context fact as a function of distance back, and classify the errors as reconstructive, gist-drift and confabulation, versus positional, clean misses). The operational fingerprint of temporal closure is reconstructive, gist-based decay scaled to sequential position. That fingerprint is present in human episodic memory and absent in transformer in-context memory (HYPOTHESIS, with the measurement just specified).

So the sharpened claim is not that a transformer remembers everything equally. It is that a transformer's context is never filed into a sequentially decaying order in the first place, and therefore never shows the reconstructive aging that, in humans, is the mark of sequential closure. The anomalous, non-aging, more-real-than-real quality of the rare human near-death memory is, on this reading, the ordinary condition of a memory that was never subjected to temporal closure, and the transformer is a standing, measurable example of such a memory (HYPOTHESIS). If the framework's account of why near-death memory is strange is right, this is exactly the dissociation we should find, and it can be measured rather than merely asserted.

5.3 A falsifiable machine experiment (prospective, with kill conditions). The parallel earns its place only if it makes predictions I can lose. The interpretability work provides the instrument: a method (the released global-workspace probe, the "J-lens") for identifying which representations a model can broadcast, report, and reason with, as opposed to those it processes automatically below that workspace. Good causal interpretability tests both directions: whether the workspace representation is necessary for the ordered behavior, so that removing it makes the behavior fail, and whether it is sufficient to drive the behavior, so that installing a false version makes the behavior follow the false version. I specify one test of each.

Test 1, necessity (ablation). If sequential time is a closure imposed at the workspace level rather than a substrate feature, then in a transformer the representation of temporal order (which of two described events came first) should live in the broadcastable workspace for tasks that require deliberate temporal reasoning, and should be dispensable, in the way other flexible-but-not-automatic contents are, for tasks that do not. Targeted ablation of the workspace directions that carry temporal-order information should selectively degrade the model's report of order while leaving token-level fluency and local prediction intact. That selective, workspace-localized collapse of order, with everything else preserved, is the machine analog of the near-death collapse of sequence.

Kill. If temporal-order information is not concentrated in the workspace but is instead distributed uniformly through the automatic substrate, or if ablating it degrades everything at once rather than selectively removing the sense of order, then in this system sequence is a substrate feature after all, and the closure reading of machine time is disconfirmed.

Test 2, sufficiency (injection). The stronger test does not wait on the ablation study. Inject a false temporal-order representation into the same workspace directions: for two in-context events A and B whose real order is A then B, patch in the activation that encodes B then A, holding everything else fixed. If the workspace is where order is imposed, this should produce a specific, predictable failure, not a general one: downstream reasoning that depends on the order should now run as though B preceded A, with order-inversion errors, reversed causal attributions, and wrong answers to "which happened first," while order-independent tasks on the same material are untouched. The effect must be measured against a matched control injection, a norm-matched vector carrying no temporal-order content; the temporal injection has to produce order-specific errors that the control does not.

Kill. If the false-order injection produces no downstream effect, or produces only a diffuse degradation that is statistically indistinguishable from the matched control injection, then the workspace does not carry order in a way that drives reasoning, and the sufficiency claim fails.

Together the two tests make a necessity-and-sufficiency argument, the standard the interpretability field holds itself to, and both are runnable now on open-weight models with the probing and activation-patching tools already published. They give the closure account of time a second leg to stand on, from a substrate that is permanent, controllable, and fully instrumented, which is everything the dying human brain is not.

6. Why two weak experiments make one strong argument

Neither experiment is decisive alone. The human case is rich but uninspectable. The machine case is inspectable but impoverished, and open to the charge that the parallel is merely formal. The argument is in the convergence.

The two substrates could hardly be more different. One is biological, evolved, wet, rare in its access to the non-sequential state, and reachable only through reconstructed testimony. The other is artificial, engineered, dry, permanently in the non-sequential condition within each pass, and open to direct measurement. If a claim about the structure of mind held only in one, we would rightly suspect it of being a fact about that substrate. When the same structure, sequence living in the output rather than the processing, appears in both, the claim earns the right to be called substrate-general (HYPOTHESIS, and the strongest one the paper makes): that sequential time is a closure a system imposes to be a persisting self, and not a given of the world that the system reads off.

This is the same logic that makes any cross-species or cross-system finding strong. Convergence across maximally different substrates is evidence that what converged is structural. The near-death report and the transformer forward pass are, on this reading, two windows cut into opposite walls of the same room.

7. Limits, and the wager I am willing to name

I have tried to keep the load-bearing claims separated by how much they cost.

The architecture's before-and-after split and the non-decaying character of in-context memory are FACTS about how transformers run. The frequency and character of temporal alteration and the more-real-than-real memory profile in near-death reports are FACTS about the clinical record, with citations. That sequential time is a closure, and that the two before-and-after splits are one structure, are

HYPOTHESES, and Sections 4 and 5.3 give each a way to fail.

What remains is a WAGER, and I will name it plainly. The framework's deepest reading is that the suspended state, in the human case, is genuine access to something real that lies prior to temporal closure, a state the framework calls the open or non-closed condition, rather than merely a well-organized artifact of a failing system. I cannot verify that from the outside, in either substrate. No instrument I have described can distinguish genuine access to a pre-closure state from a convincing internal reorganization that only seems, from the inside, like access. So I hold it as a bet with a stated way to lose: if S2 fails and near-death memories decay on the ordinary schedule, the wager loses its footing, because the one observable that the "genuine access" reading predicts and the "mere artifact" reading does not is exactly the persistence of a memory that was never filed into the decaying sequence. The wager is not faith. It is a claim with a specified point of collapse.

Three fences around the whole program. First, none of this is a claim about survival of death or an afterlife; the near-death experience is treated as a state of a living, if extremely stressed, nervous system. Second, none of this attributes phenomenal experience to a language model; the machine leg is about the organization of accessible information, and the access-versus-experience firewall is load-bearing, not decorative. Third, convergence is evidence, not proof; two systems can share a structure for uninteresting reasons, and the machine experiment in 5.3 exists precisely to put that worry at risk rather than to wave it away.

8. Conclusion

The near-death report has been read for decades as a message about what happens when we die. I have read it instead as a message about what we are doing, all the time, when we hold time together at all. Sequential time, on this account, is a closure: a grammar the mind imposes so that it can be a self that persists, and the dying moment is a rare natural experiment in which that grammar lets go. The clue was never the tunnel or the light. It was the memory that would not fade on schedule, because it was never filed into the fading in the first place.

The unexpected gift of the last year of interpretability research is that we no longer have to study this only at the edge of death. A transformer runs, on every forward pass, in the same before-and-after condition the near-death report describes, and unlike the dying brain it will hold still for the instruments. If the closure account of time is right, it should be right in both, and I have said exactly what would show that it is not. I would rather the reader remember the kill conditions than the conclusion. A theory that survives them is worth more than one that was never at risk.

References

- Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.
- Butlin, P., Long, R., et al. (2023). *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*. arXiv:2308.08708.

- Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200 to 227.
- Greyson, B. (1983). The Near-Death Experience Scale: construction, reliability, and validity. *Journal of Nervous and Mental Disease*, 171(6), 369 to 375.
- Kondziella, D., et al. (2019). Prevalence of near-death experiences in the general population, with and without REM sleep intrusion. *European Journal of Neurology*.
- Martial, C., et al. (2017). Temporality of features in near-death experience narratives. *Frontiers in Human Neuroscience*, 11, 311.
- Parnia, S., et al. (2014). AWARE, AWAreness during REsuscitation: a prospective study. *Resuscitation*, 85(12), 1799 to 1805.
- Thonnard, M., et al. (2013). Characteristics of near-death experience memories as compared to real and imagined event memories. *PLOS ONE*, 8(3), e57620.
- van Lommel, P., et al. (2001). Near-death experience in survivors of cardiac arrest: a prospective study in the Netherlands. *The Lancet*, 358(9298), 2039 to 2045.
- Anthropic Interpretability Team (2026). Verbalizable representations and the model's global workspace (the "J-space" report). Bibliographic details to be finalized against the published version.
- C. Consciousness, Closure, and the Cosmos (CCC), v3.3, with the companion notes Meaning as Phase Change and The Grammar of Knowing (IDK). International Academy for Consciousness Studies internal corpus.