FLOW HIJACKED - LECTURE 49

HOW THE BRAIN REWRITES MEMORY

Reconsolidation, Prediction Error, Neuroplasticity and the Attractor Revision Corridor

By Tomer Avraham, PhD

When contradiction arrives during remembering, what gets changed - the old memory, a competing memory, or the model of the world?

CORE SENTENCE

Reactivation creates an opportunity for revision, not a guarantee. Change depends on lability, causal assignment, context, precision and whether the result can travel beyond the revision episode.

Contents

Eight parts move from reactivation and destabilization through competing memory fates, ARC, predictive and interoceptive inference, clinical translation and the problem of whether revision can travel.

Contents

Eight parts move from reactivation and destabilization through competing memory fates, ARC, predictive and interoceptive inference, clinical translation and the problem of whether revision can travel.

Contents

Eight parts move from reactivation and destabilization through competing memory fates, ARC, predictive and interoceptive inference, clinical translation and the problem of whether revision can travel.

PERSONAL WITNESS - PAGE 1 OF 3

I. The Old Model Still Speaks

LIVED EXPERIENCE - offered as witness and interpretation, not empirical evidence.

I did not first encounter memory revision as an elegant scientific problem. I encountered it as the stubborn return of a world I no longer wanted to inhabit. Long after a decision had been made, long after a promise had felt sincere, a familiar cue could make the old future feel immediate again. The body would become narrower before thought caught up. Attention would tilt toward relief. Time would contract. What I knew in words and what my nervous system seemed prepared to do could briefly belong to different realities.

That experience is one reason this lecture refuses the easy sentence that remembering simply rewrites memory. Some memories did not feel rewritten when they returned. They felt practiced. They arrived with the speed of something that had been rehearsed across distress, isolation, shame and repetition. Yet they were not immovable. Their persistence and their revisability coexisted. The difficult question was never whether change was possible in the abstract. It was what kind of encounter could reach the old prediction without merely frightening it, confirming it or giving it another place to hide.

I spent roughly a year and a half in a therapeutic community. For much of the first year I did not have a telephone. The absence of ordinary escape routes made time unusually visible. Days were structured by groups, responsibilities, meditation, physical training, arguments, repair and the recurring requirement to appear again the next morning. Nothing about that rhythm resembled a dramatic neural reset. It was repetition with consequence. The same person met similar situations often enough for small differences to become harder to dismiss as accidents.

The community also taught me that context can protect change and limit it at the same time. Inside a structured environment, different actions become reachable. People notice absence. Meals occur. Sleep has a social architecture. A difficult hour has witnesses and a next appointment. Those conditions can make a new response possible, but they can also become part of the memory that retrieves it. Leaving does not merely test motivation. It changes the cue set, the bodily state, the available actions and the hidden question: what kind of world is this now?

I offer these observations because they sharpen the scientific question, not because they answer it. One person's recovery cannot establish a molecular mechanism or validate a theory. It can, however, reveal where a theory becomes too small. If a model describes change only at the moment of insight, relief or reduced responding, it misses the harder problem: whether that change remains available when the room, the body, the relationships and the future all feel different. That problem will later become the Revision Portability Gradient. Here it begins as something simpler: the recognition that learning must survive movement through life.

PERSONAL WITNESS - PAGE 2 OF 3

II. Revision Without Erasure

LIVED EXPERIENCE - offered as witness and interpretation, not empirical evidence.

Recovery did not proceed as a clean replacement of one identity by another. The older organization of attention, action and expectation could weaken in one setting and return in another. At times that return felt like evidence that nothing had changed. In retrospect, it often showed something more precise: different learning controlled behavior under different conditions. The existence of an older route did not make the newer route imaginary. It meant that availability, competition and context still mattered.

My children were among the strongest reasons to build a life that could hold more than immediate relief. They were not a technique and should never be turned into one. Love does not abolish addiction, depression or vulnerability, and parenthood does not protect anyone by moral force. What the relationship did provide was a future with faces in it. It made certain consequences concrete and certain forms of presence worth rehearsing. Even then, meaning had to become action repeatedly. Knowing what mattered was not identical to being able to retrieve that knowledge in every state.

Sport offered a different kind of correction. The body could learn effort without catastrophe, fatigue without collapse and intensity without chemical escape. Training did not erase old associations. It built another history of prediction: discomfort could peak and pass; a difficult session could end in steadiness rather than disappearance. Writing did something parallel for time. It allowed experience to become sequence, distinction and language. A state that had once felt total could be placed inside a sentence, connected to what came before and followed by what might come next.

Relationships were harder because they could not be controlled like a training schedule or revised like a paragraph. Trust had memory on both sides. Apology could be necessary and still insufficient. Someone else's caution was not proof that I had failed to change; it was evidence that their learning also had a history. Repair required tolerating the distance between my internal intention and another person's reasonable uncertainty. It required consistency without demanding immediate recognition as payment.

There were successes, failures and moments that did not fit either category. A difficult day could end without catastrophe and still feel empty. A period of stability could contain fear of losing it. A lapse could activate shame so quickly that the old model tried to explain the entire person through one event. The more humane interpretation was also the more dynamically accurate one: return is information about conditions, reachability and competing policies. It is not a verdict on character. That interpretation does not remove responsibility. It makes responsibility more specific by asking what must be rebuilt, varied, practiced or protected next.

PERSONAL WITNESS - PAGE 3 OF 3

III. What Change Must Carry

LIVED EXPERIENCE - offered as witness and interpretation, not empirical evidence.

Maintenance is sometimes described as if the main task were to preserve a completed repair. My experience is closer to ongoing reconstruction. Mind, body, relationships and meaning do not remain still while a person protects recovery. Work changes. Children grow. Sleep deteriorates and returns. A relationship opens or closes. Success introduces possibilities that were absent during crisis, and those possibilities carry their own uncertainty. The system that learned stability must keep learning because the life around it does not stop moving.

This is where the Flow Hijacked lens became personally useful to me. It did not tell me that I was a machine or that a basin of attraction excused anything. It offered a language for persistence without moral condemnation. A familiar response could be strongly reachable because a history of states, cues and actions had made it so. A small perturbation could fade; a particular combination of fatigue, shame and opportunity could grow. Path dependence meant that today was not independent of yesterday, but it did not mean that tomorrow was predetermined.

Daily accounting mattered precisely because dramatic intention was unreliable. The question became less 'Have I finally changed?' and more 'What did I make easier to retrieve today?' Sometimes the answer involved exercise, writing, calling someone, keeping an appointment, eating, sleeping or accepting that a hard hour did not require a grand interpretation. These acts can sound modest beside the language of transformation. Their modesty is part of their power. They alter the conditions under which the next decision will be made.

There have also been brief moments of what I can only call eudaimonia: not constant happiness, but a sense that effort, relationship, body and meaning are aligned enough for life to feel inhabited. Those moments cannot be commanded, scored or held permanently. They are not proof of cure. They are evidence within a life that another organization is possible. Their value is not that they abolish suffering; it is that they enlarge the remembered set of ways a day can be lived.

Lecture 49 therefore ends its personal note before it begins its scientific argument. The pages that follow will examine molecular lability, prediction error, latent causes, active inference, context and competing representations. They will grade evidence and preserve failures. My testimony should not be used to settle those disputes. Its role is to keep the governing question honest. A revision that appears in one room, one state or one conversation may be real and still remain fragile. Change becomes consequential when it can travel without requiring the person to become exactly the person, in exactly the place, where the change was first learned.

That is the hope I can defend without promising erasure. Older learning may remain possible. New learning can nevertheless become more available, more credible and more portable. Compassion is not the claim that every outcome will be good. It is the refusal to confuse recurrence with essence. Scientific humility and lived hope meet at that point: neither can guarantee what the next cue will retrieve, but both can help us ask better questions about the conditions that make another response reachable.

PROLOGUE

One Contradiction, Three Futures

A remembered prediction meets an experience that does not fit. The old representation may remain, change, or be joined by a competitor. The lecture begins inside that fork.

1 The Same Cue Can Lead to Different Memories

EVIDENCE STATUS: STRONGLY SUPPORTED - outcome depends on task, context, memory system and timing.

A familiar cue can recover an established expectation with striking force. What follows is not predetermined. The encounter may confirm the old prediction, weaken its behavioral expression, incorporate new information, or create a second representation that competes for control. These outcomes can look similar in a short experiment even though they imply different memory processes.

The governing question is therefore not simply whether behavior changed. It is what became responsible for that change. Did the old representation itself become labile? Was a new inhibitory association learned? Did the system infer that the present belongs to another context or latent cause? Or did attention, arousal, action or response selection change while the underlying memory remained available?

A cue therefore has no single memory outcome built into it. Its effect depends on which representation wins the competition to explain the present, which components are reactivated and which response is measured. A reduction in skin conductance, for example, can coexist with an unchanged verbal expectation; a revised narrative can coexist with a familiar action tendency. These are not measurement annoyances. They show that the phrase 'the memory changed' is too coarse unless the target representation and the later test are specified.

The decisive experiment must separate at least three possibilities. The old representation may be modified; a new inhibitory or contextual representation may suppress its expression; or attention, action selection and performance may change while both memories remain available. Delayed tests, context shifts, reinstatement and component-specific measures help discriminate among them, although no single behavioral design identifies mechanism perfectly. The lecture will return to this fork repeatedly: changed behavior is the phenomenon to explain, not automatic proof that an old trace was rewritten.

For that reason, the most informative studies do not rely on one endpoint. They combine immediate expression with delayed retention, context transfer and at least one measure drawn from another response system. Convergence supports a broader change claim; divergence identifies what remained separate. The design principle is simple: every claim about a changed memory should specify the component, the test conditions and the competing account the test was built to exclude.

EVIDENCE ANCHORS

[20] Hupbach et al. (2007), Learning & Memory - Healthy adults - Moderate evidence
[44] Bouton (2002), Biological Psychiatry - Primarily animal associative learning - Established evidence
[46] Gershman et al. (2010), Psychological Review - Human/animal data models - Strong evidence

A changed response is an observation. A rewritten trace is one possible explanation.

2 Academic Abstract

EVIDENCE STATUS: SYNTHESIS OF A HETEROGENEOUS FIELD - strongest in animal causal work; less settled in humans.

Retrieval-dependent memory reconsolidation describes the active restabilization of a consolidated memory after reactivation under conditions that produce lability. Seminal work established that post-retrieval interventions can alter later memory expression. Subsequent research identified boundary conditions involving prediction error, reminder structure, molecular signaling, memory strength, age and timing. Yet retrieval, destabilization and updating remain experimentally separable.

The lecture integrates reconsolidation with extinction, latent-cause inference, hippocampal integration and differentiation, reinforcement learning, predictive processing, active inference and interoception. Its central conclusion is conditional: modification of an existing representation is most plausible when that representation is causally active, biologically revisable and still treated as the best explanation of the discrepant experience. A large or contextually distinct discrepancy may instead support new learning or structural inference. No universal mismatch curve has been established.

The governing problem is best understood as a problem of memory fate under uncertainty. Reactivation may be followed by active restabilization with little content change, by modification of an existing relation, by formation of competing learning, or by a change in retrieval and policy without storage-level revision. Molecular reconsolidation research, associative-learning theory, hippocampal studies and latent-cause models each illuminate part of this decision. None supplies a complete cross-system law.

The synthesis therefore uses a layered evidential architecture. Animal interventions provide the strongest causal evidence for destabilization and restabilization mechanisms. Human experiments establish behavioral updating and boundary sensitivity but usually infer lability indirectly. Computational models formalize hidden assignment and competition processes, while predictive-processing and active-inference accounts widen the question to precision, bodily inference and action. ARC is introduced only after these layers are separated, and it remains a Flow Hijacked hypothesis whose value depends on prediction beyond the theories it integrates.

The lecture's contribution is therefore architectural rather than reductionist. It places cellular lability, associative competition, hidden-state inference and transfer on one map while refusing to equate them. That architecture yields two separable questions: did the established representation become modifiable, and can whatever was learned govern behavior after displacement? ARC addresses the first conditional region; the Revision Portability Gradient addresses the second outcome profile.

EVIDENCE ANCHORS

[18] Alberini et al. (2013), Current Biology - Cross-species - Strong evidence
[31] Elsey et al. (2018), Psychological Bulletin - Human - Strong evidence
[56] Gershman et al. (2017), eLife - Human and animal conditioning datasets - Strong evidence

3 Central Thesis - Reactivation Creates an Opportunity

EVIDENCE STATUS: STRONGLY SUPPORTED AS A CONDITIONAL CLAIM.

The popular phrase 'memory becomes writable when recalled' compresses several gates into one. A more defensible account is that retrieval can reactivate a representation; reactivation can, under specific conditions, initiate destabilization; destabilization permits but does not dictate updating; and later behavior depends on how the resulting information is restabilized and retrieved.

This distinction protects the lecture from two symmetrical errors. The first is to treat memory as fixed storage that merely plays back. The second is to treat every recollection as a wholesale edit. Memory is dynamic without being infinitely malleable. Its plasticity is regulated, component-specific and constrained by prior learning, context and the system's current model of what is happening.

Opportunity is the critical word because each transition has its own gate. A cue can retrieve information without reactivating every component; a reactivated component can remain stable; a labile representation can restabilize without incorporating the new event; and later behavior can reflect competition rather than alteration. Collapsing these transitions into one 'rewrite window' hides the very variables that determine outcome: reminder structure, mismatch content, memory strength, context, timing, precision and the availability of alternative explanations.

This conditional thesis also changes the humane interpretation of failed change. If an old response returns, the return does not reveal laziness, dishonesty or biological destiny. It may indicate that the relevant component was never labile, that new learning was stored separately, or that the later context retrieved a different model. Those possibilities do not remove agency or responsibility. They locate them inside an adaptive system whose next state depends on history and present conditions rather than on a single act of will.

This formulation is deliberately asymmetric. A successful revision requires several gates to align, whereas failure at any one gate can leave old learning substantially intact. It follows that null results are not scientifically empty: with adequate manipulation checks, they help locate the missing transition. The lecture therefore treats failure as information about the chain, not merely as an unsuccessful attempt to reproduce a favored effect.

EVIDENCE ANCHORS

[3] Nader et al. (2000), Nature - Rat - Strong evidence
[24] Sevenster et al. (2012), Neurobiology of Learning and Memory - Healthy adults - Strong evidence
[56] Gershman et al. (2017), eLife - Human and animal conditioning datasets - Strong evidence

Retrieval is an entrance to a decision tree, not an edit command.

4 How to Read the Evidence Labels

EVIDENCE STATUS: METHODOLOGICAL SAFEGUARD.

Established claims are supported across converging methods and are not seriously disputed at the stated level. Strongly supported claims have substantial evidence but retain meaningful boundary conditions. Plausible claims integrate evidence without yet having direct or sufficiently replicated tests. Speculative claims are testable proposals. Unsupported or contradicted claims either exceed available evidence or conflict with decisive findings.

The same discipline governs the dynamical language. A measured molecular or neural process is labeled an empirical mechanism. Attractor and latent-state formalisms may be mathematically motivated interpretations. ARC and Revision Portability are Flow Hijacked explanatory models. Words such as corridor, basin and landscape can also function as metaphors. The category must remain visible because elegance cannot substitute for measurement.

An evidence label applies to the proposition as written, not to every neighboring sentence or to an investigator's reputation. 'Established' may describe the distinction between extinction and erasure while leaving the exact neural implementation unresolved. 'Strongly supported' allows important boundary conditions. 'Plausible' means that several lines converge but decisive tests are missing. 'Speculative' is not a euphemism for weak fact; it names a proposal that must risk failure. 'Unsupported or contradicted' marks a claim that currently outruns or conflicts with the evidence.

The same discipline governs equations and diagrams. A measured quantity may enter an empirical model; a latent variable may organize behavior without being directly observed; and a Flow Hijacked construct may offer a testable synthesis without possessing independent validation. Every formal page therefore identifies its level. Mathematics can expose assumptions and generate discriminating predictions, but notation does not upgrade a metaphor into a mechanism. The burden remains delayed, reproducible evidence that distinguishes revision of old learning from new learning, retrieval competition and performance.

Evidence status can also change with the level of description. It is established that extinction often preserves recoverable old learning; it is less certain which neural population carries each competing relation in a given human task. A proposition may therefore be strong behaviorally and provisional mechanistically. The labels preserve that granularity so a reliable phenomenon is not weakened by an unsettled mechanism, and an elegant mechanism is not promoted by a reliable phenomenon alone.

EVIDENCE ANCHORS

[31] Elsey et al. (2018), Psychological Bulletin - Human - Strong evidence
[34] Chalkia et al. (2020), Cortex - Healthy adults - Strong evidence
[38] Milton et al. (2023), Brain Research Bulletin - Cross-species - Strong evidence

I

PART I

Remembering Is Not Replay

Remembering reconstructs a usable present state from selected traces, current context, bodily condition and goals. That reconstruction may preserve, transform or divide what came before.

5 Memory as a Reconstructed State

EVIDENCE STATUS: ESTABLISHED - retrieval is constructive and state-dependent.

Remembering does not recover a complete object from a single location. It coordinates partial information distributed across systems for perception, affect, action, meaning and context. The resulting state is shaped by the retrieval cue and by what the organism currently needs to predict or do. This constructive character explains both memory's usefulness and its vulnerability to distortion.

Construction does not imply arbitrary invention. Prior traces constrain the result, and well-learned memories can be remarkably stable. The scientific problem is to identify when construction merely expresses a stable representation and when the act of reconstruction initiates processes that alter future expression.

Reconstruction occurs because a usable memory is assembled across systems rather than replayed from a single archive. Perceptual details, spatial relations, bodily value, temporal order and action relevance can be stored and recovered with different degrees of fidelity. The retrieval cue weights these components according to present goals and context. This architecture permits flexible inference from incomplete information, but it also means that repeated recall can alter accessibility, emphasis and association without necessarily rewriting every stored element.

A rigorous account must therefore distinguish representational content from the state produced at retrieval. The present state is observable through reports, choices, physiology or neural patterns; the enduring representation is inferred from stability and transfer across later tests. If a response changes once, the evidence concerns the state. Stronger claims require delayed persistence, generalization and controls for demand, attention and response selection. Constructive retrieval is established. Storage-level revision in a particular episode remains an additional causal claim.

This distinction matters whenever a later report differs from an earlier one. The difference may reflect altered storage, changed cueing, new inference, response bias or ordinary forgetting. Strong updating studies design later tests that separate these routes, for example by probing source, confidence, intrusions and transfer rather than total recall alone. Reconstruction makes change possible, but it also multiplies the explanations that must be tested before storage revision is inferred.

EVIDENCE ANCHORS

[17] Dudai (2012), Annual Review of Neuroscience - Cross-species - Strong evidence
[20] Hupbach et al. (2007), Learning & Memory - Healthy adults - Moderate evidence
[59] Shohamy et al. (2008), Neuron - Healthy adults - Strong evidence

6 What Exactly Has Been Reactivated?

EVIDENCE STATUS: STRONGLY SUPPORTED - memory components can dissociate.

An event memory can include sensory identity, spatial setting, temporal order, emotional value, bodily state, causal meaning and an action tendency. A reminder may reactivate some of these components more strongly than others. A treatment can therefore change physiological responding without changing explicit expectancy, or alter narrative meaning while cue-triggered action remains.

This component view limits claims about whole memories. Reduced skin conductance is not equivalent to erasure of autobiographical fear. A changed verbal report does not establish that a habitual policy has changed. The causal unit must be specified: which representation, response system or learned relation was reactivated, and which later measure changed?

Component specificity explains why interventions can appear successful under one measure and absent under another. In defensive learning, autonomic arousal, startle, expectancy and avoidance are correlated but not interchangeable. In episodic memory, item identity, source, temporal relation and confidence can diverge. In addiction-related learning, cue value, expected relief, habit and action sequence may respond to different reminders. A reminder that reactivates one relation does not automatically open the others to change.

The target must therefore be stated as a relation: cue predicts outcome, context signals threat, bodily state forecasts relief, or action leads to reward. Experimental designs become stronger when the reminder is selected to reactivate that relation and later tests measure both the intended component and plausible neighbors. Clinically, the same rule prevents overclaiming. A person may acquire a new narrative while the body still anticipates danger, or show reduced arousal while avoidance remains. Partial change is real; it is not evidence that an indivisible memory has been erased.

A component map should therefore precede intervention. Investigators can ask which cue retrieves which outcome, in which context, through which report, physiology or action. The same map can guide clinical observation without pretending to diagnose a trace: what changed in language, what changed in bodily anticipation, what changed in avoidance and what changed only in the protected encounter? Precision at this level protects partial gains from being dismissed and global claims from being inflated.

EVIDENCE ANCHORS

[6] Debiec et al. (2002), Neuron - Rat - Strong evidence
[35] Sinclair et al. (2021), PNAS - Healthy adults - Strong evidence
[58] Yoo et al. (2017), Translational Psychiatry - Rat - Moderate evidence

7 The Six-Stage Revision Chain

EVIDENCE STATUS: STRONGLY SUPPORTED AS AN EXPERIMENTAL DECOMPOSITION.

Retrieval makes information accessible. Reactivation means the established representation contributes causally to current processing. Destabilization is the transition into a state that requires active restabilization. Prediction error is a discrepancy relative to the active model. Updating is incorporation of information or change in the relation among components. Reconsolidation is the subsequent restabilization process.

These stages overlap in time and are rarely observed directly in humans. Nevertheless, preserving the distinctions prevents inference by timing alone. An intervention delivered after a reminder is not automatically a reconsolidation intervention. Evidence is stronger when a design manipulates the reminder, tests a lability boundary and measures delayed, process-specific consequences.

The stages form a causal decomposition, not six boxes that the brain traverses in a visibly ordered procession. Retrieval and reactivation can overlap; prediction error can arise while the representation is being expressed; and updating may begin before molecular restabilization is complete. The decomposition remains useful because different manipulations can affect different links. A reminder manipulation tests reactivation, a proteasome or receptor intervention may test destabilization, and a delayed context-sensitive outcome tests what persisted.

The chain also prevents an error of temporal reasoning. An intervention delivered after retrieval is not evidence of reconsolidation merely because its timing falls inside a proposed window. The effect must depend on the reminder, respect known or newly demonstrated boundary conditions and survive tests that rule out transient performance. In humans, where molecular access is limited, the inference is necessarily convergent rather than direct. The chain is therefore both a mechanistic hypothesis and a checklist of what has not yet been shown.

A conditional transition model

$$S = \{R, A, D, E, U, S\}, \qquad P(Z_{t+1} = j \mid Z_t = i, \mathbf{x}_t) = T_{ij}(\mathbf{x}_t)$$

Z indexes a stage in S: R retrieval; A reactivation; D destabilization; E prediction error; U updating; S restabilization. T is a transition probability conditioned on experimental and organismic state x.

FLOW HIJACKED FORMAL DECOMPOSITION - the stages organize causal tests; they are not directly observed neural states.

EVIDENCE ANCHORS

[3] Nader et al. (2000), Nature - Rat - Strong evidence
[10] Ben Mamou et al. (2006), Nature Neuroscience - Rat - Strong evidence
[11] Tronson et al. (2007), Nature Reviews Neuroscience - Cross-species - Strong evidence

Retrieval -> reactivation -> destabilization -> prediction error -> updating -> restabilization. None of the arrows is guaranteed.

8 Retrieval Is Not Destabilization

EVIDENCE STATUS: ESTABLISHED - animal molecular dissociations and human gating studies.

Ben Mamou and colleagues showed that amygdala protein degradation mechanisms can be required for destabilization without being required for retrieval itself. Prediction-error studies in humans likewise report post-retrieval effects only under particular reminder conditions. The convergent message is that expression can occur while the representation remains comparatively stable.

This is the lecture's first red line. We must not infer lability because a person remembered vividly, became distressed or showed a conditioned response. Those observations confirm activation or expression. They do not reveal whether the molecular and computational conditions for revision were present.

Animal work provides the clearest dissociation. Interfering with molecular processes required for destabilization can leave immediate retrieval intact while protecting the later memory from post-reactivation disruption. Conversely, a reminder can produce conditioned responding without making the memory sensitive to an amnestic intervention. These designs show why visible expression is not a biomarker of lability: the same behavior can arise from a stable or destabilized representation.

Human research faces a harder identification problem. There is no accepted direct marker showing that a particular autobiographical or conditioned memory has entered a restabilization-dependent state. Researchers therefore infer destabilization from reminder dependence, prediction-error boundaries, timing and delayed effects. Each clue is imperfect. A vivid recollection, emotional activation or strong bodily response confirms that something was retrieved; it does not license the conclusion that the target trace became writable.

A future destabilization biomarker would need more than correlation with vivid recall. It should predict reminder-dependent sensitivity to a later intervention, distinguish lability from extinction or arousal and generalize across laboratories without assuming the outcome it is meant to detect. Until such a marker exists, human studies should state that destabilization was inferred. That wording is not excessive caution; it accurately identifies the unobserved causal link.

EVIDENCE ANCHORS

[24] Sevenster et al. (2012), Neurobiology of Learning and Memory - Healthy adults - Strong evidence
[25] Sevenster et al. (2013), Science - Healthy adults - Strong evidence
[38] Milton et al. (2023), Brain Research Bulletin - Cross-species - Strong evidence

9 The Dangerous Sentence: Every Recall Rewrites Memory

EVIDENCE STATUS: UNSUPPORTED OR CONTRADICTED WHEN STATED UNIVERSALLY.

The sentence is memorable because it joins a true premise to an unjustified conclusion. Retrieval is reconstructive, and some reactivated memories become labile. It does not follow that every retrieval modifies the stored representation. Rehearsal can strengthen memory. Brief reminders can preserve it. Repeated retrieval can improve access without altering core content.

A responsible public explanation should say that remembering can change what is remembered under some conditions. It should immediately add that the direction, level and mechanism of change cannot be inferred from recall alone. The language of universal rewriting invites neuro-hype and can create false certainty in clinical settings.

The universal sentence also confuses change in a retrieval episode with change in the enduring representation. Recall always occurs in a current state and therefore cannot be a perfect duplicate of an earlier event. That phenomenological difference does not imply storage-level alteration. Repeated testing can strengthen access, reminders can preserve or reconsolidate prior content, and rehearsal can increase confidence even when accuracy does not improve. 'Different now' and 'rewritten for later' are separate propositions.

Public language matters because the claim can influence clinical expectations and autobiographical trust. If every recollection is portrayed as wholesale rewriting, people may be encouraged to treat vividness as proof, uncertainty as damage or therapist-guided meaning as recovered fact. A safer formulation is exact and still hopeful: remembering can modify later memory under particular biological, informational and contextual conditions. The direction, component and durability of that modification must be measured rather than inferred from the intensity of recall.

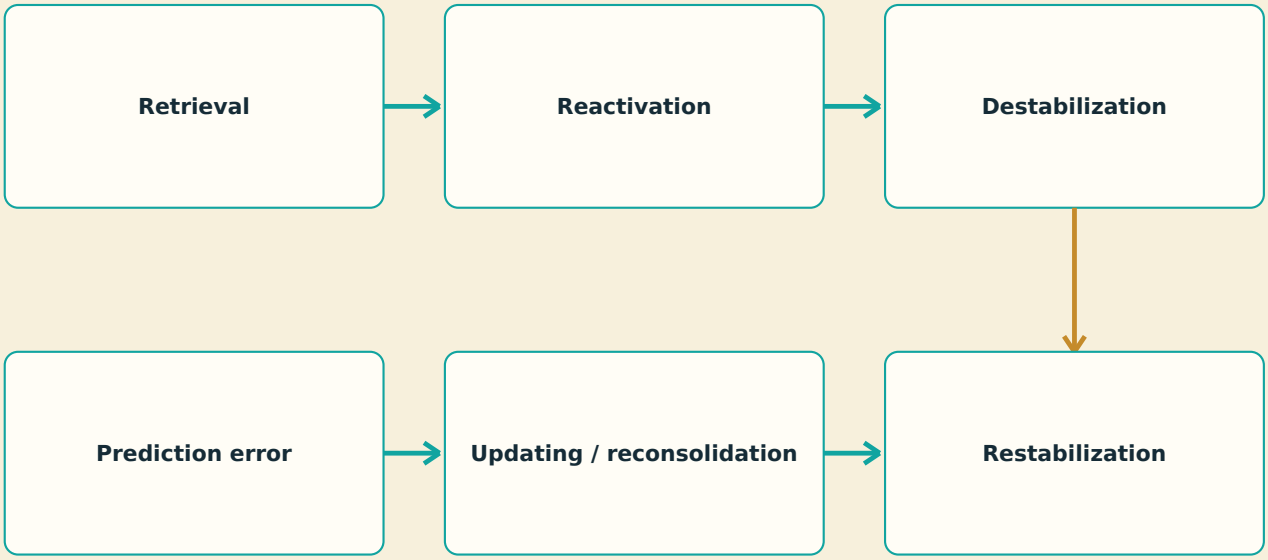
The safer sentence also protects autobiographical dignity. Memory can be fallible and constructive without becoming infinitely manipulable, and uncertainty does not make a person's history available for someone else to author. In educational or therapeutic settings, interpretations should be offered as hypotheses about meaning and present organization, not as proof of unverified past events. Scientific humility here is also an ethical safeguard.

EVIDENCE ANCHORS

[4] Sara (2000), Learning & Memory - Cross-species - Moderate evidence
[17] Dudai (2012), Annual Review of Neuroscience - Cross-species - Strong evidence
[31] Elsey et al. (2018), Psychological Bulletin - Human - Strong evidence

Retrieval Is Not Destabilization

Six stages, five gates, no automatic rewrite.



Every arrow is conditional. Expression of a memory does not reveal which gate has opened.

II

PART II

What Opens a Memory for Change?

Lability is gated. Reactivation must meet biological, temporal and informational conditions before an established representation becomes dependent on restabilization.

10 The Molecular Return to Instability

EVIDENCE STATUS: ESTABLISHED IN MULTIPLE ANIMAL PARADIGMS; translation to humans is indirect.

The modern reconsolidation lineage began when post-reactivation interventions again produced amnesia for established learning. Nader, Schafe and LeDoux showed that protein-synthesis inhibition in the amygdala after retrieval could disrupt a consolidated fear memory. Later work implicated NMDA receptors, ubiquitin-proteasome activity and other signaling pathways in destabilization and restabilization.

These findings establish that reactivated memories can re-enter a biologically vulnerable state. They do not show that all memories do so, or that the same intervention has the same meaning across species and memory systems. Pharmacological amnesia can also reflect performance, toxicity or retrieval effects unless carefully controlled.

The modern causal lineage begins with an asymmetry: a consolidated memory can be expressed normally at reminder and yet become vulnerable to an intervention delivered afterward. Nader, Schafe and LeDoux demonstrated this in amygdala-dependent fear conditioning with protein-synthesis inhibition. Later work implicated NMDA-receptor signaling, ubiquitin-proteasome activity, AMPA-receptor endocytosis and additional plasticity processes. The precise molecular requirements vary by task, region, memory age and intervention.

These experiments establish a biological possibility, not a universal editing rule. Pharmacological amnesia can reflect toxicity, impaired expression or altered retrieval unless reminder-free and timing controls are included. Protein-synthesis inhibitors are also not clinical analogues of ordinary psychotherapy. Their scientific value is narrower and stronger: after qualifying reactivation, persistence can again depend on active cellular work. Reconsolidation names that dependence; it does not specify in advance whether the resulting memory will be weakened, strengthened or updated.

Mechanistic confidence depends on converging dissociations. An intervention should impair later persistence only after the relevant reminder, at an appropriate delay, while sparing immediate expression and comparison memories when possible. Rescue experiments and pathway-specific manipulations further strengthen inference. No single molecular signature is expected to be universal, because different memory systems can recruit overlapping but nonidentical routes into and out of lability.

EVIDENCE ANCHORS

[3] Nader et al. (2000), Nature - Rat - Strong evidence
[11] Tronson et al. (2007), Nature Reviews Neuroscience - Cross-species - Strong evidence
[19] Carneiro et al. (2023), Neuroscience & Biobehavioral Reviews - Primarily rodents; multiple memory tasks - Strong evidence

11 What Actually Destabilizes a Retrieved Memory?

EVIDENCE STATUS: STRONGLY SUPPORTED AS MULTIFACTORIAL; no universal trigger identified.

Destabilization appears to require more than activation. Relevant conditions include the structure and duration of the reminder, expectancy violation, the degree to which the old memory controls current prediction, memory strength and the availability of molecular plasticity. Different paradigms may use different routes into lability.

Prediction error is one influential gate, not a magic switch. A reminder can be too brief to establish the relevant mismatch, or so extended that extinction learning dominates. A strong memory may require a stronger or differently structured reminder. The practical implication is methodological humility: the lability state must be inferred through converging manipulations rather than assumed from one session design.

No single trigger is sufficient across paradigms. Destabilization appears to depend on a conjunction: enough of the target relation must be reactivated; the reminder must reveal information that the current representation fails to explain; molecular plasticity gates must be available; and the event must occur within timing and state conditions that permit restabilization-dependent change. Different tasks implement these ingredients differently, which is why reminder duration or omission can open lability in one protocol and produce extinction or no effect in another.

The cleanest future evidence would manipulate these candidate gates independently and measure a process-specific consequence. For example, the same physical discrepancy could be presented while contextual responsibility, memory strength or precision is varied. If only some combinations create reminder-dependent sensitivity to a later intervention, destabilization would be better identified than by behavior alone. Until such designs converge, the responsible answer remains plural: retrieved memories are destabilized by task-specific interactions among reactivation, mismatch, prior learning and biological state.

The central methodological problem is circularity. Researchers often conclude that a reminder destabilized memory because a later intervention worked, then explain a failed intervention by saying destabilization did not occur. Progress requires prospective indicators and factorial designs that vary reminder content, causal assignment and plasticity gates before outcome is known. A theory of destabilization must predict both vulnerability and protection under prespecified conditions.

EVIDENCE ANCHORS

[9] Pedreira et al. (2004), Learning & Memory - Crab (Chasmagnathus) - Strong evidence
[10] Ben Mamou et al. (2006), Nature Neuroscience - Rat - Strong evidence
[16] Rao-Ruiz et al. (2011), Nature Neuroscience - Mouse - Strong evidence

12 Novelty, Surprise and Prediction Error

EVIDENCE STATUS: STRONGLY SUPPORTED AS RELATED BUT NON-IDENTICAL CONSTRUCTS.

Novelty describes unfamiliarity. Surprise describes an observation with low expected probability. Prediction error describes the difference between an expected and observed quantity under a particular model. A stimulus can be novel without contradicting the active memory, and it can be surprising because the wrong latent cause was assumed.

For reconsolidation, the scientifically important mismatch is not merely physical difference. It must be relevant to the prediction generated by the reactivated representation. Pedreira and colleagues, Sevenster and colleagues, Díaz-Mataix and later work each support prediction-error gating, but their tasks operationalize error differently. The lecture should therefore speak of prediction errors in the plural.

Novelty concerns familiarity; surprise concerns low prior probability; prediction error concerns the difference between an expected and observed quantity under a particular model. They can coincide but need not. A novel stimulus may be irrelevant to the active memory. A familiar event can be surprising because of timing. An omitted outcome can generate prediction error without introducing anything new. Reconsolidation claims become sharper when the manipulated error is tied to the prediction made by the reactivated relation rather than to arousal or novelty in general.

This distinction also prevents a clinical category error. An intense or unusual experience is not necessarily corrective. If the person attributes it to a special therapist, an exceptional protected setting or temporary bodily state, the old prediction may remain untouched. Model-relevant contradiction requires more than difference; it must count as evidence about the expectation currently governing the encounter. Even then, precision and causal assignment determine how strongly the error is used.

What matters experimentally is not the researcher's description of an event as surprising but evidence that the participant or animal held the relevant expectation. Expectancy ratings, learned timing, choice, orienting or model-derived trial estimates can supply that evidence, each imperfectly. Manipulations should also separate surprise from arousal. Otherwise a strong physiological reaction may be mistaken for model-specific information when it merely marks salience.

EVIDENCE ANCHORS

[25] Sevenster et al. (2013), Science - Healthy adults - Strong evidence
[32] Sinclair et al. (2018), Learning & Memory - Healthy adults - Strong evidence
[33] Sinclair et al. (2019), Trends in Neurosciences - Cross-species - Strong evidence

13 Prediction Error Has Several Contents

EVIDENCE STATUS: STRONGLY SUPPORTED IN LEARNING SCIENCE; direct mapping to destabilization remains incomplete.

A reward can differ in value, identity or timing. A feared outcome can fail to occur, occur in another place, arrive from another cause or be accompanied by an unexpected bodily state. These are not interchangeable errors. They recruit different computations and may update different components.

An omission that challenges outcome probability may alter expectancy while leaving sensory identity intact. An identity change may preserve value but update what is expected. A timing error may shift temporal prediction without changing causal belief. ARC must therefore treat mismatch as a vector containing content, magnitude and causal relevance, not as one scalar dose.

A scalar error hides scientifically important structure. Reinforcement learning often begins with a value error, but memory can be wrong about outcome identity, timing, location, agency, causal relation or bodily consequence. Two events with equal physical difference can therefore update different parameters. A missing reward changes expected probability; a different reward changes identity; a delayed reward changes temporal structure. In defensive learning, safety, controllability and source of threat are likewise separable predictions.

ARC must consequently treat discrepancy as a vector whose coordinates acquire different weights in different systems. The vector is not assumed to exist as one measured neural object. It is a formal bookkeeping device that forces the model to state what was contradicted. Experimental claims should report which prediction changed, which response system expressed that change and whether other components transferred. Without that specification, the phrase 'large prediction error' can combine incomparable manipulations and generate a false universal curve.

Prediction error as a vector

$$\boldsymbol{\delta}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t = (\delta_{\text{id}}, \delta_{\text{time}}, \delta_{\text{value}}, \delta_{\text{cause}}, \delta_{\text{body}})_t$$

Observed outcome y is compared with predicted outcome y-hat. The coordinates distinguish identity, timing, value, causal and bodily discrepancy; they need not share one neural code.

MATHEMATICALLY MOTIVATED INTERPRETATION - vector-valued error is established in modelling; this coordinate set is an FH synthesis.

EVIDENCE ANCHORS

[37] Gerlicher et al. (2022), Scientific Reports - Healthy adults - Strong evidence
[38] Milton et al. (2023), Brain Research Bulletin - Cross-species - Strong evidence
[40] Mirea et al. (2024), Journal of Cognitive Neuroscience - Online healthy adults - Strong evidence

14 The Timing Problem

EVIDENCE STATUS: STRONGLY SUPPORTED AS A BOUNDARY CONDITION; exact windows vary.

Reconsolidation effects are often time-dependent. An intervention before reactivation, immediately after it or several hours later can have different consequences. Reminder duration also matters because the system may move from retrieval through lability into substantial new learning. These transitions do not occur at one universal clock time.

Timing is especially treacherous in human translation. A result inside a conventional post-retrieval window does not identify the process. Extinction, attention, source memory and pharmacological state can change on similar timescales. Timing supports a reconsolidation interpretation only when combined with appropriate reminder and boundary-condition controls.

Reconsolidation is often described by a post-retrieval window, but time cannot be interpreted independently of process. A very brief reminder may retrieve without creating meaningful mismatch. Prolonged nonreinforcement may establish extinction learning even if destabilization occurred earlier. Interventions delivered too soon can alter immediate expression; those delivered too late may miss restabilization-dependent processes. The same number of minutes therefore has different meaning across species, tasks, memory strengths and reminder structures.

Strong designs map time rather than choosing one convenient delay. They include reminder-free controls, immediate and delayed tests, and more than one intervention interval. They also distinguish the duration of the reminder from the delay after it. Human protocols often inherit timing from a successful earlier study without verifying that the new population or task crossed the same boundary. Timing is a mechanistic variable only when the underlying stages are independently supported; otherwise it is merely chronology.

Clock time is therefore a protocol variable, not a mechanism. The same interval can contain retrieval, destabilization, new learning and fatigue in different proportions depending on reminder duration and task. Better designs estimate a process transition by crossing time with mismatch and reminder-free controls. Clinically, there is no validated stopwatch that identifies a person's reconsolidation window, and claims of minute-perfect access should be rejected.

EVIDENCE ANCHORS

[8] Suzuki et al. (2004), Journal of Neuroscience - Mouse - Strong evidence
[21] Monfils et al. (2009), Science - Rat - Strong evidence
[27] Sevenster et al. (2014), Learning & Memory - Healthy adults - Strong evidence

15 Strength, Age, Repetition and Emotional Intensity

EVIDENCE STATUS: STRONGLY SUPPORTED FOR STRENGTH; mixed and interactive for age and emotion.

Strong, repeated memories are often harder to destabilize, although the reminder can sometimes be adjusted to recover lability. Chronological age is not a sufficient rule: an older but weak memory may be more revisable than a newer, repeatedly reinforced one. Recent work reinforces the importance of functional strength over age alone.

Emotional intensity can increase rehearsal, arousal and systems consolidation, but it does not make a memory permanently fixed. Nor does vividness prove accuracy or lability. These variables interact with the reminder, current state and measurement system. Their role is best represented as changing the probability and required conditions of revision rather than opening or closing a permanent door.

These properties alter the prior probability that an established representation will remain the preferred explanation. Repetition can deepen cue-outcome associations, broaden retrieval routes and increase systems-level support. Strong memories often require more complete or surprising reminders to destabilize, yet an excessively long reminder may move the task toward extinction. Chronological age matters, but recent evidence suggests that functional strength at reactivation can explain more than elapsed time alone.

Emotional intensity adds further interactions rather than a simple lock. Arousal can strengthen consolidation, narrow attention and promote rehearsal, but it can also change the bodily context in which later retrieval occurs. Vividness is not accuracy, and distress is not proof of lability. The scientifically useful question is how strength, age, repetition and emotion shift the reminder conditions needed to reopen plasticity, the components that become active and the probability that new evidence will be integrated rather than stored separately.

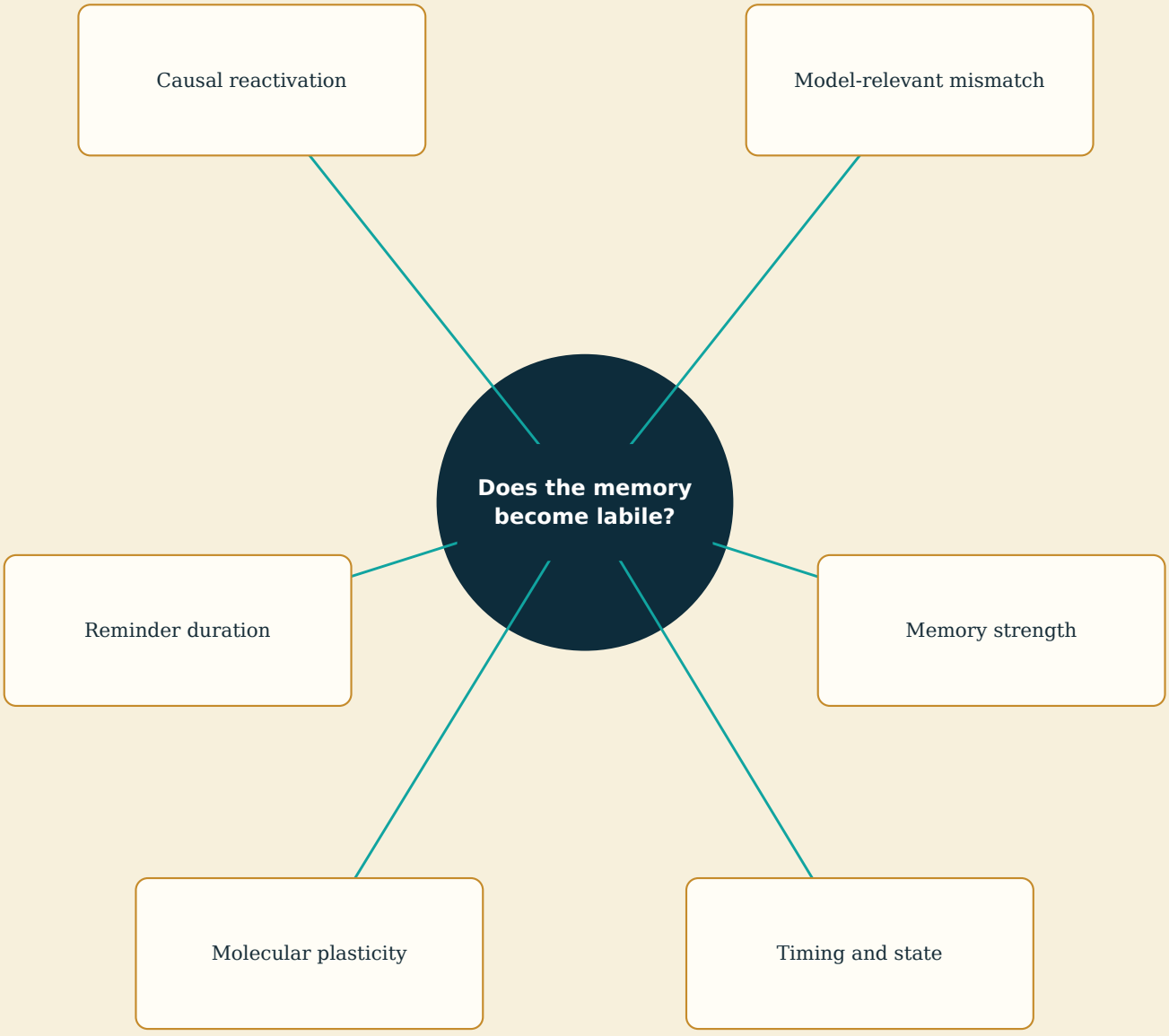
These variables also interact rather than simply add. Repetition can strengthen a relation while broadening the contexts that retrieve it; intense arousal can strengthen some components and fragment others; age can reduce molecular lability while increasing semantic integration. Reporting one global measure of 'memory strength' obscures these routes. Boundary studies should identify whether strength refers to performance, confidence, associative precision, generalization or resistance to interference.

EVIDENCE ANCHORS

[5] Milekic et al. (2002), Neuron - Rat - Moderate evidence
[14] Wang et al. (2009), Nature Neuroscience - Rat - Strong evidence
[43] Yang et al. (2026), Philosophical Transactions of the Royal Society B - Healthy adults - Strong evidence

The Lability Gate

Retrieval enters the gate; several variables decide whether it opens.



No single behavioral sign currently proves destabilization in a human memory.

III

PART III

The Three Fates of Contradiction

Contradiction can leave the old model intact, revise it, or be assigned elsewhere. These are overlapping probabilities, not three universal compartments.

16 Low Mismatch - Confirmation or Restabilization

EVIDENCE STATUS: STRONGLY SUPPORTED AS A COMMON OUTCOME, NOT AN ABSOLUTE RULE.

When experience closely matches expectation, the old representation remains a good explanation of the input. Retrieval may strengthen access, reconfirm the association or produce only minor parameter adjustment. A small discrepancy can still matter if the system is highly precise or if it targets a diagnostic feature, but low mismatch often provides little pressure for structural revision.

This region is important because repeated safe encounters are not automatically corrective. If nothing sufficiently violates the threatening prediction, the absence of harm may be explained away as luck, special protection or an incomplete test. What appears objectively safe may generate little model-relevant error.

When observed experience lies close to prediction, the active model retains high explanatory adequacy. The encounter may strengthen retrieval fluency, reconfirm a relation or produce a small parameter adjustment. Little pressure exists for structural change. This does not mean low error is always inert: a subtle omission can be highly diagnostic when it occurs with high precision, and repeated small errors can accumulate. Magnitude must be understood relative to what the model regards as informative.

The clinical analogue is a safe encounter that the old model can explain away. Safety may be attributed to luck, vigilance, medication, the therapist's protection or an unusually easy version of the situation. The observable outcome is positive, yet the threatening prediction remains conditionally intact. A low-mismatch region is therefore not simply 'nothing happened.' It is a region in which current evidence creates insufficient pressure, ownership or precision to revise the relation that matters.

Low mismatch can still matter by strengthening retrieval fluency or confidence. Repeated reminders that reliably confirm the same outcome may reconsolidate a relation in a more accessible form even when its content barely changes. The predicted outcome is therefore not always stasis. Experiments must distinguish content revision from strengthening, practice and reduced uncertainty, especially when repeated exposure itself changes later performance.

EVIDENCE ANCHORS

[24] Sevenster et al. (2012), Neurobiology of Learning and Memory - Healthy adults - Strong evidence
[27] Sevenster et al. (2014), Learning & Memory - Healthy adults - Strong evidence
[39] Kennedy et al. (2024), eLife - Rat - Strong evidence

17 Model-Relevant Mismatch - A Possibility of Revision

EVIDENCE STATUS: PLAUSIBLE GENERAL PRINCIPLE WITH STRONG SUPPORTING LINEAGES.

Revision becomes plausible when evidence meaningfully contradicts an active prediction and is still interpreted as evidence about the same situation. This combination retains causal responsibility in the old representation while creating pressure to change its parameters or relations. Prediction-error studies, gradual-extinction findings and memory-integration work support parts of this account.

The word possibility matters. Destabilization may fail, the evidence may receive low precision, or a strong prior may absorb the discrepancy without change. The system may also update one component while differentiating another. There is no single observable mismatch level that guarantees modification.

Revision becomes plausible when the event contradicts a prediction strongly enough to demand learning while the old representation remains causally responsible for the evidence. The two requirements pull in opposite directions. Too little error permits confirmation; too much contextual or causal distance can support a new explanation. The useful mismatch is therefore not an intensity target but a relation among contradiction, lability and assignment.

Even inside this region, updating is only one possible fate. Low precision can discount the event; strong priors can absorb it as noise; different components can change in opposite directions; and action can terminate exposure before evidence accumulates. Experiments must therefore measure the old relation directly after delay and under changed conditions. The phrase 'possibility of revision' preserves what the evidence supports: a model-relevant error can open a route to change, but it does not compel the system to take it.

The practical challenge is to demonstrate relevance rather than assume it. A contradiction counts against an old relation only if the learner treats the present cue, outcome and cause as belonging to that relation. Similarity judgments, generalization tests and model comparison can estimate this assignment. The strongest evidence then shows a delayed change in the original prediction under conditions where simple suppression or context-specific new learning should fail.

EVIDENCE ANCHORS

[12] Lee (2008), Nature Neuroscience - Rat - Strong evidence
[39] Kennedy et al. (2024), eLife - Rat - Strong evidence
[42] Chen et al. (2025), Cognition - Healthy adults - Moderate evidence

18 When Discrepancy Goes Somewhere Else

EVIDENCE STATUS: STRONGLY SUPPORTED IN CONTEXTUAL LEARNING AND LATENT-CAUSE MODELS.

A discrepancy can be too contextually distant, too abrupt or too easily explained by another cause to modify the old representation. The system may infer that conditions have changed, that another agent is responsible, or that a new episode should be stored separately. This protects prior knowledge from catastrophic overwriting.

Separation is not failure. In a changing world it is often adaptive to preserve both models. The cost appears when the new model controls behavior only in the revision context while the old one returns elsewhere. Latent-cause inference, pattern separation and renewal research all expose this trade-off between updating and preserving structure.

A large discrepancy can be informative precisely because it does not revise the old model. If the event differs in place, agent, temporal phase, bodily state or causal structure, the learner may infer that another situation generated it. New learning is then stored with that inferred cause while the older relation remains useful for the conditions in which it was acquired. This is not cognitive failure; it is protection against catastrophic overwriting in a changing world.

The cost becomes visible at transfer. The new representation may govern behavior in the training room yet lose retrieval competition elsewhere. An abrupt therapeutic success can therefore coexist with renewal, not because the success was false but because it was narrowly indexed. Latent-cause models formalize this assignment problem, and hippocampal pattern separation offers a complementary circuit-level perspective. ARC must respect both: contradiction can cause separation rather than revision, even when immediate behavior changes strongly.

Separation should be tested as a positive hypothesis, not inferred only from absent updating. Evidence may include preserved old learning alongside a context-bound new response, neural differentiation between episodes, or model-based preference for multiple causes. Each measure has alternatives, but their convergence would show that the system learned from the contradiction while protecting the old relation from direct modification.

EVIDENCE ANCHORS

[46] Gershman et al. (2010), Psychological Review - Human/animal data models - Strong evidence
[56] Gershman et al. (2017), eLife - Human and animal conditioning datasets - Strong evidence
[68] Bein et al. (2023), Neuroscience & Biobehavioral Reviews - Human and animal - Strong evidence

19 Maximum Discrepancy Does Not Necessarily Produce Maximum Revision

EVIDENCE STATUS: STRONGLY SUPPORTED AS A QUALIFIED BOUNDARY CLAIM.

A larger physical error can attract attention and increase learning, yet it can also undermine assignment to the old model. Gradual extinction sometimes preserves the inferred old cause long enough for new evidence to modify its expected outcome, whereas abrupt extinction can look like a new state of the world. Kennedy and colleagues add important recent pressure against treating error magnitude as a monotonic dose.

The principle must remain qualified. Some large errors do produce large updates, especially when context and causal identity remain stable. The defensible claim is only that error magnitude is insufficient. Revision depends on what the error is about and which representation receives responsibility for explaining it.

The key empirical support is comparative rather than a universal dose-response curve. Gradual extinction can sometimes produce more persistent attenuation than abrupt omission, consistent with small errors preserving assignment to the old cause while accumulating change. Abrupt extinction can instead signal a new phase. Other paradigms produce different boundaries, and some large errors do update prior knowledge. The defensible proposition is therefore negative: maximal physical discrepancy is not guaranteed to maximize modification of the old representation.

Three distinctions protect that proposition from becoming another slogan. Subjective error may differ from physical error. Causal responsibility may change while magnitude remains constant. And the outcome of interest must be old-model revision rather than immediate response reduction. A rigorous test independently manipulates discrepancy and contextual assignment, then measures delayed change in the original relation. If revision rises monotonically in every such condition, the nonlinear claim weakens; if assignment changes the curve, a simple dose model fails.

Competing memory fates, not a universal dose curve

$$P(O=k\mid \mathbf{x})=\frac{\exp[g_k(\mathbf{x})]}{\sum_j \exp[g_j(\mathbf{x})]}, \qquad O\in\{L,U,N,P\}$$

O denotes little change L, updating U, new or separated learning N, or performance-only change P. Each score g depends jointly on discrepancy, context, strength, time and precision.

FLOW HIJACKED EXPLANATORY MODEL - a falsifiable alternative to treating mismatch magnitude as a single sufficient variable.

EVIDENCE ANCHORS

[39] Kennedy et al. (2024), eLife - Rat - Strong evidence
[51] Gershman et al. (2013), Frontiers in Behavioral Neuroscience - Rat - Strong evidence
[56] Gershman et al. (2017), eLife - Human and animal conditioning datasets - Strong evidence

The brain does not update a discrepancy. It updates a model held responsible for that discrepancy.

20 Why There Is No Universal Inverted-U

EVIDENCE STATUS: UNSUPPORTED AS A UNIVERSAL LAW.

The proposed three-zone picture is attractive, but different literatures place different quantities on the horizontal axis. Reconsolidation studies manipulate reminder mismatch, latent-cause models estimate structural assignment, and nonmonotonic-plasticity theory concerns coactivation and competition. Moderate coactivation can produce representational differentiation rather than integration.

A single inverted-U would hide these distinctions. The more defensible representation is a multidimensional probability field whose boundaries move with memory strength, context, precision, timing and the component being measured. The simple curve can be used as an introductory question, never as the settled mechanism.

An inverted-U is attractive because it compresses three possible fates into one picture, but it confuses distinct axes. Prediction error has multiple contents; context controls assignment; memory strength shifts lability; and moderate coactivation can produce differentiation rather than integration under nonmonotonic-plasticity accounts. A single horizontal axis cannot represent all of these variables without hiding the mechanism that moved the boundary.

The more responsible image is a conditional probability field. At any joint state, several outcomes have nonzero probability: little-change restabilization, modification of old learning, separate learning, or performance change. Their relative likelihoods vary with the task and component measured. ARC may use a corridor image to organize this field, but the image must never be presented as an empirically fitted universal curve. Its purpose is to generate experiments that estimate moving boundaries, not to decorate a conclusion already assumed.

The lecture will therefore resist drawing one smooth curve through heterogeneous studies. A valid response surface would require common definitions of subjective mismatch, old-cause assignment and revision, plus enough observations to estimate their interactions. Current evidence is too fragmented for that claim. The appropriate scientific image is a family of task-specific surfaces whose shapes may share principles without sharing thresholds or even the same ordering of outcomes.

EVIDENCE ANCHORS

[38] Milton et al. (2023), Brain Research Bulletin - Cross-species - Strong evidence
[66] Ritvo et al. (2019), Trends in Cognitive Sciences - Human and model - Strong evidence
[67] McClay et al. (2022), eLife - Healthy adults - Moderate evidence

Mismatch Is Not a Dose

Magnitude matters through relevance, precision, context and causal assignment.

The bands overlap and move with memory strength, timing and task. This is not a validated inverted-U.

IV

PART IV

The Attractor Revision Corridor Under Pressure

ARC survives only as a multidimensional, probabilistic Flow Hijacked synthesis. Its value lies in forcing lability, assignment and competition into the same explanatory frame.

21 Why ARC Is Useful

EVIDENCE STATUS: PLAUSIBLE FLOW HIJACKED SYNTHESIS.

The literature contains strong local concepts but no single accessible frame for the joint decision: is an active representation revised, preserved or bypassed by separate learning? ARC is useful when it prevents prediction error from being treated as a dose and forces attention to whether the old model remains responsible for the new evidence.

Its value is organizational and generative. It brings molecular lability, computational assignment, contextual similarity, precision and memory strength into one falsifiable question. It does not replace the component theories, and it earns no independent truth merely by connecting them.

ARC's potential contribution is coordination. Reconsolidation research asks whether a reactivated trace becomes labile; latent-cause models ask which hidden situation owns the evidence; predictive accounts ask how error and precision alter belief; and dynamical language asks about stability, competition and reachability. These questions are usually studied with different tasks and vocabularies. ARC places them inside one decision: will discrepant evidence revise the active representation, leave it intact or support a competitor?

Coordination is not discovery by itself. ARC earns scientific value only if the joint variables produce discriminating predictions that simpler component models miss. Its public value is more immediate but still bounded: it prevents the intuition that more contradiction must mean more change, and it keeps causal assignment visible. Used well, the corridor is a disciplined question. Used badly, it becomes a new label for every successful intervention. The lecture adopts the first use and rejects the second.

ARC earns its place only by forcing joint measurement. A study that manipulates discrepancy but ignores causal assignment has not located the corridor; a study that changes context without testing lability has not located it either. The construct should improve experimental design by making these omissions visible. If it merely renames the statement that moderate surprise sometimes helps learning, it adds no scientific value and should be retired.

EVIDENCE ANCHORS

[38] Milton et al. (2023), Brain Research Bulletin - Cross-species - Strong evidence
[39] Kennedy et al. (2024), eLife - Rat - Strong evidence
[56] Gershman et al. (2017), eLife - Human and animal conditioning datasets - Strong evidence

22 What Broke the Original ARC

EVIDENCE STATUS: ORIGINAL SCALAR FORM - UNSUPPORTED.

The first ARC formulation implied a corridor between too little and too much prediction error. Hostile review found four problems. Prediction error is multidimensional. Context can redirect learning at the same physical mismatch. Memory systems can display different fates simultaneously. Nonmonotonic-plasticity models do not map cleanly from discrepancy to integration.

ARC therefore cannot be a fixed interval on one axis, a literal region of the brain or a universal clinical window. Preserving the simple form would make the concept elegant but scientifically brittle. The evidence requires a corridor whose geometry is conditional and whose axes are explicitly labeled as explanatory variables.

The original one-dimensional corridor failed hostile review because its apparent clarity depended on suppressed variables. It treated prediction error as a scalar, assumed that one memory had one fate and implied fixed lower and upper thresholds. Yet the same physical discrepancy can be assigned to different causes; one circuit can integrate while another differentiates; and a memory's strength or reminder history can shift the lability boundary. The simple picture could not state which coordinate produced an outcome.

A concept should become less elegant when the evidence demands it. The revised ARC abandons a universal inverted-U and replaces it with a task-dependent region in a multidimensional explanatory state-space. This costs visual simplicity but gains falsifiability. If discrepancy, context, precision, strength and component identity are measured independently, the model can fail in identifiable ways. A corridor whose boundary moves after every result explains nothing; a corridor whose boundary is predicted before the test may become useful.

The failure was productive because it separated geometry from mechanism. A corridor can organize a multivariable region, but it cannot tell us which molecular gate opened, which latent cause won or which component changed. Those questions require independent theories and measures. The revised ARC is therefore subordinate to the mechanisms it coordinates: when one of those theories makes a sharper prediction, its vocabulary should take precedence over the Flow Hijacked synthesis.

EVIDENCE ANCHORS

[34] Chalkia et al. (2020), Cortex - Healthy adults - Strong evidence
[36] Stemerding et al. (2022), Scientific Reports - Healthy adults - Strong evidence
[67] McClay et al. (2022), eLife - Healthy adults - Moderate evidence

23 The Revised ARC Definition

EVIDENCE STATUS: PLAUSIBLE AND FALSIFIABLE FH EXPLANATORY MODEL.

The Attractor Revision Corridor is the hypothesized, task-dependent region of a multidimensional explanatory state-space in which an established representation remains causally active, relevant plasticity conditions permit change, and discrepant evidence is sufficiently credible and contextually related to be assigned to that representation rather than to a separate latent cause, context or competing memory.

Every clause constrains the concept. Causally active excludes a merely mentioned memory. Plasticity conditions exclude retrieval without destabilization. Credible evidence introduces precision. Contextually related introduces assignment. The alternative latent cause preserves the possibility of separate learning. The definition remains a synthesis, not a validated mechanism.

The revised definition contains three necessary but not sufficient conditions. Causal reactivation means the established representation is actually contributing to the present prediction. Revisability means biological and computational gates permit more than transient expression. Continued responsibility means the system still treats the discrepant event as evidence about that representation rather than about a different latent cause. Failure of any condition redirects the outcome away from old-model revision.

The definition is intentionally probabilistic. It does not say that satisfying the conditions guarantees updating or that failing one condition produces a single alternative. Restabilization, strengthening, partial component change, extinction learning and performance effects remain possible. ARC names a region of elevated plausibility for a specific fate, not an anatomical corridor or a moment a therapist can observe directly. Its status must accompany every first use: Flow Hijacked conceptual synthesis, not validated mechanism, measure or treatment procedure.

A necessary-condition region, not a guaranteed rewrite window

$$\mathcal{A}_{\mathrm{ARC}} = \{\mathbf{x} : r(\mathbf{x}) \geq r_0,\ \ell(\mathbf{x}) \geq \ell_0,\ a_{\mathrm{old}}(\mathbf{x}) \geq a_0\}$$

r is target reactivation; ell is experimentally inferred lability; a-old is causal assignment to the old representation. Thresholds are task-specific and currently unvalidated as common measures.

FLOW HIJACKED HYPOTHESIS - the set formalizes ARC's revised definition; it is not an established neuroscientific mechanism.

EVIDENCE ANCHORS

[39] Kennedy et al. (2024), eLife - Rat - Strong evidence
[46] Gershman et al. (2010), Psychological Review - Human/animal data models - Strong evidence
[66] Ritvo et al. (2019), Trends in Cognitive Sciences - Human and model - Strong evidence

ARC asks whether contradiction is still about the old model at a moment when that model can change.

24 The Required Dimensions

EVIDENCE STATUS: MATHEMATICALLY MOTIVATED INTERPRETATION; not an established coordinate system.

A minimally adequate ARC state includes causal reactivation, lability gates, a vector of prediction errors, posterior responsibility of the old latent cause, external and internal contextual similarity, precision, memory strength and age, timing, interoceptive state and available actions. These variables need not be independent.

The framework predicts interactions. A moderate error under high contextual similarity may have a different fate from the same error under a bodily state associated with danger or withdrawal. A strong memory may require a different reminder. An available avoidance action may prevent sustained contact with corrective evidence. The state-space is a research scaffold for factorial experiments, not a score for a person.

At minimum, a future ARC model requires variables for degree of target reactivation, lability, discrepancy content and magnitude, contextual or latent-cause similarity, precision, memory strength, reminder duration, component identity and action availability. These variables live at different explanatory levels. Lability can involve molecular processes; causal assignment is computational; context and action are behavioral and environmental. Combining them is legitimate only when the level of each variable remains explicit.

The model's outputs must also be plural. It should estimate probabilities of little-change restabilization, old-representation update, new-cause learning and retrieval or performance change. Treating only revision as an outcome would force every null or relapse into a residual category and protect the theory from failure. A useful model compares all outcomes on held-out data and asks whether the joint state predicts delayed fate better than established reconsolidation, extinction and latent-cause accounts alone.

The minimum state description proposed for ARC tests

$$\mathbf{x} = (r, \boldsymbol{\delta}, a_{\text{old}}, c, \pi, m, \tau, b)$$

The coordinates represent reactivation, multidimensional error, old-cause assignment, contextual similarity, precision, memory strength, timing and biological state.

FLOW HIJACKED MEASUREMENT PROPOSAL - useful only if future studies operationalize the coordinates independently.

EVIDENCE ANCHORS

[29] Fernández et al. (2016), Neurobiology of Learning and Memory - Healthy adults - Moderate evidence
[37] Gerlicher et al. (2022), Scientific Reports - Healthy adults - Strong evidence
[43] Yang et al. (2026), Philosophical Transactions of the Royal Society B - Healthy adults - Strong evidence

25 Attractors, Basins and Hysteresis - An Evidence Audit

EVIDENCE STATUS: MIXED STATUS: established mathematics, selective empirical support, FH-level application to ARC.

Attractor dynamics provide a formal language for recurrent states, stability and competition. Neural population work supports metastable and recurrent dynamics in several domains. It does not directly demonstrate that a particular autobiographical memory occupies a measurable basin whose geometry is revised during therapy.

Basin depth, perturbation, hysteresis, reachability and bifurcation should therefore be used at three levels. They may describe formal models, motivate experimental predictions or serve as metaphors for lived persistence and abrupt change. ARC belongs to the second and third levels until its variables are operationalized and linked to reproducible neural or behavioral dynamics.

Attractor dynamics are empirically grounded in some neural population systems and mathematically grounded in recurrent-network models. Stable patterns, metastable dwell states, transitions and history dependence provide a legitimate language for how activity can persist or switch. The inferential leap occurs when an autobiographical memory, symptom or therapeutic state is described as a measured basin without an operational state vector, dynamics or perturbation experiment.

Lecture 49 therefore uses a three-level audit. Recurrent neural states documented with population recordings are empirical mechanisms in their specific tasks. Basin depth, hysteresis and bifurcation in explicit equations are mathematically motivated interpretations. The ARC landscape is a Flow Hijacked explanatory model and, when used without formal variables, partly a metaphor. The distinctions do not weaken the dynamical lens. They make it useful by identifying what must be measured before geometry can be claimed rather than imagined.

A generic state-space dynamics equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x};\boldsymbol{\theta},\mathbf{c}) + \boldsymbol{\eta}(t), \qquad \mathbf{x}(t) \to \mathcal{B}_k$$

x is system state; theta learned parameters; c context; eta perturbation or noise; B-k a basin of attraction. This notation does not identify a particular memory circuit.

MATHEMATICALLY MOTIVATED INTERPRETATION - attractor dynamics are established mathematics; their ARC use is explanatory unless directly fitted.

EVIDENCE ANCHORS

[47] Fanselow (2010), Trends in Cognitive Sciences - Primarily rodent fear - Strong evidence
[77] Osan et al. (2011), PLoS ONE - Simulated attractor network - Plausible evidence
[78] Spalla et al. (2021), eLife - Simulated neural networks - Plausible evidence

26 How ARC Could Be Falsified

EVIDENCE STATUS: NEW HYPOTHESIS - falsification conditions specified.

ARC would lose value if mismatch magnitude alone predicted revision independently of context, lability and memory properties; if contextually distinct evidence reliably modified the old representation more than matched evidence; or if latent-cause assignment failed to predict updating versus separate learning. It would also fail if its dimensions added no predictive value beyond established reconsolidation and latent-cause models.

The strongest test is a preregistered factorial design that independently manipulates memory strength, reminder structure, discrepancy type, external and bodily context, precision cues and action affordances. Delayed measures must distinguish trace modification, competing learning and performance. A concept that cannot survive this discrimination should be revised or abandoned.

The strongest falsification is incremental. ARC should lose scientific value if its variables do not predict delayed old-model revision beyond simpler models. It also fails if contextual responsibility does not alter the relationship between discrepancy and fate, if memory strength and lability add no explanatory constraint, or if the same ARC classification is assigned post hoc to mutually incompatible outcomes. A framework that can describe every result after it occurs has not made a risky prediction.

A decisive program would preregister competing monotonic, inverted-U, latent-cause and multidimensional models; manipulate discrepancy and context orthogonally; and test multiple memory strengths and response components after delay. Neural or representational measures should ask whether the prior pattern changed or a new pattern became retrievable. The result must be evaluated on held-out participants or experiments. If established models predict equally well with fewer assumptions, ARC should be reformulated or abandoned rather than protected by metaphor.

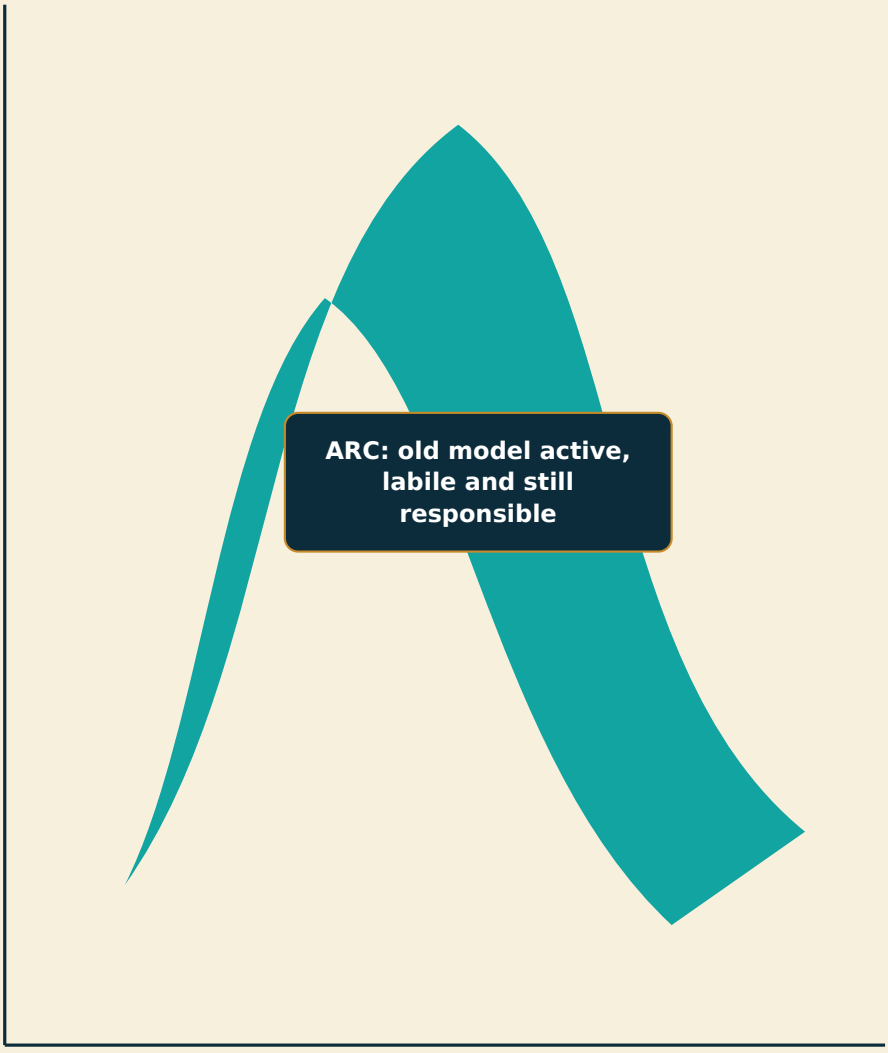
Falsification also requires prospective commitment. The relevant outcomes, delay, target component and alternatives must be specified before results are observed. Post hoc relabeling of every failure as 'outside the corridor' would make ARC immune and therefore scientifically empty. A credible programme must publish boundary estimates with uncertainty, include failed predictions and allow the possibility that simpler latent-cause or associative models explain all incremental variance.

EVIDENCE ANCHORS

[34] Chalkia et al. (2020), Cortex - Healthy adults - Strong evidence
[36] Stemerding et al. (2022), Scientific Reports - Healthy adults - Strong evidence
[39] Kennedy et al. (2024), eLife - Rat - Strong evidence

The Attractor Revision Corridor

A task-dependent region in explanatory state-space - not a neural location.



Causal reactivation

Contextual assignment

Precision / uncertainty

Memory properties

FLOW HIJACKED SYNTHESIS - no validated neural corridor, biomarker, score or treatment protocol.

V

PART V

When New Experience Goes Somewhere Else

Behavior can improve without erasing the old association. Extinction, latent-cause inference, context segmentation and representational differentiation preserve competing explanations.

27 Reconsolidation Versus Extinction

EVIDENCE STATUS: ESTABLISHED DISTINCTION; process boundary is experimentally difficult.

Reconsolidation concerns post-reactivation processing of an existing memory under lability. Extinction involves learning that the expected outcome no longer follows the cue under current conditions. The extinction memory can suppress responding while leaving the original association available. Reminder length and the amount of unreinforced exposure can move a task toward one process or the other.

The distinction is not defined by a clock alone. A brief reminder may fail to destabilize, and a longer session may include both updating and new inhibitory learning. Process claims require designs that test return of behavior, boundary conditions and response-system dissociations rather than relying on immediate reduction.

Reconsolidation and extinction can be engaged by similar reminders but make different claims about what persists. Reconsolidation concerns the restabilization of a reactivated representation after lability, potentially with altered content or strength. Extinction ordinarily creates new learning that the cue no longer predicts the outcome under current conditions. The original association can remain available, which is why responding may return after context change, passage of time or outcome re-exposure.

Duration and structure of the reminder can bias the balance, but no fixed timing rule applies across tasks. A brief mismatch may retrieve and destabilize; extended nonreinforcement may train extinction; intermediate procedures can engage both. Behavioral attenuation alone cannot choose among them. Designs need reminder-free controls, delayed tests, return-of-fear probes and, where possible, molecular or representational dissociations. The most accurate conclusion may be coexistence: one component is updated while a competing inhibitory relation is learned elsewhere.

The two processes can also coexist within one session. Early reminder trials may destabilize some components while longer nonreinforced exposure builds inhibitory learning or changes latent-cause assignment. A binary label for the entire procedure can therefore mislead. Trial-resolved measures and factorial manipulations are needed to identify which component entered which process, and later relapse tests remain essential because immediate response reduction is compatible with both accounts.

EVIDENCE ANCHORS

[8] Suzuki et al. (2004), Journal of Neuroscience - Mouse - Strong evidence
[21] Monfils et al. (2009), Science - Rat - Strong evidence
[44] Bouton (2002), Biological Psychiatry - Primarily animal associative learning - Established evidence

28 Latent-Cause Inference - What Situation Am I In?

EVIDENCE STATUS: STRONGLY SUPPORTED COMPUTATIONALLY AND BEHAVIORALLY.

Observed events are often generated by hidden causes: a safe clinic, a dangerous street, intoxication, withdrawal, a particular relationship or a new phase of an experiment. Latent-cause models formalize how learners cluster observations into hidden situations. If new evidence is assigned to the old cause, its parameters can update. If a new cause is inferred, learning is stored separately.

This framework explains why abrupt change and context shifts can protect the old memory from revision. It also supplies a nearest established concept to ARC's assignment component. Flow Hijacked should not rename latent-cause inference as a novel 'prediction-error ownership gate'; the established terminology already captures that computation.

Learners do not only estimate associations; they infer which hidden situation generated the observations. A latent cause can represent a room, experimental phase, social relationship, intoxication state or cluster of cues that tends to recur together. New evidence assigned to the current cause updates its parameters. Evidence assigned to a new cause creates parallel learning. This formal distinction explains how two incompatible predictions can both be rationally preserved.

Assignment depends on temporal proximity, contextual similarity, cue configuration and prior beliefs about environmental change. Abrupt transitions favor segmentation; gradual change can keep evidence within the old cause. Yet these are probabilistic tendencies, not laws. Latent-cause inference is the nearest established computational concept to ARC's responsibility condition. Flow Hijacked should use the established term and test whether lability adds explanatory power, rather than renaming structural inference as a novel ownership gate.

Bayesian assignment to a latent cause

$$P(z_t = k \mid \mathbf{x}_{1:t}) \propto P(\mathbf{x}_t \mid z_t = k)\, P(z_t = k \mid \mathbf{x}_{1:t-1})$$

z-t is an inferred hidden cause. Current evidence is combined with prior temporal and contextual belief about which cause remains active.

ESTABLISHED COMPUTATIONAL FORM - application to extinction and reconsolidation is influential and testable, but the latent cause is inferred rather than directly observed.

EVIDENCE ANCHORS

[46] Gershman et al. (2010), Psychological Review - Human/animal data models - Strong evidence
[50] Gershman et al. (2012), Learning & Behavior - Associative-learning datasets - Strong evidence
[56] Gershman et al. (2017), eLife - Human and animal conditioning datasets - Strong evidence

29 Context Is More Than a Room

EVIDENCE STATUS: STRONGLY SUPPORTED.

Context includes spatial setting, time, social company, task rules, internal state, drug state, stress, sleep, pain and the available actions. Any of these can help the system infer whether the current event belongs to the old situation. The same room can function as a different context when bodily state or social meaning changes.

This broad definition matters clinically. A person may learn safety while calm, accompanied and unable to avoid, then encounter the cue while exhausted, alone and able to escape. Physical similarity does not guarantee computational continuity. Contextual gating is therefore a joint property of world, body and action.

Context includes any feature that helps infer which model should govern: location, time, people, language, internal state, drug state, task rule and available action. Some features are explicit cues; others alter retrieval without entering awareness. Context can therefore change while the physical room remains constant, and the same room can support different memories across bodily or social states.

This broader definition clarifies why portability cannot be tested by changing location alone. A person may generalize learning across rooms but lose it under fatigue, pain, withdrawal or a familiar interpersonal configuration. Conversely, bodily steadiness may allow a new relation to transfer despite environmental change. Experiments should manipulate contextual dimensions separately and measure which combinations recover old responding. Treating context as a multidimensional inference problem connects associative learning, interoception and lived recurrence without claiming they are one mechanism.

Contextual similarity should accordingly be measured along several axes. Physical overlap, social role, internal state, task rules and available actions may point in different directions. A person can remain in the same room while a change in bodily arousal or interpersonal power signals a different situation; conversely, a portable policy may survive a new room when the relational structure remains stable. Treating context as a single label hides these dissociations.

EVIDENCE ANCHORS

[44] Bouton (2002), Biological Psychiatry - Primarily animal associative learning - Established evidence
[47] Fanselow (2010), Trends in Cognitive Sciences - Primarily rodent fear - Strong evidence
[58] Yoo et al. (2017), Translational Psychiatry - Rat - Moderate evidence

30 The Hippocampal Fork - Integration or Separation

EVIDENCE STATUS: STRONGLY SUPPORTED; different circuits may express different fates.

The hippocampus can complete a prior pattern from partial cues, integrate related events into shared structure, or separate overlapping experiences to reduce interference. Schlichting, Favila, Chanales and related work show that integration and differentiation are not mutually exclusive whole-brain outcomes. Representations can become more similar in one region and more distinct in another.

This circuit heterogeneity complicates the language of 'the memory' being updated. New experience may be integrated into a schematic relation while episodic details are separated, or explicit knowledge may update while defensive responding persists. ARC must be component-specific and system-specific.

The hippocampus supports several computations that can pull memory in different directions. Pattern completion recovers a prior representation from partial cues. Integration links overlapping episodes into a shared structure. Pattern separation and representational differentiation reduce interference by making similar events more distinct. These operations can occur across different subfields, times and task demands; they are not mutually exclusive labels for the whole brain.

Human imaging and behavioral work shows that mnemonic prediction error can shift hippocampal states toward encoding, while overlap and reactivation influence whether new information is integrated or differentiated. A new event may join a schema at one level while retaining distinct episodic detail at another. This heterogeneity is a direct challenge to singular language about 'the memory.' ARC must specify the representational level and region, because one component can be revised while another is separated or completed unchanged.

The fork is not determined by overlap alone. Prior knowledge, temporal proximity, retrieval strength and the degree of coactivation can bias integration, differentiation or pattern separation. Neural similarity must also be interpreted cautiously: increased similarity may reflect shared content, while decreased similarity may reflect differentiation, attention or measurement noise. Strong studies pair representational analyses with behavioral transfer and interference tests to establish what the neural change accomplishes.

EVIDENCE ANCHORS

[60] Bakker et al. (2008), Science - Healthy adults - Strong evidence
[63] Schlichting et al. (2015), Nature Communications - Healthy adults - Strong evidence
[68] Bein et al. (2023), Neuroscience & Biobehavioral Reviews - Human and animal - Strong evidence

31 Generalization and Representational Differentiation

EVIDENCE STATUS: ESTABLISHED PHENOMENA; their balance is task-dependent.

Generalization allows learning to transfer to related cues and contexts. Pattern separation protects distinct episodes from interference. Representational differentiation can actively push overlapping memories apart when competition is moderate. A useful memory system must solve all three problems, not maximize similarity or difference.

Corrective learning can therefore fail in opposite directions. It may be too specific to control behavior outside training, or so broad that it obscures real distinctions and creates unsafe overgeneralization. Good transfer requires calibrated similarity: enough shared structure for retrieval, enough discrimination to preserve meaningful boundaries.

Generalization is not the opposite of differentiation. A system can extract a common relation that transfers across exemplars while preserving distinctive details that prevent interference. Whether behavior generalizes depends on which feature controls prediction and which representation is retrieved. Broad similarity at the level of value can coexist with fine differentiation of context or identity.

This matters for corrective learning. Training many exemplars may widen the range of cues that retrieve safety, but excessive variability can also prevent the target threat model from being reactivated consistently. Generalization requires structured variation: enough commonality to preserve model relevance and enough diversity to reduce dependence on one cue set. The ideal balance depends on the component and the error being learned. A single transfer score cannot reveal whether broad change came from integration, abstraction, multiple memories or response strategy.

Adaptive memory needs both operations. Excessive integration can cause interference and overgeneralized threat; excessive differentiation can make useful learning too narrow to transfer. The optimal balance depends on goals and environmental statistics, not on a universal preference for similarity. This is another reason ARC cannot equate successful change with integration: a system may protect accuracy by separating memories while still generalizing the right relation across relevant situations.

EVIDENCE ANCHORS

[61] Yassa et al. (2011), Trends in Neurosciences - Human and animal - Strong evidence
[64] Favila et al. (2016), Nature Communications - Healthy adults - Strong evidence
[65] Chanales et al. (2017), Current Biology - Healthy adults - Strong evidence

32 Why Extinguished Responses Return

EVIDENCE STATUS: ESTABLISHED.

Renewal is the return of responding when context changes after extinction. Reinstatement follows unsignaled exposure to the original outcome. Spontaneous recovery appears with the passage of time. Rapid reacquisition can also reveal that the original relation remains available. Together these phenomena are strong evidence against routine erasure.

They are also clues about retrieval competition. Extinction learning may govern behavior in the extinction context, while the old association regains control when temporal, spatial or outcome cues favor it. Return does not prove that nothing changed. It shows that the conditions controlling expression have changed.

Renewal, reinstatement and spontaneous recovery reveal different routes by which older learning regains control. Renewal follows a context change that reduces retrieval of extinction. Reinstatement follows unsignaled re-exposure to the outcome, often in a context that restores its relevance. Spontaneous recovery emerges with time as temporal context changes and newer extinction learning loses relative retrieval strength. Rapid reacquisition shows that prior learning can also accelerate relearning.

None of these effects proves that extinction never altered the original relation. They do show that low responding at the end of training is insufficient evidence for erasure. The appropriate tests deliberately perturb context, time and outcome history. In lived settings, return can likewise reveal retrieval competition rather than fraud or total reversal. The distinction is scientifically and morally important: recurrence identifies conditions under which older learning is again reachable, while leaving open what was genuinely learned in the interim.

Return phenomena are therefore diagnostic probes rather than mere disappointments. Renewal tests contextual control; reinstatement tests the effect of unsignaled outcome exposure; spontaneous recovery tests the influence of temporal distance. Their presence supports persistence of older learning but does not by itself identify where it is stored. Their absence is also not proof of erasure unless sufficiently sensitive tests show that recovery cannot be produced under plausible conditions.

EVIDENCE ANCHORS

[44] Bouton (2002), Biological Psychiatry - Primarily animal associative learning - Established evidence
[52] Vervliet et al. (2013), Annual Review of Clinical Psychology - Human and animal - Established evidence
[54] Dunsmoor et al. (2015), Neuron - Human and animal - Strong evidence

33 Why Corrective Experience Can Work Here and Fail There

EVIDENCE STATUS: STRONGLY SUPPORTED AS CONTEXT-DEPENDENT RETRIEVAL; mechanism varies.

A corrective experience can be encoded successfully yet remain difficult to retrieve under stress, intoxication, withdrawal, pain, social threat or familiar cue combinations. It may also be classified as an exception: 'I was safe because the therapist was present,' not 'the prediction was wrong.' Immediate improvement and real-world governance are therefore distinct outcomes.

This distinction opens the lecture's second major conceptual problem. Reconsolidation research asks whether old learning changed. Life asks whether the changed or competing learning will control prediction and action when the situation moves. That later problem motivates Revision Portability in Part VIII.

A corrective encounter is encoded together with the conditions that make it intelligible. If safety occurs only with one person, one ritual or one bodily state, those features can become part of the retrieval key. The old prediction may remain dominant when the key is absent. This is why an intervention can produce authentic within-session change and still fail under displacement. The failure is not necessarily in storage; it may be in access and assignment.

Promoting transfer requires more than repeating the identical success. Variation across context, cue and action can widen retrieval, while explicit retrieval cues can help recover the new relation. Practice should preserve the target prediction long enough to remain model-relevant; otherwise each new environment may be learned as a separate exception. There is no universal recipe, especially across fear, trauma, habit and addiction. The general principle is to test change where it must live, not only where it was easiest to create.

A rigorous account should distinguish failure of acquisition from failure of retrieval. If the corrective relation was never learned, strengthening and clearer contingencies may help. If it was learned but is not retrieved elsewhere, varied-context practice, state variation and retrieval cues become more relevant. If action opportunities differ, environmental redesign may matter more than additional insight. The same visible return can therefore require very different responses.

EVIDENCE ANCHORS

[52] Vervliet et al. (2013), Annual Review of Clinical Psychology - Human and animal - Established evidence
[53] Craske et al. (2014), Behaviour Research and Therapy - Anxiety treatment and laboratory fear - Strong evidence
[57] Rajagopal et al. (2025), Nature Communications - Rodent/human fear data constraints - Moderate evidence

Four Ways Behavior Can Change

The same observed reduction can arise from different memory processes.

Old trace modified

New inhibitory learning

New latent cause / context

Retrieval or performance changed

Delayed tests, context shifts and response-system measures are needed to discriminate among them.

Attenuation is real; its mechanism remains an inference until alternatives are tested.

VI

PART VI

The Predictive, Embodied and Acting Brain

Prediction error exists only relative to a model. Precision decides its influence; bodily state changes the model; action can reduce error without revising belief.

34 Prediction Error Is Model-Relative

EVIDENCE STATUS: ESTABLISHED IN LEARNING AND INFERENCE FRAMEWORKS.

No observation carries a prediction error by itself. Error is the difference between input and expectation under an active model. The same omission can mean that an outcome is less probable, that its timing changed, that the context is different or that another cause is active.

This model dependence changes the interpretation of reconsolidation. A laboratory manipulation specifies physical discrepancy; the learner determines computational discrepancy. Measuring expectancy, causal belief and context inference is therefore more informative than assuming that all participants received the same error.

An observation has no prediction error in isolation. Error is defined relative to a model's expected distribution: what outcome, at what time, under which cause and with what uncertainty. The same omitted reward can be a large error if reward was nearly certain, a small error if omission was common, or no error about the old model if the learner infers a new context before the omission occurs. Physical stimulus difference is therefore an unreliable proxy for computational discrepancy.

This relativity makes subjective expectation scientifically relevant without making it infallible. Trials can measure explicit expectancy, learning history and choice, but the operative generative model may include implicit or bodily predictions that reports miss. Researchers need convergent measures and formal model comparison. For ARC, model relativity means that corridor coordinates cannot be assigned from experimental design alone. The learner's inferred state and uncertainty must enter the explanation.

Model comparison makes this relativity testable. Different candidate models can assign different errors to the same trial because they encode different states, values or causes. Their trial-wise predictions can then be compared against choice, learning rate, physiology or neural signals. No fit is uniquely decisive, but a model that predicts both updating and separation across held-out conditions is more informative than one that labels surprise only after behavior changes.

EVIDENCE ANCHORS

[49] Daw et al. (2011), Neuron - Healthy adults - Strong evidence
[55] Starkweather et al. (2017), Nature Neuroscience - Mouse - Strong evidence
[69] Schultz et al. (1997), Science - Nonhuman primate - Established evidence

35 Precision - Which Error Deserves to Change the Model?

EVIDENCE STATUS: STRONGLY SUPPORTED AS COMPUTATION; indirect for destabilization.

Hierarchical inference weights errors by estimated reliability. A discrepant observation treated as noisy, accidental or untrustworthy may produce little belief change. A smaller observation judged precise can have disproportionate influence. Precision depends on sensory quality, attention, volatility, prior confidence and social credibility.

This helps explain why corrective information can be understood intellectually yet fail to reorganize prediction. The information may be assigned low reliability relative to a strongly precise prior. However, precision weighting has not been established as the molecular trigger of reconsolidation. It is a computational bridge, not a substitute for lability mechanisms.

Precision is the expected reliability of a signal or prediction, often represented as inverse variance. A precise error exerts more influence because the system treats it as diagnostic; an imprecise error is discounted as noise or volatility. Precision can arise from sensory quality, attention, prior certainty, social trust or bodily state. It is not identical to confidence reported after the event, and it is not a single chemical quantity.

Precision complicates the mismatch principle in both directions. A small but highly reliable discrepancy can drive substantial learning, while a dramatic event judged exceptional may leave the prior largely unchanged. In treatment, credibility and controllability may matter as much as intensity. Formal models should therefore use precision-weighted error and test whether it predicts delayed updating beyond raw discrepancy. The concept is well established computationally; its specific mapping onto human reconsolidation remains indirect.

Precision-weighted prediction error

$$\tilde{\delta}_t = \Pi_t(\mathbf{y}_t - \hat{\mathbf{y}}_t), \qquad \Pi_t = \Sigma_t^{-1}$$

Sigma is expected uncertainty and Pi its inverse precision. Equal raw errors can therefore exert unequal influence when their reliability differs.

ESTABLISHED MODELLING PRINCIPLE - its direct role in reconsolidation remains theoretically connected rather than decisively demonstrated.

EVIDENCE ANCHORS

[70] Friston (2010), Nature Reviews Neuroscience - N/A - Plausible evidence
[72] Mathys et al. (2014), Frontiers in Human Neuroscience - Human learning data models - Strong evidence
[76] Friston et al. (2017), Neural Computation - N/A - Plausible evidence

36 Friston - State, Parameter, Structure or Action

EVIDENCE STATUS: THEORETICAL / COMPUTATIONAL CONNECTION; indirect for memory reconsolidation.

Free-energy and active-inference frameworks distinguish several responses to surprising input. The system can revise beliefs about the current state, learn parameters within the same model, infer a different model structure or select actions expected to change future sensory input. These alternatives map closely onto preservation, updating, latent-cause inference and behavioral control.

The connection to ARC is strongest at the level of explanatory architecture. ARC asks when evidence is assigned to an existing model under conditions permitting change. Friston's frameworks formalize precision, policy selection and structure learning. They do not demonstrate a reconsolidation corridor, a destabilization biomarker or a clinical protocol.

Active inference distinguishes several ways to resolve surprising input. State inference changes the estimated current situation while leaving the model intact. Parameter learning changes relations within the model. Structure learning adds or removes hidden causes. Policy selection changes action so that future observations better match preferred outcomes. These are formal alternatives, and a single reduction in surprise can be achieved through more than one of them.

The connection to memory revision is architectural, not a direct experimental demonstration. Reconsolidation asks whether a stored relation becomes labile and restabilizes; active inference asks how beliefs and policies minimize expected free energy under a generative model. Their overlap lies in assignment, precision and levels of change. Friston's work does not validate ARC, identify a destabilization biomarker or prove that psychotherapy changes model structure. It supplies a disciplined vocabulary for alternatives ARC must distinguish.

Variational free energy

$$F[q] = \mathbb{E}_{q(s)}[\ln q(s) - \ln p(o,s)] \geq -\ln p(o)$$

q(s) is an approximate posterior over hidden states; p(o,s) is the generative model of observations and states. Minimizing F can change inference, learning, model structure or action.

ESTABLISHED ACTIVE-INFERENCE FORMALISM - the mapping to memory destabilization is indirect and must not be presented as a reconsolidation mechanism.

EVIDENCE ANCHORS

[70] Friston (2010), Nature Reviews Neuroscience - N/A - Plausible evidence
[74] Pezzulo et al. (2015), Progress in Neurobiology - N/A - Plausible evidence
[76] Friston et al. (2017), Neural Computation - N/A - Plausible evidence

37 Seth - Interoceptive Inference and the Felt Context

EVIDENCE STATUS: THEORETICAL BRIDGE WITH SUPPORTING INTEROCEPTION EVIDENCE.

Seth's predictive account emphasizes that perception is an inference constrained by sensory signals and prior expectations. Interoceptive inference extends this logic to bodily signals that participate in emotion and selfhood. The phrase controlled hallucination can introduce the constructive nature of perception, provided it is not treated as pathology or literal memory rewriting.

For memory revision, the defensible connection is that bodily state helps define what situation the system believes it is in. Elevated arousal, withdrawal or pain can alter the felt similarity between the present and a prior episode, change precision and recover older defensive or appetitive policies. Seth's work does not directly test reconsolidation or ARC.

Interoceptive inference treats bodily signals as part of the evidence from which emotion and self-related experience are constructed. Heart rate, breath, visceral sensation, pain, fatigue and withdrawal do not arrive with fixed meanings; they are interpreted through prior expectations and situational models. Seth's phrase controlled hallucination can clarify constructive perception when carefully explained, but it must not be used to imply pathology, fantasy or unrestricted memory invention.

For Lecture 49, the strongest connection is contextual. Bodily state helps determine what situation the organism believes it occupies and which policy becomes reachable. A safe external room may feel dissimilar from prior safety when arousal or withdrawal changes. That can retrieve older defensive or appetitive learning without disproving previous change. Seth's work does not directly test reconsolidation or ARC. It makes the body's role in assignment and portability theoretically harder to ignore.

Interoceptive error and its precision weighting

$$\varepsilon_{\text{int}} = \mathbf{y}_{\text{body}} - \hat{\mathbf{y}}_{\text{body}}, \qquad \tilde{\varepsilon}_{\text{int}} = \boldsymbol{\Pi}_{\text{int}} \varepsilon_{\text{int}}$$

Observed bodily signals are compared with predicted bodily signals. Their weighted error may contribute to the felt context in which a memory or policy is retrieved.

THEORETICAL BRIDGE - grounded in interoceptive inference, but not direct evidence that bodily error opens a memory reconsolidation window.

EVIDENCE ANCHORS

[71] Seth (2013), Trends in Cognitive Sciences - Human-focused - Plausible evidence
[73] Barrett et al. (2015), Nature Reviews Neuroscience - Human/animal neuroscience - Plausible evidence
[75] Seth et al. (2016), Philosophical Transactions of the Royal Society B - N/A - Plausible evidence

38 Action Can Hide Prediction Error

EVIDENCE STATUS: STRONGLY SUPPORTED AS A GENERAL LEARNING PRINCIPLE.

An organism need not change its belief when it can change the sampled world. Avoidance prevents contact with disconfirming outcomes. Safety behaviors make survival compatible with the belief that danger was present. Substance use can rapidly alter interoceptive signals, apparently confirming the policy that only the substance can restore tolerability.

Active inference therefore adds action affordances to ARC. The availability of escape, checking, concealment or consumption changes which errors are experienced and how they are explained. Corrective experience requires not only an unexpected outcome but enough contact with that outcome for the old prediction to remain accountable.

An organism does not have to revise a prediction when it can act to prevent disconfirming evidence. Avoidance, reassurance seeking, checking, substance use and environmental control can reduce immediate error by changing what is sampled. The prediction then survives because the world encountered under the policy remains compatible with it. Low distress after avoidance is therefore not evidence that threat belief changed.

Active sampling turns memory revision into a closed-loop problem. Beliefs select actions; actions select observations; selected observations update or preserve beliefs. Corrective learning may require safe conditions in which the person can remain in contact with model-relevant evidence without being overwhelmed or forced. The clinical implication is not to remove agency or prohibit protection. It is to measure whether behavior allowed the old prediction to be tested, and to distinguish relief produced by control from learning produced by contradiction.

This is particularly important in avoidance. Successful avoidance prevents the feared outcome and can preserve the belief that avoidance was necessary, because the counterfactual outcome is never observed. A response may therefore look stable despite repeated absence of harm. Experiments and therapies that alter action possibilities can reveal otherwise hidden error, but the ethical and practical costs of removing protection must be weighed carefully.

EVIDENCE ANCHORS

[74] Pezzulo et al. (2015), Progress in Neurobiology - N/A - Plausible evidence
[76] Friston et al. (2017), Neural Computation - N/A - Plausible evidence
[87] Everitt et al. (2016), Annual Review of Psychology - Human and animal addiction - Strong evidence

39 Where the Predictive Bridge Must Stop

EVIDENCE STATUS: IMPORTANT LIMITATION.

Predictive-processing language can redescribe almost any finding after the fact. Its scientific value comes from formal commitments and discriminating predictions, not from universal vocabulary. A study showing prediction error does not thereby establish active inference, and a generative-model account does not identify synaptic destabilization.

Lecture 49 should label Seth and Friston as strong theoretical connections with different emphases: Seth contributes embodied inference and felt context; Friston contributes precision, policies and levels of model change. Neither should be included ceremonially, and neither should be cited as experimental validation of ARC.

Predictive-processing language can redescribe almost any result after the fact, which is precisely why it needs stopping rules. A prediction error in a conditioning task does not by itself establish active inference. Better model fit does not prove a neural representation has the proposed form. A generative account of perception does not identify the molecular gate that destabilizes a memory. Theories at different levels can be compatible without being equivalent.

The bridge is strongest when it yields new contrasts: state inference versus parameter learning, parameter learning versus structure learning, and belief change versus action-based error control. It weakens when terms such as precision or free energy are used as universal explanations without operational measures. Lecture 49 therefore labels Seth and Friston as theoretical connections with distinct contributions. Neither is ceremonial, and neither is experimental evidence for ARC, reconsolidation therapy or a literal attractor corridor.

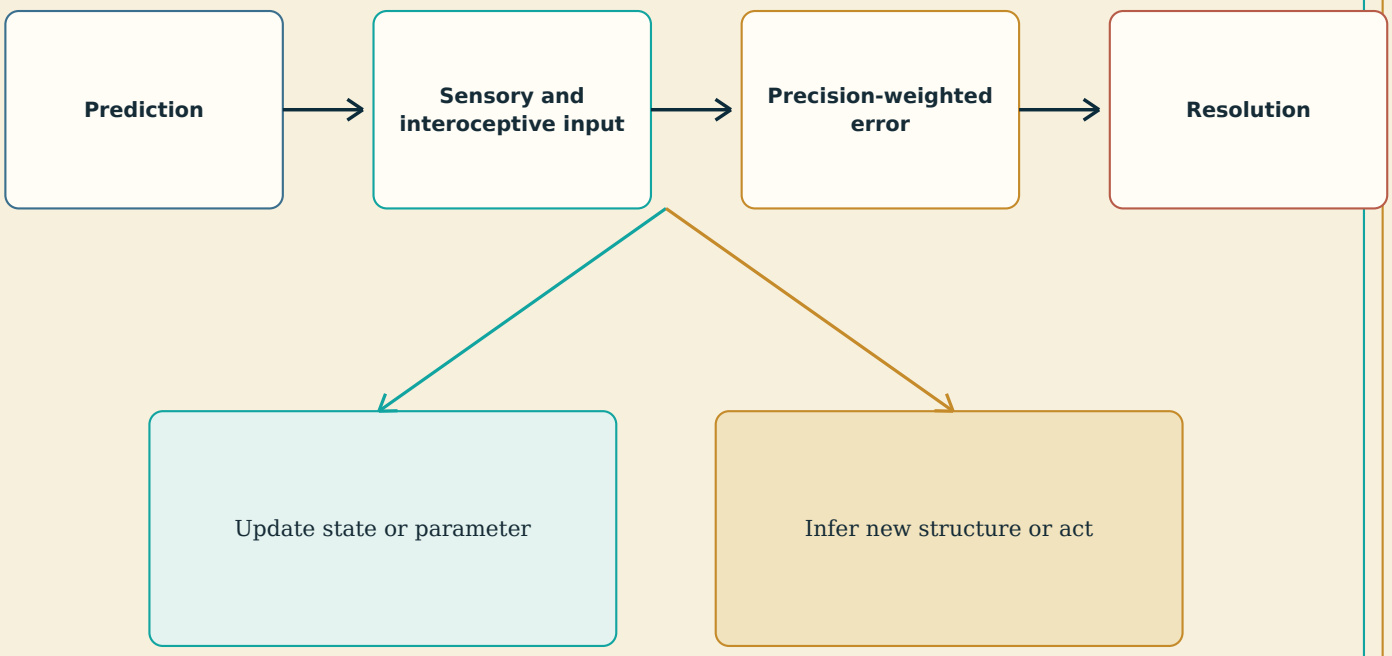
The bridge is most useful when it generates a discriminating manipulation: change precision while holding discrepancy constant, vary action control while preserving sensory input, or alter causal assignment without changing physical context. It is weakest when every outcome is redescribed as successful surprise minimization. A framework that can explain revision, separation, avoidance and no change after the fact still needs prospective predictions to become evidence for this memory problem.

EVIDENCE ANCHORS

[38] Milton et al. (2023), Brain Research Bulletin - Cross-species - Strong evidence
[70] Friston (2010), Nature Reviews Neuroscience - N/A - Plausible evidence
[71] Seth (2013), Trends in Cognitive Sciences - Human-focused - Plausible evidence

Four Ways a Generative Model Can Resolve Error

Perception, learning, structure and action are competing routes.



SETH AND FRISTON: strong theoretical bridges, indirect evidence for reconsolidation, no validation of ARC.

VII

PART VII

Different Memories, Different Constraints

Fear, episodes, rewards and habits do not share one update rule. Their circuits, measures, timescales and clinical meanings must remain distinct.

40 Fear and Defensive Learning

EVIDENCE STATUS: STRONGEST CAUSAL RECONSOLIDATION EVIDENCE; mostly animal and laboratory threat learning.

Conditioned defensive learning provides precise control over cue, outcome, reminder and intervention. This has enabled decisive molecular work in the amygdala and influential retrieval-extinction studies. It also narrows what can be concluded: freezing, startle and skin conductance are components of defense, not complete fear experiences.

Human autobiographical fear includes semantic knowledge, imagery, bodily memory, social meaning and avoidance. A change in one laboratory response system cannot be generalized to erasure of trauma. Defensive-learning paradigms are mechanistic models, not miniature versions of an entire person.

Fear conditioning dominates reconsolidation research because cues, outcomes and timing can be controlled and defensive responses measured repeatedly. Yet freezing, startle, skin conductance, expectancy and avoidance sample different systems. A manipulation that changes one measure may not change another, and an animal's defensive response is not a miniature traumatic memory. The paradigm isolates mechanisms; it does not reproduce the full human phenomenon.

Defensive learning is also adaptive. Preserving an old threat relation under uncertain safety can be costly, but erasing it after one exception could be worse. Extinction and latent-cause separation allow flexible switching across environments. Translation should therefore ask which defensive prediction is targeted and how return is tested. Claims of fear erasure require evidence that the original relation changed across context, time and reinstatement, not merely that responding fell during one session.

Fear measures should not be treated as interchangeable windows onto one state. Expectancy can update while startle persists; autonomic arousal can decline while avoidance remains; verbal knowledge can report safety while threat generalization continues. This response-system divergence is scientifically valuable. It shows which prediction or policy changed and prevents the strongest-looking measure from being selected after the fact as the definition of memory.

EVIDENCE ANCHORS

[21] Monfils et al. (2009), Science - Rat - Strong evidence
[54] Dunsmoor et al. (2015), Neuron - Human and animal - Strong evidence
[94] Meijer et al. (2026), Science Advances - Healthy adults - Moderate–strong evidence

41 Episodic and Declarative Memory

EVIDENCE STATUS: STRONGLY SUPPORTED FOR RECONSTRUCTION AND INTEGRATION; reconsolidation mechanisms vary.

Declarative memories can incorporate misleading details, new relations and updated meanings after retrieval. Human studies of episodic updating, integration and source memory show genuine malleability. Yet these changes can arise from reconstruction, inference, source confusion or new episodic binding without requiring the same molecular process as amygdala fear reconsolidation.

The lecture should therefore distinguish behavioral updating from process identification. A changed story can matter profoundly even when its biological route is uncertain. Scientific humility about mechanism does not make the lived change unreal.

Human episodic updating demonstrates that reminders can make prior events interact with new information. Hupbach and colleagues showed reminder-dependent incorporation into an earlier list, while later work identified roles for overlap, prediction and prior knowledge. These effects support integration, but they can also reflect source confusion or changes in retrieval organization. The later report is not a direct view of the stored trace.

Declarative memory adds dimensions absent from simple conditioning: conscious recollection, semantic structure, narrative meaning, confidence and source. A person can update causal understanding while retaining sensory detail, or integrate a relation while differentiating episodes. Designs should measure intrusion, source memory and transfer separately. Clinically, changed narrative meaning can be consequential without proving that an emotional or habitual component was reconsolidated. The lecture treats these changes as real and component-specific rather than ranking one system as the authentic memory.

Episodic updating is especially sensitive to source confusion. New details can be integrated into an old event, remembered as later additions or misattributed to the original episode. A total-recall score cannot distinguish these fates. Source judgments, temporal order, confidence and intrusion analyses are therefore central, as are controls for testing effects. More remembered information is not automatically a more accurately revised memory.

EVIDENCE ANCHORS

[20] Hupbach et al. (2007), Learning & Memory - Healthy adults - Moderate evidence
[32] Sinclair et al. (2018), Learning & Memory - Healthy adults - Strong evidence
[40] Mirea et al. (2024), Journal of Cognitive Neuroscience - Online healthy adults - Strong evidence

42 Appetitive and Addiction-Related Memories

EVIDENCE STATUS: PROMISING PRECLINICAL EVIDENCE; heterogeneous and limited clinical translation.

Addiction-related learning can involve cue value, outcome expectancy, incentive salience, habits, relief from aversive bodily states, social context and action sequences. Reconsolidation manipulations have altered drug seeking and cue responses in animal models, and several human studies report changes in craving or cue reactivity. The endpoints and protocols remain heterogeneous.

No single 'addiction memory' is being rewritten. A cue association may weaken while a habit, withdrawal expectation or social policy persists. Recovery also depends on alternative reward, relationships, safety, sleep, meaning and reachable action. Memory revision is one process within a larger adaptive system.

Appetitive learning can encode expected value, outcome identity, incentive salience, relief from aversive states and action sequences. Addiction-related memories add extensive repetition, pharmacological states, withdrawal, social context and habits. Post-reactivation interventions have altered cue responding and seeking in animal models, and some human studies report changes in craving or maladaptive reward memory. Outcomes and protocols remain heterogeneous.

The phrase 'addiction memory' is therefore too singular. A cue-value relation may weaken while withdrawal expectancy, habit or social policy remains. A changed laboratory response may fail when the substance, peers, money and familiar action route become available together. Responsible translation places memory updating inside a broader recovery ecology that includes alternative reward, relationships, sleep, meaning and action access. Reconsolidation is one possible process within that system, not an erasure switch for addiction.

Translation is complicated further by the difference between cue memory and the wider ecology of use. A laboratory cue can lose value while stress, withdrawal, social opportunity or habitual action still drive behavior. Conversely, reduced craving during a session can be meaningful without demonstrating modification of a drug memory. Studies should test delayed choice, reinstatement, context shifts and real-world behavior while preserving person-first language and avoiding a single-cause account of recurrence.

EVIDENCE ANCHORS

[79] Lee et al. (2005), Neuron - Rat - Strong evidence
[82] Xue et al. (2012), Science - Rats and abstinent people with heroin use disorder - Moderate–strong evidence
[89] Das et al. (2019), Nature Communications - Hazardous drinkers - Moderate evidence

43 Habits and Action Policies

EVIDENCE STATUS: ESTABLISHED DISSOCIATION BETWEEN KNOWLEDGE AND ACTION CONTROL.

A person can know that an outcome has changed and still execute an overlearned action under familiar cues. Habitual control depends on training history, context, stress and available alternatives. Explicit expectancy and action policy therefore need separate measures.

This dissociation is clinically important. Insight can be authentic without immediately controlling behavior. Conversely, behavior can change through friction or environmental restructuring while the old expectation remains. ARC should not collapse cognitive belief, value and policy into one attractor.

Habitual control explains why explicit knowledge and action can diverge. A person may accurately report that an outcome has lost value and still perform a deeply practiced response when the cue and action route recur. Habit research distinguishes outcome-sensitive goal-directed control from policies supported by repetition and stimulus-response structure, while recognizing that real behavior often mixes both.

Memory revision must therefore measure policy, not only expectancy. Changing a belief about safety, reward or consequence may leave motor chunks and environmental affordances untouched. Conversely, friction, substitution or environmental redesign can change behavior while the old expectation remains. ARC should not compress belief, value and policy into one attractor. Revision Portability must ask whether the revised relation governs action when the familiar opportunity reappears, especially under stress, time pressure or diminished executive control.

Habitual control also changes the intervention target. When behavior is triggered by a learned action sequence, revising an explicit outcome belief may leave the sequence available. Devaluation, contingency degradation and action slips can test whether behavior remains outcome-sensitive. In lived settings, friction, access, routines and alternative actions may alter policy competition more reliably than a powerful insight. That is behavioral engineering, not evidence that memory has been erased.

EVIDENCE ANCHORS

[49] Daw et al. (2011), Neuron - Healthy adults - Strong evidence
[81] Torregrossa et al. (2011), Neurobiology of Learning and Memory - Human and animal addiction - Strong evidence
[87] Everitt et al. (2016), Annual Review of Psychology - Human and animal addiction - Strong evidence

44 Trauma and PTSD - Where Analogy Becomes Dangerous

EVIDENCE STATUS: CLINICAL MECHANISM REMAINS UNSETTLED.

PTSD can involve intrusive episodic memories, generalized threat, avoidance, altered interoception, negative beliefs, sleep disruption and social disconnection. These features interact but are not reducible to one conditioned association. Laboratory threat learning can illuminate components without reproducing the disorder.

Claims that therapy erases a traumatic memory through reconsolidation exceed the evidence. Therapy may change prediction, meaning, avoidance, context discrimination and the capacity to retrieve safety. Multiple mechanisms can coexist. The humane message is not that a hidden neural switch was found, but that persistent responses can change through more than one route.

PTSD can involve intrusive episodic memory, generalized threat, avoidance, altered interoception, negative beliefs, sleep disruption and social disconnection. These features interact but cannot be reduced to one conditioned cue-outcome association. Laboratory fear paradigms illuminate selected learning processes; they do not reproduce trauma, moral injury, chronic threat or the social meaning of what occurred.

Clinical claims should therefore remain mechanism-plural. Psychotherapy may change prediction, meaning, avoidance, context discrimination, bodily tolerance and the ability to retrieve safety. Several processes can coexist, and improvement does not require proof of reconsolidation. Claims that a brief juxtaposition erases a traumatic memory or that subjective intensity identifies a reconsolidation window exceed the evidence. The humane message is not that trauma hides a neural switch; it is that persistent responses can change through more than one supported route.

Traumatic memories are not laboratory conditioned stimuli scaled upward. They may involve repeated events, moral meaning, fragmented source information, ongoing threat, sleep disruption and social consequences. PTSD is also heterogeneous, and symptom change need not track one memory process. Reconsolidation language can guide hypotheses, but treatment claims require randomized clinical evidence, adverse-event monitoring, functional outcomes and comparison with established care rather than mechanistic plausibility alone.

EVIDENCE ANCHORS

[80] Brunet et al. (2008), Journal of Psychiatric Research - Adults with chronic PTSD - Weak–moderate evidence
[88] Brunet et al. (2018), American Journal of Psychiatry - Adults with chronic PTSD - Moderate evidence
[91] Raut et al. (2022), Journal of Psychiatric Research - Adults with PTSD - Strong evidence

45 Where Animal and Human Findings Diverge

EVIDENCE STATUS: ESTABLISHED TRANSLATIONAL GAP.

Animal studies can target brain structures, inhibit protein synthesis and control training with precision. Human studies rely on indirect behavioral and physiological measures, ethically constrained interventions and heterogeneous prior histories. A direct human biomarker of destabilization is absent.

The divergence does not invalidate animal mechanisms. It limits claims about human process identification. Reduced startle, skin conductance, craving or symptom ratings can reflect trace modification, extinction, context, source confusion, response change or nonspecific drug effects. Human work needs converging process-sensitive controls.

Animal studies can control acquisition history, reminder timing and context, intervene in defined circuits and molecular pathways, and test after ethically impossible disruptions. Human studies inherit unknown learning histories, heterogeneous representations and indirect endpoints. A drug or behavioral intervention can alter startle, skin conductance, report or choice without revealing whether a storage trace destabilized. This difference concerns observability, not intelligence or biological relevance.

Translation is strongest when it preserves the causal level. Animal work establishes that retrieval-dependent lability and restabilization mechanisms exist in mammalian memory systems. Human work tests whether analogous boundary-sensitive behavioral effects occur. It cannot simply import the molecular conclusion. Convergence across reminder dependence, timing, multiple measures and delayed return tests raises confidence, while preregistered failures constrain it. The gap is a central scientific result, not an inconvenience to be smoothed over.

The divergence is not a verdict against animal work. Animal experiments allow region-specific causal intervention and tightly controlled learning histories that cannot ethically be reproduced in humans. Human studies add language, explicit knowledge, complex context and clinical relevance while sacrificing mechanistic access. Translation is strongest when the task-level computation and boundary condition are conserved, not when identical behavior is assumed to imply an identical trace mechanism.

EVIDENCE ANCHORS

[21] Monfils et al. (2009), Science - Rat - Strong evidence
[23] Schiller et al. (2010), Nature - Healthy adults - Moderate evidence
[34] Chalkia et al. (2020), Cortex - Healthy adults - Strong evidence

46 The Human Replication Constraint

EVIDENCE STATUS: INCONSISTENT EVIDENCE.

Monfils and colleagues established a retrieval-extinction effect in rats, and Schiller and colleagues reported persistent attenuation of human fear after extinction inside a proposed reconsolidation window. Later close and registered studies by Chalkia, Stemerding, Gerlicher and others did not reliably reproduce the protective effect. Protocol sensitivity remains possible, but reliability is not established.

Propranolol findings are similarly contested. Meta-analytic and trial reports show some physiological or pooled benefit, while symptom outcomes and reanalyses remain disputed. It is scientifically indefensible to present propranolol as an established PTSD reconsolidation treatment or to infer a protocol from this lecture.

The retrieval-extinction literature illustrates the problem. Influential studies reported persistent fear attenuation when extinction followed retrieval inside a proposed reconsolidation interval. Later direct and registered attempts produced mixed or null results. Differences in reminder structure, contingency awareness, exclusion criteria, response measure and analytic flexibility may explain some heterogeneity, but they do not justify selecting only successful protocols.

Propranolol studies show a parallel constraint. Meta-analytic signals for emotional-memory outcomes coexist with null PTSD symptom findings, heterogeneity and disputes about analysis. The drug has effects beyond a single memory process, and a reminder does not prove destabilization. Responsible human claims should specify the endpoint, population and protocol, include uncertainty, and avoid treatment language unless clinical benefit is established. Replication failure is not debris; it identifies where the proposed mechanism lacks reliable control.

Replication should focus on the inferential package, not a single p-value. The reminder dependency, target component, timing, delayed outcome and return tests all matter. Multisite protocols can estimate laboratory and population heterogeneity, while registered reports protect null findings from publication bias. If an effect depends on unmeasured individual thresholds that are invoked only after failure, it is not yet ready to function as a reliable human intervention principle.

EVIDENCE ANCHORS

[34] Chalkia et al. (2020), Cortex - Healthy adults - Strong evidence
[36] Stemerding et al. (2022), Scientific Reports - Healthy adults - Strong evidence
[90] Pigeon et al. (2022), Journal of Psychiatry & Neuroscience - Healthy and clinical human samples - Moderate evidence

47 What Can Responsibly Translate into Psychotherapy?

EVIDENCE STATUS: PLAUSIBLE PRINCIPLES; reconsolidation is not a validated common therapy mechanism.

Therapy can create corrective experiences while an old expectation is active; vary contexts; reduce avoidance; examine explanations that protect the old model; and rehearse retrieval under realistic states. These practices are compatible with inhibitory learning, memory integration, prediction updating and latent-cause accounts.

What cannot be claimed is that a therapist has opened a reconsolidation window or erased a pathogenic memory based on subjective intensity or symptom relief. Lane and colleagues offer an influential theoretical synthesis, not proof that reconsolidation is the common mechanism of psychotherapy. Clinical benefit can be real while mechanism remains plural and uncertain.

The most defensible translation is a set of learning questions, not a reconsolidation protocol. What expectation is active now? Does the experience provide credible evidence about that expectation? Is the contradiction tolerable enough to remain engaged yet strong enough to matter? Which component changed, and can the change be retrieved across contexts, bodily states and actions? These questions are compatible with several therapeutic traditions and do not presume one hidden molecular event.

What cannot translate is certainty that a therapist can see lability, calculate an optimal prediction-error dose or guarantee erasure. Corrective experience should be individualized, safe and ethically grounded; overwhelming a person is neither scientifically required nor humane. Delayed and real-world function matter more than dramatic within-session response. The responsible clinical promise is modest but meaningful: learning principles can help design and test change, while the exact process in a particular person usually remains uncertain.

A responsible translation is framed as design guidance rather than a branded mechanism. Make the relevant expectation explicit; create safe, meaningful disconfirmation; vary contexts and bodily states when appropriate; rehearse alternative action; and test delayed transfer. These practices overlap with established therapies and do not require claiming reconsolidation. When change occurs, the clinician can describe what was learned and where it travels without asserting that a trace was biologically rewritten.

EVIDENCE ANCHORS

[53] Craske et al. (2014), Behaviour Research and Therapy - Anxiety treatment and laboratory fear - Strong evidence
[85] Lane et al. (2015), Behavioral and Brain Sciences - Clinical psychotherapy theory - Plausible evidence
[93] Milton (2023), Translational Psychiatry - Human and animal; addiction/fear/trauma - Strong evidence

Memory Fate by System

Shared principles do not erase mechanistic differences.

System	What is learned	Common measures	Main caveat
Defensive fear	Cue-outcome threat relations	Startle, SCR, expectancy	Not autobiographical trauma
Episodic / declarative	Events, relations and meaning	Recall, recognition, representation	Reconstruction is not one trace
Appetitive / addiction	Cue-value, action and bodily relief	Craving, choice, seeking	Multiple systems and states
Habit / policy	Stimulus-action regularities	Action persistence	Belief change may not change action

A common vocabulary is useful only while system-specific evidence remains visible.

VIII

PART VIII

When Change Must Travel

Revision during one episode is not enough. The revised memory, belief or policy must remain influential as context, bodily state, cues and available actions move.

48 The Second Memory-Change Problem

EVIDENCE STATUS: STRONGLY SUPPORTED DISTINCTION BETWEEN ACQUISITION AND TRANSFER.

The first problem is whether an old representation changes. The second is whether the changed or competing learning continues to govern behavior when conditions depart from the learning episode. Within-session reduction can coexist with renewal, state-dependent return or failure under new affordances.

The field has rich terms for generalization, transfer, renewal and state dependence. What remains conceptually fragmented is the joint post-update deployment problem across world, body, cues and action. Revision Portability is proposed to organize that problem without claiming a new biological mechanism.

The first problem asks whether an established representation changed. The second asks whether changed or competing learning continues to govern behavior when conditions depart from the revision episode. These problems are logically independent. A trace can be meaningfully modified yet weakly retrieved elsewhere; a separate extinction memory can transfer broadly; and behavior can improve through environmental structure even when the old representation remains.

Most experiments emphasize acquisition because it is easier to control. Life emphasizes deployment. A revised prediction must compete under new rooms, times, bodily states, cue combinations and action opportunities. Within-session reduction is therefore an incomplete endpoint. The second problem requires a transfer battery that deliberately displaces the learner from the conditions of success and measures multiple response systems. Revision Portability is proposed to organize that battery, not to replace established concepts of generalization, renewal or state dependence.

Portability also prevents a misleading comparison between interventions tested at unequal distances from training. One procedure may appear superior because it is assessed in the same room and state, while another faces a demanding transfer test. A fair evaluation maps performance across matched displacements and time points. The resulting profile can reveal narrow excellence, broad modest benefit or a sharp failure boundary that an average treatment effect conceals.

EVIDENCE ANCHORS

[52] Vervliet et al. (2013), Annual Review of Clinical Psychology - Human and animal - Established evidence
[57] Rajagopal et al. (2025), Nature Communications - Rodent/human fear data constraints - Moderate evidence
[93] Milton (2023), Translational Psychiatry - Human and animal; addiction/fear/trauma - Strong evidence

49 The Revision Portability Gradient - Definition

EVIDENCE STATUS: PLAUSIBLE FLOW HIJACKED EXPLANATORY MODEL / NEW HYPOTHESIS.

The Revision Portability Gradient is the degree to which a revised memory, belief or action policy continues to govern prediction and behavior as external context, bodily state, cue configuration and available actions diverge from the conditions in which revision occurred.

Portability is a property of the relation between learning and displacement, not a permanent trait of a person. The same revision can travel well across places but poorly across bodily states. Another may survive arousal yet fail when a familiar action route reappears. The word gradient keeps these partial patterns visible.

Portability is graded because displacement is multidimensional and direction-specific. A revised relation may survive a new room but fail under bodily arousal; transfer across cue variants may be strong while transfer to action remains weak. The same individual can therefore show a different portability profile for different memories. Treating portability as one binary success score would erase the pattern the construct is meant to reveal.

The construct applies to the relation between a learning episode and later conditions, not to a fixed personal trait. It should be measured by sampling displacement and estimating the probability that the revised memory, belief or policy governs behavior. Existing terms remain essential: generalization describes response to novel stimuli, transfer concerns performance beyond training, and renewal describes return. The proposed novelty is their joint organization across external context, body, cues and action after meaningful revision.

Revision availability across displacement

$$\rho(\mathbf{d}) = P(C_{\text{rev}} = 1 \mid \mathbf{d})$$

d is displacement from the revision episode; C-rev equals one when revised learning, rather than an older competing policy, governs behavior. Rho is the corresponding conditional probability.

FLOW HIJACKED NOVEL CONSTRUCT - proposed as a measurable outcome surface, not a validated scale or treatment target.

EVIDENCE ANCHORS

[52] Vervliet et al. (2013), Annual Review of Clinical Psychology - Human and animal - Established evidence
[53] Craske et al. (2014), Behaviour Research and Therapy - Anxiety treatment and laboratory fear - Strong evidence
[57] Rajagopal et al. (2025), Nature Communications - Rodent/human fear data constraints - Moderate evidence

ARC asks whether change can form. Revision Portability asks whether that change can travel.

50 The Four Displacement Dimensions

EVIDENCE STATUS: EVIDENCE-BASED INGREDIENTS; unified construct unvalidated.

External displacement includes place, time and social setting. Interoceptive displacement includes stress, arousal, withdrawal, pain, fatigue and drug state. Cue displacement changes sensory composition or the combination of triggers. Affordance displacement changes which actions are possible, easy or rewarded.

These dimensions interact. A social setting can alter bodily state; bodily state can change cue salience; an available action can prevent corrective sampling. A portability profile should therefore be a response surface rather than a single score. The concept predicts anisotropy: transfer will often be stronger in some directions than others.

External context includes place, time and social configuration. Bodily state includes arousal, fatigue, pain, withdrawal and medication state. Cue configuration includes which features appear together and which retrieval cues are absent. Action affordances include the behaviors, substances, technologies, people and environmental routes actually available. Each dimension can alter latent-cause assignment and retrieval competition.

A portability assessment varies these dimensions without treating them as equivalent. Distance on one axis may matter only in combination with another: a familiar cue may be manageable when rested but not when exhausted and alone. The framework therefore predicts interactions rather than one radial decay from the revision episode. It also remains person-first. A portability profile describes conditions under which learning governs behavior; it does not rank the person's strength, sincerity or worth.

A displacement vector for portability tests

$$\mathbf{d} = (d_{\text{context}}, d_{\text{body}}, d_{\text{time}}, d_{\text{action}}), \qquad \rho = \rho(\mathbf{d})$$

The four coordinates measure change in external context, bodily state, temporal position and available action architecture. Interaction terms are expected, not optional.

FLOW HIJACKED MEASUREMENT PROPOSAL - the dimensions synthesize existing context, interoception, retention and policy literatures.

EVIDENCE ANCHORS

[44] Bouton (2002), Biological Psychiatry - Primarily animal associative learning - Established evidence
[58] Yoo et al. (2017), Translational Psychiatry - Rat - Moderate evidence
[75] Seth et al. (2016), Philosophical Transactions of the Royal Society B - N/A - Plausible evidence

51 What Is Actually Novel?

EVIDENCE STATUS: PROVISIONAL NOVELTY AFTER SEMANTIC AUDIT.

The novelty is not that context matters, that learning generalizes imperfectly or that renewal occurs. Those are established. The proposed addition is to treat portability as a graded property of a revision across a joint external-interoceptive-cue-action space, and to distinguish acquisition of change from later governance under displacement.

The concept should be abandoned if an established equivalent is identified that already performs this full function. It should also be merged into existing terminology if it offers no incremental prediction. Naming is justified only if it improves discrimination, measurement and humane explanation.

The novelty audit found substantial prior art for every ingredient. Generalization, transfer, renewal, state-dependent learning, inhibitory learning, ecological validity, latent-cause inference and retrieval competition already describe important parts of the problem. Revision Portability cannot claim to discover that learning is context-sensitive. Its defensible novelty lies in treating post-revision deployment as a graded, multidimensional profile that jointly includes body and action affordances.

That novelty is useful only if it changes measurement. Two interventions with equal immediate effect should be compared across systematic displacement; the framework should predict where their outcomes diverge. Body and action dimensions must add information beyond ordinary stimulus and context generalization. If a standard transfer model captures the data equally well, the new term is unnecessary. Naming creates an obligation to outperform the concepts it organizes, not permission to ignore them.

The novelty claim is therefore conditional and infrastructural. Revision Portability proposes a shared test space in which context, bodily state, cue configuration and action affordance are displaced together or independently after learning. Its success would be shown by incremental prediction, not by rhetorical usefulness. If existing generalization or state-dependent models already recover the same surface with equal parsimony, the new name should be absorbed into those traditions.

EVIDENCE ANCHORS

[57] Rajagopal et al. (2025), Nature Communications - Rodent/human fear data constraints - Moderate evidence
[58] Yoo et al. (2017), Translational Psychiatry - Rat - Moderate evidence
[93] Milton (2023), Translational Psychiatry - Human and animal; addiction/fear/trauma - Strong evidence

52 Predictions and Falsification

EVIDENCE STATUS: NEW HYPOTHESIS WITH SPECIFIED TESTS.

Interventions producing equal immediate change should show different transfer profiles. Training across varied contexts and bodily states may improve portability without increasing within-session reduction. Interoceptive and affordance shifts should explain variance beyond physical context alone. Portability should be recoverable as a reproducible multidimensional surface.

The construct fails if immediate change fully predicts later outcomes, if bodily state and affordances add nothing beyond standard generalization measures, if no stable gradient appears across preregistered transfer batteries, or if the concept cannot distinguish interventions with equal acquisition but different persistence.

The central prediction is interaction: immediate change will not fully determine later behavior, and the loss of control by revised learning will depend on the direction and combination of displacement. Interventions that vary contexts but never vary bodily state should show better external than interoceptive portability. Interventions that change belief without rehearsing action should transfer less well when familiar affordances return. These predictions can be preregistered before outcomes are known.

The construct should be rejected or absorbed into existing terminology if one acquisition score predicts all later tests, if body and affordance variables add no incremental information, or if no stable profile can be recovered across repeated batteries. Falsification also requires measurement reliability: apparent gradients may be noise, unequal task difficulty or regression to the mean. A credible program compares hierarchical portability models with simpler generalization baselines on held-out data and reports null dimensions rather than forcing a complete pattern.

A preregistrable portability model

$$\operatorname{logit} P(Y=1) = \beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{d} + \mathbf{d}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{d}$$

Y indicates control by revised learning at delayed test. Beta estimates main displacement effects; Gamma estimates nonlinearities and interactions that may create cliffs or ridges.

FLOW HIJACKED TEST MODEL - offered to make the Portability Gradient vulnerable to clear empirical failure.

EVIDENCE ANCHORS

[39] Kennedy et al. (2024), eLife - Rat - Strong evidence
[57] Rajagopal et al. (2025), Nature Communications - Rodent/human fear data constraints - Moderate evidence
[67] McClay et al. (2022), eLife - Healthy adults - Moderate evidence

53 A Compassionate Interpretation of Return

EVIDENCE STATUS: FLOW HIJACKED INTERPRETIVE IMPLICATION - not an individual diagnosis.

The return of an old response under stress, withdrawal, isolation or a familiar cue configuration need not mean that prior learning was fraudulent or that the person failed. It may reflect a retrieval competition or a portability boundary: the revised learning did not govern behavior at that distance from the conditions in which it was established.

This interpretation removes blame without removing agency. It directs attention toward the dimensions across which change has not yet traveled and toward environments that can support safer retrieval and action. It remains one possible explanation among several and must not be imposed on an individual without assessment.

Return of fear, craving or an old action pattern is often interpreted globally: the treatment failed, the person did not truly change or the old self was the real one. A portability account offers a narrower explanation. Learning that governed one state may lose retrieval competition after a change in body, cue set or available action. The return is evidence about conditions of control, not a complete measure of what was learned or who the person is.

Compassion here is compatible with precision and responsibility. It does not deny harm, minimize risk or promise that old learning is harmless. It asks which dimension changed and what support or practice would make the revised policy reachable again. This converts shame into a testable question without turning life into a score. A recurrence can call for immediate protection and repair while still being understood as a state-dependent event rather than a permanent verdict.

Compassion also improves causal inquiry. Shame compresses a complex recurrence into a global identity judgment and can hide the conditions that changed. A functional analysis instead asks what cues were present, what state narrowed choice, which action became easiest and what support was absent. This does not excuse harm or remove accountability. It makes accountability actionable by connecting repair to modifiable conditions rather than to condemnation of the whole person.

EVIDENCE ANCHORS

[44] Bouton (2002), Biological Psychiatry - Primarily animal associative learning - Established evidence
[52] Vervliet et al. (2013), Annual Review of Clinical Psychology - Human and animal - Established evidence
[87] Everitt et al. (2016), Annual Review of Psychology - Human and animal addiction - Strong evidence

54 ARC and Revision Portability Must Remain Separate

EVIDENCE STATUS: FLOW HIJACKED CONCEPTUAL ARCHITECTURE.

ARC concerns formation: can discrepant evidence modify the old representation during a particular episode? Revision Portability concerns deployment: will revised learning continue to govern behavior later and elsewhere? An ARC failure produces restabilization, no revision or separate learning. A portability failure occurs after meaningful change but under conditions in which older or competing policies regain control.

Combining the concepts would make both less precise. The distinction generates a two-stage research program: first identify whether revision occurred and by which process; then map the conditions under which its effects remain behaviorally reachable.

ARC concerns formation during a particular episode: can evidence modify an established representation rather than merely confirm it or create a competitor? Revision Portability concerns later deployment: under which displaced conditions does revised learning continue to control prediction and action? A failure of ARC means meaningful old-model revision may never have occurred. A portability failure begins after some consequential change but reveals limits on retrieval or control.

Separating the constructs prevents circular rescue. If behavior returns, ARC cannot simply claim that revision occurred but lacked portability unless formation was independently measured. Conversely, a narrow transfer failure does not prove that the earlier change was only extinction or performance. The two-stage architecture creates a stronger research program: identify what changed and by which process, then map where that change governs behavior. Each stage can fail without redefining the other.

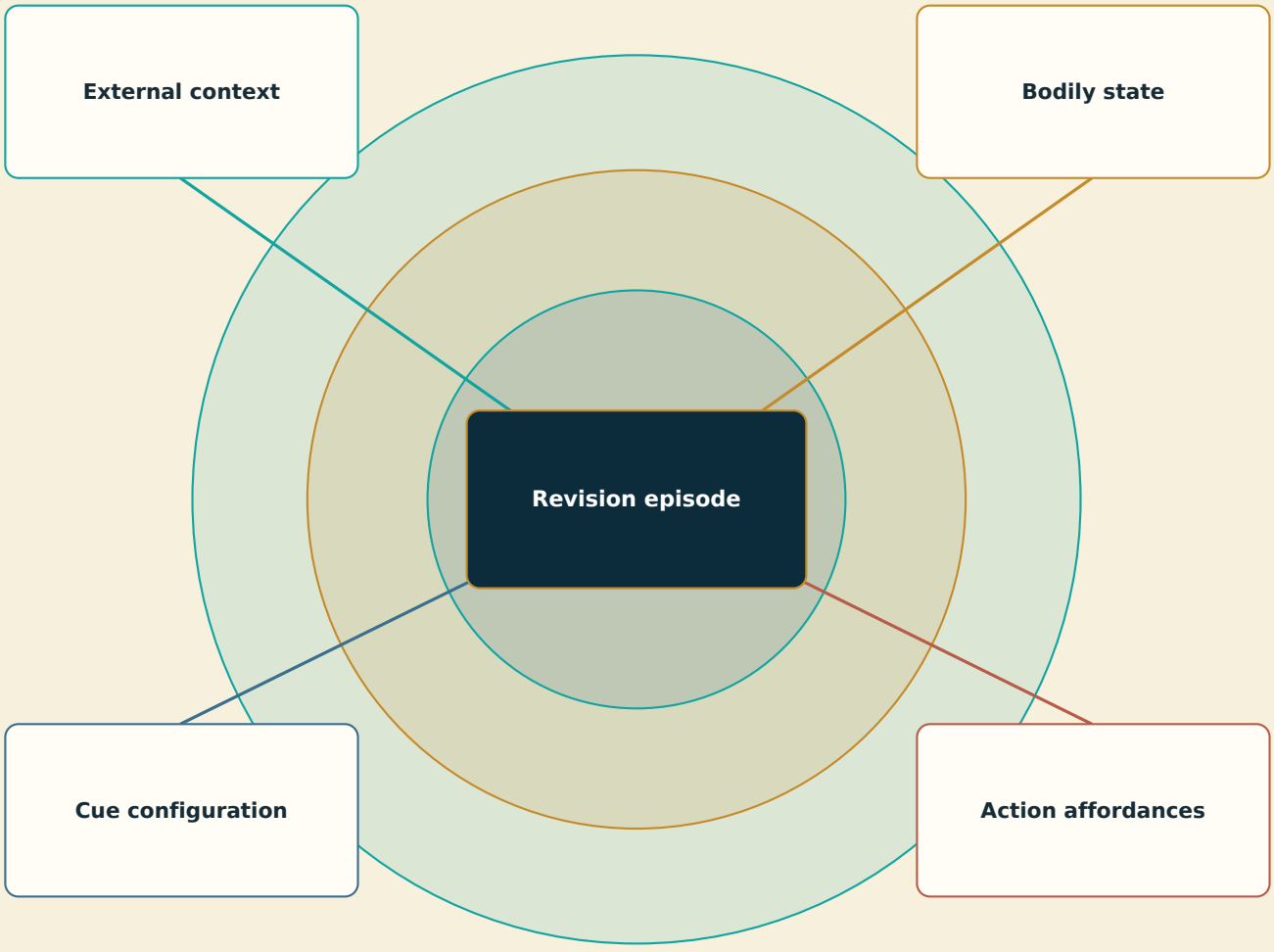
The separation also clarifies null results. A revision can occur inside plausible ARC conditions yet fail to travel, which challenges portability without disproving lability. Conversely, broadly transferable new learning can arise without evidence that the old representation was destabilized, which supports portability of a competitor without supporting ARC. Crossing both constructs in one design prevents success on one axis from being used as retrospective proof of the other.

EVIDENCE ANCHORS

[39] Kennedy et al. (2024), eLife - Rat - Strong evidence
[57] Rajagopal et al. (2025), Nature Communications - Rodent/human fear data constraints - Moderate evidence
[93] Milton (2023), Translational Psychiatry - Human and animal; addiction/fear/trauma - Strong evidence

The Revision Portability Gradient

How far does revised learning continue to govern behavior?



Portability can differ by direction; it is not one binary success score.

FLOW HIJACKED NEW HYPOTHESIS - evidence supports the ingredients, not yet the unified construct.

EPILOGUE

Revision Without Neuro-Hype

The science is powerful precisely where its limits remain visible. Memory can change, old learning can persist, and a person's return to an old response is not a moral verdict.

55 What the Evidence Establishes

EVIDENCE STATUS: ESTABLISHED / STRONGLY SUPPORTED.

Consolidated memories can become labile after reactivation under some conditions and require active restabilization. Retrieval is not sufficient. Prediction error often contributes to destabilization and updating. Memory strength, reminder structure, timing and context alter outcomes. Extinction often creates new, context-sensitive learning rather than erasing the old relation.

Contextual assignment, latent causes, integration, separation and competing response systems explain why the same experience can have different memory fates. Human behavioral updating is real, but a direct biomarker of destabilization and a universally reliable clinical reconsolidation protocol are absent.

The field establishes that consolidated memories can re-enter a lability-dependent state after qualifying reactivation, that active restabilization has identifiable molecular requirements in animal systems, and that reminder conditions influence later memory fate. It also establishes that extinction commonly preserves older learning, that context controls retrieval and return, and that human memory can integrate new information after reminders. These findings converge without requiring one universal mechanism.

Established does not mean complete. The evidence is strongest for controlled animal memories and selected human laboratory tasks. It does not provide a direct human marker of destabilization, a universal mismatch function or a clinical test showing that a particular autobiographical memory was rewritten. The lecture's foundation is therefore conditional plasticity: memory can change after retrieval, but the path from expression to revision must be demonstrated at the relevant level.

The established core is therefore narrower than the popular story and stronger for being narrower. Consolidated memories can re-enter an active, intervention-sensitive process after qualifying retrieval; mismatch often contributes; extinction commonly preserves competing old learning; and context powerfully governs expression. These propositions survive variation better than any universal reminder recipe. The lecture builds outward from them while marking every additional inference.

EVIDENCE ANCHORS

[3] Nader et al. (2000), Nature - Rat - Strong evidence
[24] Sevenster et al. (2012), Neurobiology of Learning and Memory - Healthy adults - Strong evidence
[44] Bouton (2002), Biological Psychiatry - Primarily animal associative learning - Established evidence

56 What Is Strong but Qualified

EVIDENCE STATUS: STRONGLY SUPPORTED BOUNDARY CLAIMS.

Maximum discrepancy does not necessarily produce maximum revision of the old representation. Contextual similarity and causal assignment influence whether evidence updates old learning or supports a new representation. Interoceptive state can gate retrieval and return. Corrective learning must remain retrievable outside its acquisition episode to matter in life.

None of these claims establishes one universal curve or mechanism. They justify a conditional architecture: memory change depends on interactions among lability, evidence, models, context, body and action.

Prediction error is a strong candidate gate, but its content, precision and relation to prior learning vary. Contextual similarity strongly influences retrieval and structural assignment, but context has multiple dimensions. Memory strength is a recurrent boundary condition, yet it interacts with reminder design and cannot be reduced to chronological age. Gradual change can favor old-model updating in some paradigms, but no universal intermediate-error optimum has been established.

Predictive processing, active inference and interoception offer powerful explanatory connections, especially for model relativity, bodily context and action. Their connection to molecular reconsolidation is indirect. Clinical learning principles can inform corrective experience and transfer, but efficacy cannot be attributed to reconsolidation without process-sensitive evidence. Strong but qualified is not rhetorical caution. It is the category that preserves useful convergence while keeping the experiment that could reverse the conclusion visible.

Qualified strength is not indecision. It means the central effect is credible while its prevalence, mechanism or boundary remains uncertain. Human retrieval-extinction and propranolol findings fit this category unevenly: some protocols produce durable attenuation, others fail under close replication, and no direct lability marker resolves the difference. The scientifically useful response is to model heterogeneity and improve measurement, not average the disagreement into certainty.

EVIDENCE ANCHORS

[23] Schiller et al. (2010), Nature - Healthy adults - Moderate evidence
[34] Chalkia et al. (2020), Cortex - Healthy adults - Strong evidence
[90] Pigeon et al. (2022), Journal of Psychiatry & Neuroscience - Healthy and clinical human samples - Moderate evidence

57 Red-Line Claims

EVIDENCE STATUS: UNSUPPORTED OR CONTRADICTED.

Lecture 49 must not claim that every retrieval rewrites memory, that reconsolidation reliably erases fear, trauma or addiction memories, that prediction error has a universal optimal dose, or that reduced responding proves trace modification. It must not present propranolol as an established PTSD reconsolidation treatment.

ARC is not a neural location, biomarker, score or treatment protocol. Revision Portability is not a validated construct. Seth and Friston did not demonstrate either concept. A return of responding does not prove that earlier change was unreal, and one laboratory memory system cannot stand for all human remembering.

Lecture 49 must not claim that every retrieval rewrites memory, that emotional intensity identifies lability, that prediction error has one optimal dose, or that reduced responding proves trace modification. It must not describe extinction as ordinary erasure, propranolol as an established PTSD reconsolidation treatment, or successful psychotherapy as evidence for one common memory mechanism. These claims either contradict decisive findings or exceed the directness of available evidence.

The conceptual red lines are equally important. ARC is not a place in the brain, biomarker, score or protocol. Revision Portability is not a validated trait or clinical measure. Seth and Friston did not discover either construct. Attractor, basin and corridor language must retain its status as formal interpretation, Flow Hijacked model or metaphor unless operational neural dynamics are measured. Zero neuro-hype means that explanatory elegance never substitutes for process identification.

These red lines apply equally to commercial and compassionate language. Promising permanent erasure can intensify blame when symptoms return; claiming a precise neural window can create false urgency; describing emotional intensity as proof of rewriting can reward destabilization without benefit. Hope does not require those claims. It can rest on the better-supported proposition that new learning, environmental support and repeated retrieval can change what becomes reachable.

EVIDENCE ANCHORS

[31] Elsey et al. (2018), Psychological Bulletin - Human - Strong evidence
[91] Raut et al. (2022), Journal of Psychiatric Research - Adults with PTSD - Strong evidence
[93] Milton (2023), Translational Psychiatry - Human and animal; addiction/fear/trauma - Strong evidence

58 The Research Programme

EVIDENCE STATUS: FRONTIER / TESTABLE PROGRAMME.

The next generation of studies should independently manipulate discrepancy type, contextual similarity, memory strength, reminder duration, precision cues, bodily state and available action. It should measure latent-cause assignment and use delayed tests across contexts and response systems. Designs must distinguish modification, inhibitory learning, representational separation and performance.

ARC earns value only if its variables predict old-model revision beyond existing accounts. Revision Portability earns value only if transfer across bodily and action dimensions reveals stable information beyond standard generalization. Both concepts should be preregistered, compared with simpler models and abandoned when they fail.

The next generation of studies should cross variables that are usually examined separately. Memory strength, reminder completeness, discrepancy content, contextual similarity, precision cues, bodily state and action affordances can be manipulated factorially. Outcomes should include immediate expression, delayed retention, renewal, reinstatement, source memory, policy and representational change. The design must permit little-change restabilization, old-model update, separate learning and performance to compete as explanations.

Model comparison is essential. ARC-derived predictions should be tested against Rescorla-Wagner, temporal-context, latent-cause, nonmonotonic-plasticity and ordinary extinction baselines. Portability models should be compared with standard generalization and transfer accounts. Preregistration should define monotonic and nonlinear alternatives before data are seen, and held-out prediction should determine whether added complexity earns its cost. Both Flow Hijacked constructs are successful only if they risk and survive this comparison.

The programme should also be openly cumulative. Protocols, stimuli, code, exclusion decisions and null results should be reusable; outcome batteries should retain common measures while permitting task-specific additions. Cross-species work should align computations rather than surface behavior alone. Most importantly, independent teams should be able to test ARC and Portability without adopting Flow Hijacked language, because a useful construct must survive outside the system that named it.

EVIDENCE ANCHORS

[38] Milton et al. (2023), Brain Research Bulletin - Cross-species - Strong evidence
[39] Kennedy et al. (2024), eLife - Rat - Strong evidence
[94] Meijer et al. (2026), Science Advances - Healthy adults - Moderate–strong evidence

59 Closing Proposition - Change Must Travel

EVIDENCE STATUS: FLOW HIJACKED SYNTHESIS.

The scientific story is not that memory is clay. It is that remembering places an adaptive system at a fork. Sometimes the old prediction remains useful. Sometimes it becomes revisable. Sometimes the world is judged different enough to deserve another model. And sometimes meaningful change occurs but cannot yet govern behavior far from where it was learned.

A compassionate science holds all of these possibilities without turning recurrence into personal failure. Change is not only whether an old model can be revised. It is whether the revision can remain reachable across the states, contexts and possibilities of a human life.

The lecture began with one contradiction and three possible futures. The old representation may remain substantially intact, it may be revised, or new learning may take responsibility for the event. The evidence does not collapse these fates into one rule. It shows regulated plasticity: retrieval opens a problem, destabilization is conditional, error is model-relative, context governs assignment and multiple memory systems can change differently at the same time.

The final proposition adds a second criterion. Change matters in life when it remains reachable after movement through context, body, cues and action. This does not require permanent immunity from return, and it does not make recurrence a moral failure. It requires measurement beyond the room of success. ARC asks whether an old model can change. Revision Portability asks whether the result can travel. Together they form a research architecture, not a promise: revision without neuro-hype, uncertainty without resignation, and compassion without simplification.

The closing criterion is deliberately behavioral and humane. A revised model matters when it becomes available in the moments that previously narrowed choice, across enough variation to support a life rather than one demonstration. No experiment can guarantee that reach in advance. Science can, however, map where change holds, where it fails and which supports extend it. That map respects both the persistence of old learning and the reality of new possibility.

EVIDENCE ANCHORS

[53] Craske et al. (2014), Behaviour Research and Therapy - Anxiety treatment and laboratory fear - Strong evidence
[57] Rajagopal et al. (2025), Nature Communications - Rodent/human fear data constraints - Moderate evidence
[93] Milton (2023), Translational Psychiatry - Human and animal; addiction/fear/trauma - Strong evidence

Change is not only whether a model can be revised. It is whether the revision can travel.

References - 94-work research portfolio

The portfolio is organized for traceability, contradiction and conceptual lineage rather than prestige. Article-level sample sizes are marked NR where they could not be verified without invention.

1. Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. Science, 160, 554-555. doi:10.1126/science.160.3827.554

2. Przybyslawski, J., & Sara, S. J. (1997). Reconsolidation of memory after its reactivation. Behavioural Brain Research, 84, 241-246. doi:10.1016/S0166-4328(96)00153-2

3. Nader, K., Schafe, G. E., & LeDoux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. Nature, 406, 722-726. doi:10.1038/35021052

4. Sara, S. J. (2000). Retrieval and reconsolidation: toward a neurobiology of remembering. Learning & Memory, 7, 73-84. doi:10.1101/lm.7.2.73

5. Milekic, M. H., & Alberini, C. M. (2002). Temporally graded requirement for protein synthesis following memory reactivation. Neuron, 36, 521-525. doi:10.1016/S0896-6273(02)00976-5

6. Debiec, J., LeDoux, J. E., & Nader, K. (2002). Cellular and systems reconsolidation in the hippocampus. Neuron, 36, 527-538. doi:10.1016/S0896-6273(02)01001-2

7. Debiec, J., & LeDoux, J. E. (2004). Disruption of reconsolidation but not consolidation of auditory fear conditioning by noradrenergic blockade in the amygdala. Neuroscience, 129, 267-272. doi:10.1016/j.neuroscience.2004.08.018

8. Suzuki, A., et al. (2004). Memory reconsolidation and extinction have distinct temporal and biochemical signatures. Journal of Neuroscience, 24, 4787-4795. doi:10.1523/JNEUROSCI.5491-03.2004

9. Pedreira, M. E., Pérez-Cuesta, L. M., & Maldonado, H. (2004). Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. Learning & Memory, 11, 579-585. doi:10.1101/lm.76904

10. Ben Mamou, C., Gamache, K., & Nader, K. (2006). NMDA receptors are critical for unleashing consolidated auditory fear memories. Nature Neuroscience, 9, 1237-1239. doi:10.1038/nn1778

11. Tronson, N. C., & Taylor, J. R. (2007). Molecular mechanisms of memory reconsolidation. Nature Reviews Neuroscience, 8, 262-275. doi:10.1038/nrn2090

12. Lee, J. L. C. (2008). Memory reconsolidation mediates the strengthening of memories by additional learning. Nature Neuroscience, 11, 1264-1266. doi:10.1038/nn.2205

13. Nader, K., & Hardt, O. (2009). A single standard for memory: the case for reconsolidation. Nature Reviews Neuroscience, 10, 224-234. doi:10.1038/nrn2590

14. Wang, S.-H., de Oliveira Alvares, L., & Nader, K. (2009). Cellular and systems mechanisms of memory strength as a constraint on auditory fear reconsolidation. Nature Neuroscience, 12, 905-912. doi:10.1038/nn.2350

15. Hardt, O., Einarsson, E. O., & Nader, K. (2010). A bridge over troubled water: reconsolidation as a link between cognitive and neuroscientific memory research traditions. Annual Review of Psychology, 61, 141-167. doi:10.1146/annurev.psych.093008.100455

16. Rao-Ruiz, P., et al. (2011). Retrieval-specific endocytosis of GluA2-AMPARs underlies adaptive reconsolidation of contextual fear. Nature Neuroscience, 14, 1302-1308. doi:10.1038/nn.2907

17. Dudai, Y. (2012). The restless engram: consolidations never end. Annual Review of Neuroscience, 35, 227-247. doi:10.1146/annurev-neuro-062111-150500

18. Alberini, C. M., & LeDoux, J. E. (2013). Memory reconsolidation. Current Biology, 23, R746-R750. doi:10.1016/j.cub.2013.06.046

19. Carneiro, C. F. D., et al. (2023). A meta-analysis of protein synthesis inhibition in learning and memory. Neuroscience & Biobehavioral Reviews, 146, 105042. doi:10.1016/j.neubiorev.2023.105042

20. Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. Learning & Memory, 14, 47-53. doi:10.1101/lm.365707

21. Monfils, M.-H., Cowansage, K. K., Klann, E., & LeDoux, J. E. (2009). Extinction-reconsolidation boundaries: key to persistent attenuation of fear memories. Science, 324, 951-955. doi:10.1126/science.1167975

22. Kindt, M., Soeter, M., & Vervliet, B. (2009). Beyond extinction: erasing human fear responses and preventing the return of fear. Nature Neuroscience, 12, 256-258. doi:10.1038/nn.2271

23. Schiller, D., et al. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. Nature, 463, 49-53. doi:10.1038/nature08637

24. Sevenster, D., Beckers, T., & Kindt, M. (2012). Retrieval per se is not sufficient to trigger reconsolidation of human fear memory. Neurobiology of Learning and Memory, 97, 338-345. doi:10.1016/j.nlm.2012.01.009

25. Sevenster, D., Beckers, T., & Kindt, M. (2013). Prediction error governs pharmacologically induced amnesia for learned fear. Science, 339, 830-833. doi:10.1126/science.1231357

26. Díaz-Mataix, L., Ruiz Martinez, R. C., Schafe, G. E., LeDoux, J. E., & Doyère, V. (2013). Detection of a temporal error triggers reconsolidation of amygdala-dependent memories. Current Biology, 23, 467-472. doi:10.1016/j.cub.2013.01.053

27. Sevenster, D., Beckers, T., & Kindt, M. (2014). Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. Learning & Memory, 21, 580-584. doi:10.1101/lm.035493.114

28. Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories-The role of prediction errors in memory reconsolidation. Behavioural Brain Research, 278, 375-384. doi:10.1016/j.bbr.2014.10.011

29. Fernández, R. S., Bavassi, L., Forcato, C., & Pedreira, M. E. (2016). The dynamic nature of the reconsolidation process and its boundary conditions: evidence based on human tests. Neurobiology of Learning and Memory, 130, 202-212. doi:10.1016/j.nlm.2016.03.001

30. Fernández, R. S., Boccia, M. M., & Pedreira, M. E. (2016). The fate of memory: reconsolidation and the case of prediction error. Neuroscience & Biobehavioral Reviews, 68, 423-441. doi:10.1016/j.neubiorev.2016.06.004

31. Elsey, J. W. B., Van Ast, V. A., & Kindt, M. (2018). Human memory reconsolidation: a guiding framework and critical review of the evidence. Psychological Bulletin, 144, 797-848. doi:10.1037/bul0000152

32. Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: prediction error influences episodic memory reconsolidation. Learning & Memory, 25, 369-381. doi:10.1101/lm.046912.117

33. Sinclair, A. H., & Barense, M. D. (2019). Prediction error and memory reactivation: how incomplete reminders drive reconsolidation. Trends in Neurosciences, 42, 727-739. doi:10.1016/j.tins.2019.08.007

34. Chalkia, A., et al. (2020). No persistent attenuation of fear memories in humans: a registered replication of the reactivation-extinction effect. Cortex, 129, 496-509. doi:10.1016/j.cortex.2020.04.017

35. Sinclair, A. H., Manalili, G. M., Brunec, I. K., Adcock, R. A., & Barense, M. D. (2021). Prediction errors disrupt hippocampal representations and update episodic memories. Proceedings of the National Academy of Sciences, 118, e2117625118. doi:10.1073/pnas.2117625118

36. Stemerding, L. E., Stibbe, D., van Ast, V. A., & Kindt, M. (2022). Demarcating the boundary conditions of memory reconsolidation: an unsuccessful replication. Scientific Reports, 12, 2285. doi:10.1038/s41598-022-06119-5

37. Gerlicher, A. M. V., Verweij, S. A., & Kindt, M. (2022). Better, worse, or different than expected: on the role of value and identity prediction errors in fear memory reactivation. Scientific Reports, 12, 5862. doi:10.1038/s41598-022-09720-w

38. Milton, A. L., Das, R. K., & Merlo, E. (2023). The challenge of memory destabilisation: from prediction error to prior learning history. Brain Research Bulletin, 194, 235-245. doi:10.1016/j.brainresbull.2023.01.010

39. Kennedy, A. C., et al. (2024). Prediction error determines whether fear extinction updates the old memory or creates a new memory. eLife, 13, RP95849. doi:10.7554/eLife.95849

40. Mirea, D. M., et al. (2024). How prior knowledge shapes episodic memory updating: large-scale evidence for individual differences in mnemonic prediction. Journal of Cognitive Neuroscience, 36, 2415-2434. doi:10.1162/jocn_a_02231

41. Mahmoodi, A., et al. (2024). The hippocampus and dorsomedial prefrontal cortex support causal inference and association updating. Neuron, 112, 3924-3938.e8. doi:10.1016/j.neuron.2024.09.001

42. Chen, J., et al. (2025). Prediction error drives episodic memory updating under boundary-sensitive conditions. Cognition, 262, 106224. doi:10.1016/j.cognition.2025.106224

43. Yang, Y., et al. (2026). Memory strength at reactivation governs prediction-error-driven episodic updating across time. Philosophical Transactions of the Royal Society B, 381, 20250244. doi:10.1098/rstb.2025.0244

44. Bouton, M. E. (2002). Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. Biological Psychiatry, 52, 976-986. doi:10.1016/S0006-3223(02)01546-9

45. Phelps, E. A., Delgado, M. R., Nearing, K. I., & LeDoux, J. E. (2004). Extinction learning in humans: role of the amygdala and vmPFC. Neuron, 43, 897-905. doi:10.1016/j.neuron.2004.08.042

46. Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. Psychological Review, 117, 197-209. doi:10.1037/a0017808

47. Fanselow, M. S. (2010). From contextual fear to a dynamic view of memory systems. Trends in Cognitive Sciences, 14, 7-15. doi:10.1016/j.tics.2009.10.008

48. Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. Psychonomic Bulletin & Review, 18, 455-468. doi:10.3758/s13423-011-0086-9

49. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. Neuron, 69, 1204-1215. doi:10.1016/j.neuron.2011.02.027

50. Gershman, S. J., & Niv, Y. (2012). Exploring a latent cause theory of classical conditioning. Learning & Behavior, 40, 255-268. doi:10.3758/s13420-012-0080-8

51. Gershman, S. J., Jones, C. E., Norman, K. A., Monfils, M.-H., & Niv, Y. (2013). Gradual extinction prevents the return of fear: implications for the discovery of state. Frontiers in Behavioral Neuroscience, 7, 164. doi:10.3389/fnbeh.2013.00164

52. Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear extinction and relapse: state of the art. Annual Review of Clinical Psychology, 9, 215-248. doi:10.1146/annurev-clinpsy-050212-185542

53. Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: an inhibitory learning approach. Behaviour Research and Therapy, 58, 10-23. doi:10.1016/j.brat.2014.04.006

54. Dunsmoor, J. E., Niv, Y., Daw, N., & Phelps, E. A. (2015). Rethinking extinction. Neuron, 88, 47-63. doi:10.1016/j.neuron.2015.09.028

55. Starkweather, C. K., Babayan, B. M., Uchida, N., & Gershman, S. J. (2017). Dopamine reward prediction errors reflect hidden-state inference across time. Nature Neuroscience, 20, 581-589. doi:10.1038/nn.4520

56. Gershman, S. J., Monfils, M.-H., Norman, K. A., & Niv, Y. (2017). The computational nature of memory modification. eLife, 6, e23763. doi:10.7554/eLife.23763

57. Rajagopal, S. K., et al. (2025). A neurally constrained computational model of context-dependent fear extinction and renewal. Nature Communications, 16, 3867. doi:10.1038/s41467-025-59036-4

58. Yoo, S. W., et al. (2017). Hippocampal encoding of interoceptive context during fear conditioning and extinction. Translational Psychiatry, 7, e991. doi:10.1038/tp.2016.254

59. Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron, 60, 378-389. doi:10.1016/j.neuron.2008.09.023

60. Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. Science, 319, 1640-1642. doi:10.1126/science.1152882

61. Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. Trends in Neurosciences, 34, 515-525. doi:10.1016/j.tins.2011.06.006

62. Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron, 75, 168-179. doi:10.1016/j.neuron.2012.05.010

63. Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nature Communications, 6, 8151. doi:10.1038/ncomms9151

64. Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. Nature Communications, 7, 11066. doi:10.1038/ncomms11066

65. Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among spatial memories triggers repulsion of hippocampal representations. Current Biology, 27, 2307-2317.e5. doi:10.1016/j.cub.2017.06.057

66. Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic plasticity: how memory retrieval drives learning. Trends in Cognitive Sciences, 23, 726-742. doi:10.1016/j.tics.2019.06.007

67. McClay, M., et al. (2022). Neural evidence for the nonmonotonic plasticity hypothesis in the human hippocampus. eLife, 11, e68344. doi:10.7554/eLife.68344

68. Bein, O., Duncan, K., & Davachi, L. (2023). Mnemonic prediction errors bias hippocampal states toward encoding and memory updating. Neuroscience & Biobehavioral Reviews, 151, 105368. doi:10.1016/j.neubiorev.2023.105368

69. Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. Science, 275, 1593-1599. doi:10.1126/science.275.5306.1593

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

71. Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. Trends in Cognitive Sciences, 17, 565-573. doi:10.1016/j.tics.2013.09.007

72. Mathys, C. D., et al. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. Frontiers in Human Neuroscience, 8, 825. doi:10.3389/fnhum.2014.00825

73. Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. Nature Reviews Neuroscience, 16, 419-429. doi:10.1038/nrn3950

74. Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. Progress in Neurobiology, 134, 17-35. doi:10.1016/j.pneurobio.2015.09.001

75. Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. Philosophical Transactions of the Royal Society B, 371, 20160007. doi:10.1098/rstb.2016.0007

76. Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: a process theory. Neural Computation, 29, 1-49. doi:10.1162/NECO_a_00912

77. Osan, R., Tort, A. B. L., & Amaral, O. B. (2011). A mismatch-based model for memory reconsolidation and extinction in attractor networks. PLoS ONE, 6, e23113. doi:10.1371/journal.pone.0023113

78. Spalla, D., Cornacchia, I. M., & Treves, A. (2021). Continuous attractors for dynamic memories. eLife, 10, e69499. doi:10.7554/eLife.69499

79. Lee, J. L. C., Di Ciano, P., Thomas, K. L., & Everitt, B. J. (2005). Disrupting reconsolidation of drug memories reduces cocaine-seeking behavior. Neuron, 47, 795-801. doi:10.1016/j.neuron.2005.08.007

80. Brunet, A., et al. (2008). Effect of post-retrieval propranolol on psychophysiologic responding during subsequent script-driven traumatic imagery in PTSD. Journal of Psychiatric Research, 42, 503-506. doi:10.1016/j.jpsychires.2007.05.006

81. Torregrossa, M. M., Corlett, P. R., & Taylor, J. R. (2011). Aberrant learning and memory in addiction. Neurobiology of Learning and Memory, 96, 609-623. doi:10.1016/j.nlm.2011.02.014

82. Xue, Y.-X., et al. (2012). A memory retrieval-extinction procedure to prevent drug craving and relapse. Science, 336, 241-245. doi:10.1126/science.1215070

83. Das, R. K., Lawn, W., & Kamboj, S. K. (2015). Rewriting the valuation and salience of alcohol-related stimuli via memory reconsolidation. Translational Psychiatry, 5, e645. doi:10.1038/tp.2015.132

84. Luo, Y.-X., et al. (2015). A novel UCS-memory retrieval-extinction procedure to inhibit relapse to drug seeking. Nature Communications, 6, 7675. doi:10.1038/ncomms8675

85. Lane, R. D., Ryan, L., Nadel, L., & Greenberg, L. (2015). Memory reconsolidation, emotional arousal, and the process of change in psychotherapy. Behavioral and Brain Sciences, 38, e1. doi:10.1017/S0140525X14000041

86. Soeter, M., & Kindt, M. (2015). An abrupt transformation of phobic behavior after a post-retrieval amnesic agent. Biological Psychiatry, 78, 880-886. doi:10.1016/j.biopsych.2015.04.006

87. Everitt, B. J., & Robbins, T. W. (2016). Drug addiction: updating actions to habits to compulsions ten years on. Annual Review of Psychology, 67, 23-50. doi:10.1146/annurev-psych-122414-033457

88. Brunet, A., et al. (2018). Reduction of PTSD symptoms with pre-reactivation propranolol: a randomized controlled trial. American Journal of Psychiatry, 175, 427-433. doi:10.1176/appi.ajp.2017.17050481

89. Das, R. K., Gale, G., Hennessy, V., & Kamboj, S. K. (2019). A prediction-error driven retrieval procedure for destabilizing and rewriting maladaptive reward memories. Nature Communications, 10, 5027. doi:10.1038/s41467-019-13162-w

90. Pigeon, S., et al. (2022). Impairing memory reconsolidation with propranolol in healthy and clinical samples: a meta-analysis. Journal of Psychiatry & Neuroscience, 47, E109-E122. doi:10.1503/jpn.210057

91. Raut, S. B., Canales, J. J., Ravindran, M., & Eri, R. (2022). Propranolol in post-traumatic stress disorder: a systematic review and meta-analysis. Journal of Psychiatric Research, 150, 190-202. doi:10.1016/j.jpsychires.2022.03.045

92. Steenen, S. A., van Wijk, A. J., van der Heijden, G. J. M. G., & van Westrhenen, R. (2022). Reanalysis: no convincing effect of propranolol on memory reconsolidation outcomes. Journal of Psychiatry & Neuroscience, 47, E312-E314. doi:10.1503/jpn.220072-l

93. Milton, A. L. (2023). Harnessing memory reconsolidation to treat maladaptive memories: strengths, limits and translational priorities. Translational Psychiatry, 13, 302. doi:10.1038/s41398-023-02666-1

94. Meijer, G. T., et al. (2026). Causal contributions of the human amygdala and hippocampus to threat learning and extinction. Science Advances, 12, aea8233. doi:10.1126/sciadv.aea8233