Memory erasure by dopamine-gated retrospective learning

Huijeong Jeong[1], Leo Zsembik[2,12], Farah Farouq[1,12], Risha Chakraborty[3], Nishita Belur[4], Mingkang Zhou[2], Andrea D. Sanders[2], Styra X. Wang[1], Ananya Srinivasan[5], Sylvia M. L. Cox[6,7], Eric Garr[8], Sara Brooke[9], Patricia H. Janak[8,9], Marco Leyton[6,7], Ritchie Chen[2,10,11], and Vijay Mohan K Namboodiri[1,2,11,*]

[1]Department of Neurology, University of California, San Francisco, CA, USA
[2]Neuroscience Graduate Program, University of California, San Francisco, CA, USA
[3]Department of Neuroscience, Yale University, CT, USA
[4]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA
[5]Department of Neuroscience, University of California, Berkeley, CA, USA
[6]Department of Psychiatry, McGill University, Québec, Canada
[7]Department of Neurology and Neurosurgery, McGill University, Québec, Canada
[8]Department of Psychological & Brain Sciences, Krieger School of Arts & Sciences, Johns Hopkins University, MD, USA
[9]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University, MD, USA
[10]Department of Neurological Surgery, University of California, San Francisco, CA, USA
[11]Weill Institute for Neuroscience, Kavli Institute for Fundamental Neuroscience, Center for Integrative Neuroscience, University of California, San Francisco, CA, USA
[12]These authors contributed equally to this work (co-second author)
[*]Corresponding author: VijayMohan.KNamboodiri@ucsf.edu

Erasing outdated memories is crucial for adaptive behavior. Yet once a cue–outcome association is learned, repeated cue exposure without outcome suppresses conditioned behavior without erasing the underlying memory. This allows rapid behavioral recovery when outcomes are reintroduced. Here, we confirm this limitation for standard "prospective extinction" protocols that present cues without the associated outcome, but show that true memory erasure is achieved by inverting the paradigm: presenting outcomes without associated cues, i.e., "retrospective extinction". We demonstrate that orbitofrontal cortex activity at outcome is necessary for the rapid behavioral recovery following prospective extinction, and that mesolimbic dopamine activity at outcome is necessary for retrospective extinction. These findings reconceptualize extinction mechanisms and suggest complementary strategies to mitigate relapse and erase maladaptive memories.

Introduction

Unlearning outdated or harmful memories is as critical as forming new ones, enabling adaptive behavior when cues no longer predict outcomes. For instance, when trauma-associated cues no longer predict danger (referred to as extinction of cue-trauma pairing), persistent cue-trauma memories can still provoke distress. Similarly, persistent cue-drug memories allow drug-associated environmental cues to trigger craving despite abstinence. These enduring memories drive chronic relapse cycles (1–3). Standard extinction—repeated cue presentation without the outcome—suppresses conditioned behavior but leaves the underlying memory intact (4–7), as shown by rapid reacquisition with minimal cue-outcome re-exposure (**Fig. 1A**) (8–10) or spontaneous recovery of conditioned behavior (11–13). Although modified extinction procedures have been reported to reduce behavioral recovery compared to standard extinction (14–24), results remain variable across studies (25–34), leaving unclear what conditions genuinely induce unlearning. Identifying concise, reproducible conditions that truly erase memories could transform interventions for post-traumatic stress disorder (PTSD), addiction, and related disorders involving maladaptive memories.

The maintenance of memory despite behavioral extinction raises two questions: how is the original memory maintained during behavior extinction, and how is its behavioral impact suppressed during extinction? The widely accepted explanation is that extinction forms a new memory that competes with, instead of replacing, the original memory (5–7, 11, 35) (**Fig. 1B**, Conventional model). Reward prediction error (RPE)-based latent cause models formalize this idea by proposing that animals learn cue-outcome (e.g., cue-reward) memories for inferred hidden "contexts" even when physical contexts do not change (36–40). These models need to assume hidden contexts to explain memory maintenance during extinction, because memories in these models are prospective, storing the prediction of what follows the cue, i.e., $P(\text{cue} \rightarrow \text{reward})$. To allow two distinct cue-outcome memories to coexist, the original memory is assigned to an acquisition context [$P(\text{cue} \rightarrow \text{reward} \mid \text{context}_{\text{Acq}}) = 100\%$], while a new extinction memory is assigned to an extinction context [$P(\text{cue} \rightarrow \text{reward} \mid \text{context}_{\text{Ext}}) = 0\%$]. Therefore, an inferred switch from $\text{context}_{\text{Acq}}$ to $\text{context}_{\text{Ext}}$ produces behavioral extinction despite the original memory being preserved, while a switch back to $\text{context}_{\text{Acq}}$ through rapid inference during reacquisition enables fast restoration of conditioned behavior.

An alternate view comes from retrospective learning models, i.e., models learning that cues preferentially *precede* meaningful outcomes, as exemplified by Adjusted Net Contingency of Causal Relations (ANCCR) (**Fig. 1B**, Retrospective model) (41–43). Rather than positing multiple hidden contexts, ANCCR assumes that animals only learn and store a single retrospective $P(\text{cue} \leftarrow \text{reward})$ association indexed by physical context. Animals then infer a prospective $P(\text{cue} \rightarrow \text{reward})$ prediction to drive cue-induced behavior using Bayes' rule: $P(\text{cue} \rightarrow \text{reward}) = P(\text{cue} \leftarrow \text{reward}) \times P(\text{reward})/P(\text{cue})$ (**Supplementary Note 1**). Although prospective and retrospective associations may seem similar,

Fig. 1. Retrospective extinction degrades original memory. A, Rapid reacquisition compared to original acquisition demonstrates that extinction does not erase the original memory. This raises two questions: how is the memory preserved during extinction, and if the memory is preserved, why does behavior extinguish? **B**, Hypotheses for how memory is preserved during extinction proposed by two models: latent cause models and Adjusted Net Contingency of Causal Relations (ANCCR). Latent cause models propose that distinct memories of the same cue are stored in different inferred contexts, such as a cue-reward memory in $\text{context}_{\text{Acq}}$ and a cue-no reward memory in $\text{context}_{\text{Ext}}$. A shift in inferred context from $\text{context}_{\text{Acq}}$ to $\text{context}_{\text{Ext}}$ enables behavioral extinction despite preservation of the original cue-reward memory, whereas a shift back from $\text{context}_{\text{Ext}}$ to $\text{context}_{\text{Acq}}$ enables rapid reacquisition. In ANCCR, the prospective probability $P(\text{cue} \rightarrow \text{reward})$ is the product (indicated by $\times$) of the retrospective probability $P(\text{cue} \leftarrow \text{reward})$ and the normalization factor $P(\text{reward})/P(\text{cue})$. Because the retrospective probability updates only upon reward delivery, it serves as a stable memory that is preserved throughout extinction. In contrast, a reduction in $P(\text{reward})$ during extinction and its rapid restoration during reacquisition drive corresponding decreases and rapid recovery of the prospective probability, and thus behavior. **C**, ANCCR predicts two types of extinction.

(Fig. 1. Continued) In standard extinction, the original memory $P(\text{cue} \leftarrow \text{reward})$ is preserved, but a reduction in $P(\text{reward})$ drives behavioral extinction, as illustrated in **B**. In retrospective extinction, reward delivery dissociated from the cue, either preceded by another cue or delivered without any cue, degrades the retrospective association, resulting in unlearning. **D**, Task schematic. Mice first learned to associate both CS-A and CS-B with reward. This was followed by CS-A extinction (differential across groups), CS-B extinction (identical across groups), and then reacquisition for both cues (identical). **E**, Task design for testing two types of extinction. During CS-A (red cue) extinction, Group 1 (G1) underwent standard extinction, whereas Group 2 (G2) underwent retrospective extinction. CS-B (blue cue) underwent standard extinction for both groups, allowing both between-animal and within-animal comparisons of memory persistence. **F**, Behavior was quantified as the change in lick count during the 1.25 s following cue onset ("lick to CS"). Learning was assessed either by the area under the curve (AUC) of the learning curve or by identifying the inflection point ("trial to learn," denoted by a vertical line) of the cumulative learning curve. **G**, Example lick raster plots (upper row) and peri-stimulus time histograms (PSTHs; lower row) from one representative mouse per group aligned to CS-A trials (top, G1; bottom, G2). The first column shows data from the last day of CS-B extinction, and subsequent columns (Day 1 to Day 5) show reacquisition. Graphs are aligned to cue onset, denoted by the first vertical black dotted line. The second black dotted line indicates reward delivery or omission. The G1 mouse shows increased anticipatory licking before reward on Day 1 of reacquisition, whereas the G2 mouse develops anticipatory licking only by Day 4. Gray shading indicates the anticipatory window used to calculate lick to CS. **H**, Group-averaged normalized lick to CS-A across trials during reacquisition showing faster reacquisition in G1 compared to G2. Trial 0 marks the first CS-A trial in reacquisition. Each animal's lick to CS-A was normalized to its average during the last three days of acquisition. Lines represent group means and shaded areas represent SEM. **I**, Reacquisition is more rapid than acquisition only in G1, not in G2. Area under the curve (AUC) of normalized lick to CS-A over the first six days of acquisition and reacquisition. Error bars represent SEM and light-colored lines represent individual animals. **J**, Group-averaged cumulative sum of normalized lick to CS-A during acquisition (dotted) and reacquisition (solid). Trial 0 marks the first trial of each respective phase. For reacquisition, CS-B extinction trials are shown at trial numbers less than 0. **K**, Fraction of learners (filled bars) and non-learners (open bars) for each group during acquisition and reacquisition. Animals from G1 and G2 learned similarly during acquisition. Only animals that learned during acquisition ($n = 41$) were divided into groups and included in subsequent experiments. **L**, Reacquisition was faster than acquisition in G1 but similarly slow in G2 when non-learners were excluded (7 of 19 G2 animals during reacquisition). Error bars represent SEM and individual animals are shown as scatter points. **M**, G1 shows stronger spontaneous recovery than G2. Lick to CS-A across trials within an extinction session, averaged across the first eight sessions after behavioral extinction (see Methods). Each data point represents the average lick response at a given trial number across sessions, then averaged across animals. **N**, Area under the curve (AUC) across the first five trials of the trace shown in **M**, quantifying the magnitude of spontaneous recovery. **O**, Example heatmaps (upper row) and PSTHs (lower row) of dopamine release ($\Delta F/F$) in the nucleus accumbens in response to CS-A from a representative G1 and G2 mouse. Dopamine responses to the cue emerge earlier in G1 than in G2, consistent with behavior. Gray shading indicates the anticipatory window used to calculate dopamine to CS. **P**, Reacquisition of dopamine responses is more rapid than acquisition only in G1, not in G2. Group-averaged cumulative sum of normalized dopamine to CS-A during acquisition (dotted) and reacquisition (solid). Dopamine to CS-A was calculated as the AUC of $\Delta F/F$ over the 1.25 s cue-reward window and normalized to the maximum reward response (AUC over the 1.25 s window following reward) during the first three days of acquisition. **Q**, Area under the curve (AUC) of normalized dopamine to CS-A during the first six days of acquisition and reacquisition.

p<0.05, **p<0.01, ***p<0.001; n.s., not significant. See **Supplementary Table 2 for full statistical results. Error bars and shaded regions throughout the figure represent SEM unless otherwise noted.*

they differ substantially. For example, if a reward is always preceded by a cue but the cue is followed by reward only 10% of the time, the prospective association is weak [$P(\text{cue} \rightarrow \text{reward}) = 10\%$] while the retrospective association is perfect [$P(\text{cue} \leftarrow \text{reward}) = 100\%$]. This asymmetry means that lowering reward probability all the way to zero during extinction can abolish prospective $P(\text{cue} \rightarrow \text{reward})$ association while leaving the retrospective $P(\text{cue} \leftarrow \text{reward})$ association intact. The retrospective association therefore forms a persistent memory of the previous cue-reward pairing. Behavior, which is guided by the prospective prediction of reward following the cue, is extinguished due to the overall reward rate (i.e., $P(\text{reward})/P(\text{cue})$) becoming zero during extinction (**Fig. 1C**, Standard prospective extinction). The preserved retrospective memory enables rapid restoration of behavior when the estimated reward rate rapidly recovers during reacquisition.

Intuitively, the retrospective framework helps explain why abstinence from drugs, or extinction-based approaches such as exposure therapy for aversive outcomes, fail to entirely erase maladaptive drug or trauma memories. Repeated cue exposure without outcome teaches that the outcome is unlikely to happen, suppressing conditioned responses to the cue, but leaves intact the memory that whenever the outcome occurs, it is preceded by the cue.

By contrast, cue-reward memory erasure requires abolishing the retrospective $P(\text{cue} \leftarrow \text{reward})$ association. This quantity measures the fraction of rewards that are preceded by the cue and, by definition, is updated only at the time of reward ("whenever a reward occurs, check whether it was preceded by the cue"). Therefore, ANCCR predicts that true unlearning occurs only when rewards are experienced without the associated cue preceding them (they may be fully uncued like in contingency degradation experiments, or be preceded by another cue, i.e., part of another trial type). This process, which we term retrospective extinction, is the only way to degrade the retrospective $P(\text{cue} \leftarrow \text{reward})$ association and induce true memory erasure (**Fig. 1C**, retrospective extinction).

In sum, in ANCCR, the primary stored memory is a single retrospective association indexed by the physical context, rather than multiple prospective memories indexed by hidden contexts. This distinction is the central conceptual difference between ANCCR and latent-cause RPE frameworks. The distinction between these frameworks is crucial. Beyond advancing our understanding of learning, establishing conditions for memory erasure could inform the design of more durable treatments for disorders in which maladaptive memories drive persistent harmful behavior.

Results

Behavioral learning in one-tenth the experiences with ten times the trial spacing. We began by testing whether retrospective extinction can, in fact, erase a cue-reward memory. We employed Pavlovian conditioning in head-fixed mice that randomly interleaves two distinct tones (CS-A, CS-B). During the *Acquisition* phase, both cues predicted sucrose reward with 50% probability, denoted [$\text{A} \rightarrow 50\%$ sucrose, $\text{B} \rightarrow 50\%$ sucrose], and conditioned behavior was measured using anticipatory licking (**Fig. 1D–F**). Thereafter, mice were divided into two groups to perform either standard or retrospective extinction of CS-A, while performing standard extinction of CS-B in both groups. CS-B trials therefore provide a within-animal control for comparing memory strength to CS-A. To achieve such cue-specific extinction, in the phase following *Acquisition* (*"CS-A extinction"*), Group1 (G1; standard extinction, purple) received only CS-A followed by reward omission [$\text{A} \rightarrow \text{omit}$]. In contrast, Group2 (G2; retrospective extinction, green) experienced CS-A omission trials randomly interleaved with CS-B trials [$\text{A} \rightarrow \text{omit}$, $\text{B} \rightarrow 50\%$ sucrose]. For both groups,

we continued the protocol until responding to CS-A was eliminated. Thereafter, CS-B underwent standard extinction (*"CS-B extinction"*), in which both CS-A and CS-B were unrewarded [A → omit, B → omit] (see **fig. S1** for why a simpler alternative for cue-specific retrospective extinction does not slow reacquisition). Memory was then tested in the *Reacquisition* phase, which resumed the original rule [A → 50% sucrose, B → 50% sucrose] (**Fig. 1D**), by determining whether behavioral reacquisition was faster than the original acquisition (**Fig. 1F**). In short, this paradigm selectively degrades the retrospective CS-A←reward association in G2, but not in G1 (**Fig. 1E**; results from a single cue-reward paradigm are shown later).

Because CS-A underwent either retrospective or standard extinction depending on the group, while CS-B always underwent standard extinction in both groups, we first focused on CS-A trials (results from CS-B trials are shown in **Fig. 3**). G1 animals reacquired anticipatory licking to CS-A within a few trials, learning significantly faster than during initial acquisition, and demonstrating memory preservation under standard extinction. In contrast, G2 animals reacquired CS-A behavior much more slowly than G1, at a rate matching their own acquisition, and consistent with erasure of the original memory (**Fig. 1G–L**). Interestingly, 40% of animals in G2 (compared to 0% in G1) never reacquired conditioned behavior, demonstrating the power of retrospective extinction (**Fig. 1K**, **fig. S2A,B**). Although not a core prediction of retrospective extinction, this effect may arise because, having truly unlearned a previous association, G2 may be more hesitant to act on it the next time when it appears to be active. G2 also showed reduced spontaneous recovery of licking during extinction sessions (**Fig. 1M,N**), further indicating memory erasure. Simultaneous dopamine release measurement in a subset of animals revealed parallel group differences (**Fig. 1O–Q**): rapid reacquisition of dopamine responses to CS-A in G1, but reacquisition as slow as acquisition in G2. These findings demonstrate that retrospective extinction erases cue-reward memory, evident in both behavior and dopamine responses.

We next asked whether the group difference in reacquisition could be attributed to other differences between groups. G1 and G2 did not differ in anticipatory behavior immediately before reacquisition (**fig. S2C**) or in spontaneous recovery (**fig. S2D**). Reacquisition was consistently faster in G1 than G2 regardless of the number of matching extinction sessions (**fig. S2F–H**). G2 showed slower CS-A extinction than G1 (**fig. S3A–D**), consistent with a protective effect of cross-cue generalization (**Supplementary Note 2**; **fig. S3E–H**). However, extinction rate did not correlate with reacquisition rate within either group (**fig. S2E**). Therefore, reacquisition results are not explained by these differences. Additionally, for *CS-B extinction*, ANCCR predicts slower extinction in G2 than G1, because the reward rate remained elevated during the *CS-A extinction* phase in G2 but was degraded to zero in G1 (**fig. S4A,B**). Behavior and dopamine responses supported this prediction (**fig. S4C–H**). ANCCR also predicts faster behavioral extinction of CS-B than CS-A within G1, which we likewise observed (**fig. S4I–O**). Together, these analyses confirm that the impact of retrospective extinction on memory erasure cannot be explained by differences in baseline behavior or extinction dynamics.

Suppression of dopamine response at reward during retrospective extinction prevents memory erasure. Because retrospective extinction induced unlearning as predicted by ANCCR, we next tested ANCCR's additional postulate that dopamine gates retrospective learning at rewards. In ANCCR, memory erasure is counterintuitively hypothesized to occur not due to reward omission, but instead by receipt of rewards themselves, specifically those not preceded by CS-A. To test whether retrospective extinction of CS-A in G2 depends on dopamine gating at reward on CS-B trials, we optogenetically suppressed dopamine neuron activity at reward, while leaving dopamine transients unmanipulated during interleaved CS-A→omission trials (**Fig. 2A**, G2-opto). In the experimental group, an inhibitory opsin was expressed in VTA dopamine neurons and laser was delivered for 5 s following reward delivery on CS-B trials during *CS-A extinction*. Control animals received the same laser stimulation but without opsin expression. ANCCR predicts that suppressing dopamine activity at reward on CS-B trials should prevent retrospective updates and therefore block CS-A←reward memory erasure, allowing rapid reacquisition of CS-A similar to G1. In the absence of dopamine inhibition at reward, the control group should exhibit slower reacquisition, replicating the original G2 results (**Fig. 2B**). Reacquisition to CS-B was expected to remain rapid in both groups.

Simultaneous recording of dopamine release in the nucleus accumbens confirmed effective suppression of reward-evoked dopamine, with net responses reduced below baseline (**Fig. 2C**, **fig. S5L**). In ANCCR, such below-baseline suppression at the time of reward reduces behavior to both cues by lowering the estimated reward magnitude (as observed, **fig. S5M**) without degrading the retrospective association. Consistent with prevention of memory erasure, the experimental group reacquired CS-A behavior rapidly, whereas the control group showed slower reacquisition than acquisition, similar to the original G2 (**Fig. 2D–F**, **fig. S5B,C**). Likewise, in the experimental group, reacquisition speed of the dopaminergic response to CS-A was as rapid as that to CS-B, further supporting a blockade of retrospective extinction (**fig. S5D–I**). Notably, consummatory licking was not affected by optogenetic manipulation, indicating that suppression of dopamine responses at reward did not impair reward consumption but selectively disrupted retrospective learning triggered by rewards (**Fig. 2G**). Together, these findings support that behavioral extinction in the experimental group did not arise from degradation of stored cue-reward memory like in the control group or the original G2, but by suppressing behavioral output variables like G1 (estimated reward magnitude in the experimental group and reward rate in G1). They further support a causal role for dopamine activity at rewards themselves, on trials lacking CS-A, in driving CS-A←reward memory erasure. This gating role of dopamine is consistent with the emerging consensus among alternative dopamine models proposing learning rate control (41, 43–

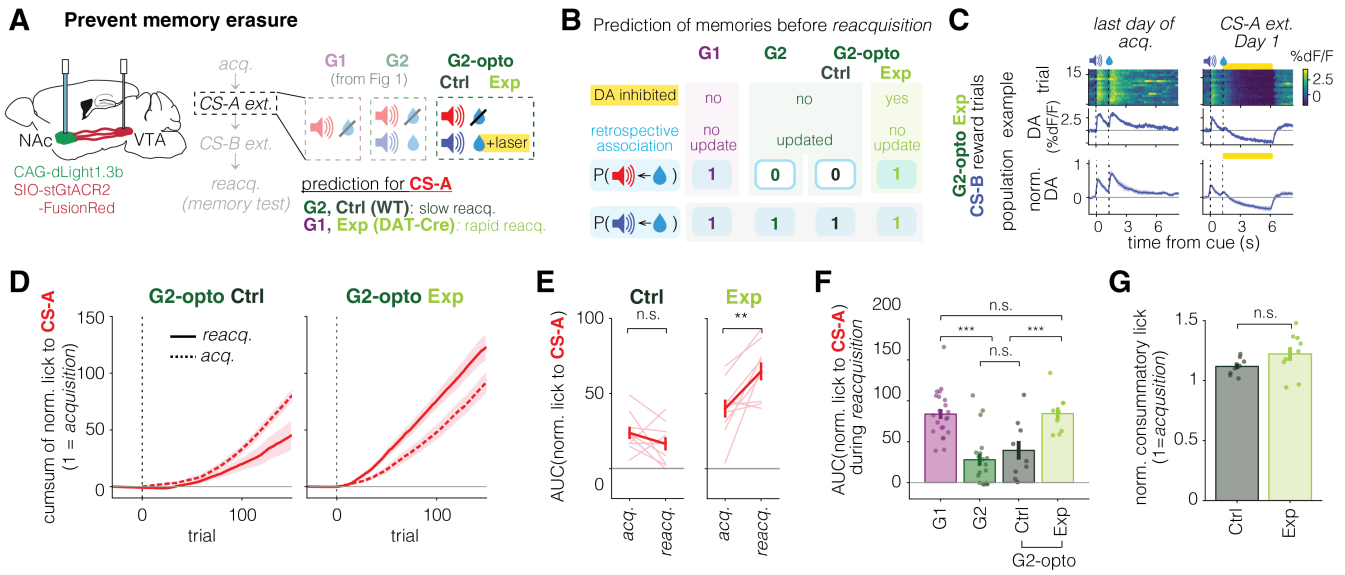

Fig. 2. Dopamine response at reward gates retrospective extinction. A, Experimental schematic. In the same CS-A extinction paradigm used for G2, dopamine neurons in the ventral tegmental area (VTA) were optogenetically suppressed for 5 s following reward delivery on CS-B trials in the experimental group (Exp). The control group (Ctrl) underwent the same procedure without expression of an inhibitory opsin. **B**, Prediction. Dopamine is hypothesized to gate updates of the retrospective association occurring at the time of reward. In the Exp group, suppressing dopamine responses during CS-B rewards prevents degradation of the CS-A←reward memory, resulting in rapid reacquisition of CS-A, similar to G1. In contrast, the Ctrl group is expected to resemble G2, showing unlearning and slow CS-A reacquisition. **C**, Example heatmaps and peri-stimulus time histograms (PSTHs) of dopamine release ($\Delta F/F$) in the nucleus accumbens during CS-B rewarded trials from a representative Exp animal (*top*) and group-averaged PSTHs across all Exp animals (*bottom*). The left column shows data from the last day of acquisition without laser stimulation, and the right column shows data from the first day of CS-A extinction with laser delivered on CS-B reward trials. Yellow bars indicate laser-on periods. Dopamine responses to reward following CS-B were strongly suppressed during the CS-A extinction phase compared to the last day of acquisition. **D**, Cumulative sum of normalized lick responses to CS-A across trials during acquisition (dotted) and reacquisition (solid). In the Ctrl group (*left*), reacquisition is slow and resembles the pattern observed in G2 without optogenetic manipulation (see **Fig. 1J**). In contrast, the Exp group (*right*) shows rapid reacquisition. Lines represent means across animals and shaded regions indicate SEM. **E**, Comparison of the area under the curve (AUC) of normalized licking between acquisition and reacquisition for CS-A. Light-colored lines represent individual animals. **F**, Comparison of AUC of normalized licking to CS-A during reacquisition with the original Groups 1 and 2 without optogenetic manipulation (**Fig. 1**), showing similarity of Ctrl to G2 and of Exp to G1. Error bars represent SEM across animals. G1 and G2 results are from the same animals shown in **Fig. 1**. **G**, Mean normalized consummatory licking during CS-A extinction, showing no difference between groups. Error bars represent SEM across animals, and the value of 1 indicates the average consummatory licking over the last three days of acquisition.

45), and supports the core postulate of reward-triggered retrospective learning in ANCCR.

Latent cause models do not explain memory erasure by retrospective extinction. Results from G1 and G2 (**Fig. 1**) and from causal dopamine inhibition in G2-opto (**Fig. 2**) are consistent with simulations of ANCCR (**fig. 3A–C**). In latent cause models, whether memory is erased or maintained depends on which latent context is inferred during extinction. This inference is implemented in many ways across distinct versions of latent cause models (**fig. S6A**). For example, the model by Redish et al. (2007) (39) infers a single discrete context on each trial based on the similarity between current observations—such as sensory cues and recent reward history—and the learned context prototypes. In this framework, persistent negative RPEs promote the inference of a new state by increasing sensitivity to the current observations (which likely contain the clues regarding the identity of the new state; **fig. S6B**). By contrast, Bayesian latent cause models (36, 37) represent context inference probabilistically, maintaining a graded belief over multiple contexts on each trial and updating these beliefs via posterior inference. In this setting, both positive and negative prediction errors can favor inference of a new context by reducing the posterior probability of existing contexts (**fig. S6C,D**). In addition to latent cause models, a recent RPE-based model termed "value RNN" (46) learns to identify and encode task structure implicitly through the dynamics of a recurrent neural network, and learns the value of cues in different contexts without explicit context inference.

Can any of these models capture the experimental results anticipated by retrospective learning? Bayesian versions of the latent cause model (36, 37) and value RNN (46) failed to account for the results from G1 and G2 (**fig. S6N–U**, **fig. S7**), and we therefore did not further evaluate their predictions for G2-opto. By contrast, the Redish et al. model (39) was compatible with the results from G1 and G2. This model correctly predicted that CS-A memory would be lost in G2 but not in G1 (**Fig. 3D–F**, **fig. S6E–J**). In this model, the presence of CS-B trials during *CS-A extinction* in G2—resembling *Acquisition*—increases the probability of inferring the acquisition context ($\text{context}_{\text{Acq}}$) instead of the extinction context ($\text{context}_{\text{Ext}}$). In G1, where extinction occurs without CS-B trials, the model infers $\text{context}_{\text{Ext}}$, preserving memory. This difference in context inference produces memory erasure in G2 but not in G1, matching behavioral outcomes (**Fig. 3D–F**, **fig. S6E–J**). However, this model fails to explain the results from G2-opto (**Fig. 3D–F**, **fig. S6M**). Nevertheless, because this discrepancy could stem from implementation details rather than core model principles (**Supplementary Note 3**), we next designed experiments to directly discriminate ANCCR from the core computations underlying all versions of latent cause models.

The core tenet of all latent cause models is that the

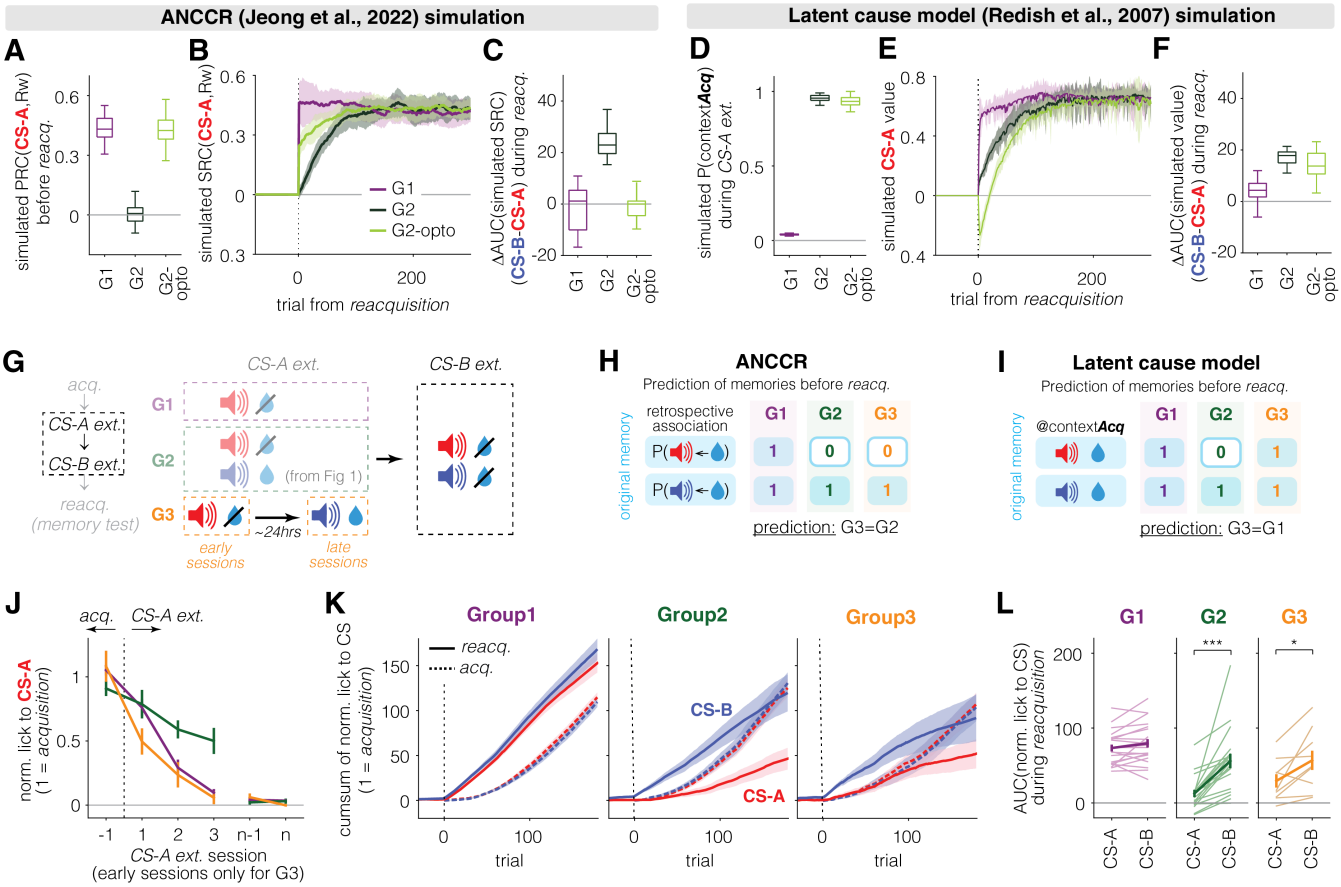

Fig. 3. Latent cause model cannot explain slower reacquisition following retrospective extinction A–C. ANCCR simulations for G1, G2, and G2-opto. **A.** Simulated retrospective association (predecessor representation contingency, PRC) between CS-A and reward after CS-A extinction. Higher value implies preserved memory, while lower value reflects unlearning. **B.** Simulated prospective association (successor representation contingency, SRC) between CS-A and reward during reacquisition. It captures the behavioral results. ANCCR predicts that dopamine inhibition at reward following CS-B (G2-opto) would cause G2 to behave like G1, consistent with the empirical results (**Fig. 2**). The lower rise of SRC (and predicted behavior) in G2-opto is because the previously experienced non-zero reward rate in G2 (which was during CS-A extinction) is half that of G1 (which was during Acquisition). **C.** Difference in AUC of simulated SRC between CS-A and CS-B during reacquisition. **D–F.** Latent cause model (Redish et al., 2007; see **fig. S6–7** for other latent cause models) simulations for G1, G2 and G2-opto. **D.** Simulated probability of inferring the acquisition context (contextAcq) during CS-A extinction. Closer to zero indicates switching to a new context (memory preserved), while higher values indicate inference of contextAcq (unlearning). In box plot, the central line indicates the median of iterations; box edges mark the 25th and 75th percentiles; whiskers extend to the most extreme data points within 1.5× the interquartile range. **E.** Simulated CS-A value across trials during reacquisition can capture behavioral results of G1 and G2. However, latent cause model predicts that dopamine inhibition at reward following CS-B (G2-opto) would not speed up behavior in G2, which contradicts the empirical results (**Fig. 2**). Lines indicate mean and shadings represent STD across iterations. Negative CS-A value early in Reacquisition is due to the negative RPE from dopamine inhibition in G2-opto. **F.** Difference in AUC of simulated values between CS-A and CS-B during reacquisition, reflecting the predicted cue-dependent difference in reacquisition speed. **G.** Task schematic for CS-A extinction and CS-B extinction. Group3 (G3) is similar to G2, but CS-A and CS-B trials were separated into distinct sessions across days during CS-A extinction. CS-B extinction, consisting of omission trials for both CS-A and CS-B, was identical across groups. **H.** ANCCR predicts G3 would behave like G2, since CS-B rewarded trials degrade the retrospective association between CS-A and reward, leading to unlearning, even though CS-B reward trials happen more than two days after behavior extinction for CS-A. **I.** Latent cause model predicts that G3 would behave like G1, since animals infer a new context during CS-A omission trials (early sessions of CS-A extinction), preserving original memory in contextAcq. **J–L.** Behavioral results support ANCCR. G1 and G2 results are from the same animals shown in **Fig. 1**. **J.** Average lick to CS-A in each session during CS-A extinction phase. Day n indicates the last day of CS-A extinction. For G3, only early sessions, those identical to CS-A extinction phase in G1, were included in analysis. **K.** Group-averaged cumulative sum of normalized lick to CS-A (red) and CS-B (blue) during acquisition (dotted) and reacquisition (solid). **L.** AUC of normalized lick to CS-A and CS-B during reacquisition. CS-B reacquisition is faster than acquisition in all groups, but CS-A reacquisition is faster only in G1 and not in G2 or G3.

hidden context inferred *at the time of behavior extinction* determines whether to maintain or degrade memory. In contrast, degradation of memory occurs *at the time of rewards* in ANCCR. To fully distinguish between these update principles for ANCCR and all latent cause models, we designed Group 3 (G3). In G3, *CS-A extinction* phase was temporally separated across sessions: the first 5 daily sessions ('early sessions') consisted solely of CS-A omission trials (like G1) [$A \rightarrow$ omit], whereas the subsequent daily sessions ('late sessions') included only CS-B trials with 50% reward probability [$B \rightarrow 50\%$sucrose] (**Fig. 3G**). Latent cause models predict that G3 will behave like G1, because early CS-A-only extinction sessions, which are identical to G1, should trigger context switching and preserve CS-A memory (**Fig. 3I**, **fig. S8A**). Because G3 undergoes the exact same CS-A extinction as G1 during early sessions, subsequent late sessions of CS-B trials should not thereafter degrade the original CS-A-reward memory. Conversely, ANCCR predicts that subsequent rewarded CS-B trials provide opportunities for retrospective extinction of the CS-A←reward, causing memory erasure as in G2 (**Fig. 3H**, **fig. S8B**). In sum, latent cause models predict that G3 will behave like G1 while ANCCR instead predicts that G3 will behave like G2. Thus, this design discriminates the theoretical frameworks by testing whether rewards delivered >2 days after complete behavior extinction of CS-A (**Fig. 3J**) retrospectively erase

the preserved CS-A memory. This is therefore a test of the central principle of reward-triggered retrospective learning.

Behavioral results strongly favor ANCCR over latent cause models. Within each group, we compared CS-A and CS-B reacquisition speeds, using CS-B as an internal control since it underwent identical standard extinction across all groups and should show uniformly rapid reacquisition under both theoretical frameworks (**Fig. 3K,L**). We compared behavioral results from G3 to those from G1 and G2. G1 animals reacquired both cues rapidly with no speed difference, consistent with preserved memory for both cue-reward associations. In contrast, both G2 and G3 exhibited selective impairment. CS-B reacquisition remained rapid, whereas CS-A reacquisition was significantly slower. Importantly, no cue differences were found during initial acquisition across any group (**fig. S8C**), confirming that reacquisition differences are due to extinction manipulations instead of variations in baseline cue salience. Together, these results contradict the core tenet of latent-cause models that persistence of the original memory after extinction is determined by the latent context inferred at the time of extinction. Instead, they support the ANCCR prediction that rewards, even if experienced more than 2 days after behavioral extinction, can retrospectively erase a preserved cue–reward memory.

To further test whether rewards alone are enough for retrospective extinction, we next tested Group4 (G4), in which the second phase of *CS-A extinction* (like in G3) included uncued rewards instead of CS-B trials. Under ANCCR, this manipulation should produce non-selective retrospective extinction degrading all reward-associated memories (**fig. S8D,E**). Consistent with this, G4 animals showed slow reacquisition for both CS-A and CS-B, supporting ANCCR's core idea that reward triggers unlearning (**fig. S8F-K**). Further, experiments in freely moving rats confirmed that retrospective degradation is reward identity specific and persistent. When different reward types (sucrose vs. food pellet) were associated with different cues, retrospective degradation selectively impaired recovery only for the cue associated with the degraded reward identity (**fig. S8L,M**). Moreover, retrospective extinction effects persisted even when followed by standard extinction, with no evidence of behavioral recovery (**fig. S8N-P**), contrasting earlier pigeon autoshaping results from a similar experiment (47, 48).

Altogether, we demonstrate cue-specific retrospective memory erasure in G2 and G3, where CS-B-reward trials degrade CS-A retrospective association. By contrast, reacquisition rate was similar for both cues when mesolimbic dopamine activity was suppressed at rewards in G2-opto. Further, CS-A retrospective extinction does not require the reward to be preceded by CS-B; it also works when the reward is uncued (**fig. S8D-M**), and when there is only a single cue in the paradigm (shown later in **Fig. 4**). Collectively, these findings show that retrospective extinction is a robust and effective mechanism for memory erasure via reward-driven updates, providing strong empirical support for ANCCR over latent cause models.

Mesolimbic dopamine signaling in contingency degradation and extinction is consistent with ANCCR. Our findings further shed light on ongoing debates regarding dopamine's role in learning, contrasting prospective reward prediction error (RPE)-based frameworks (including latent cause models) and retrospective learning models such as ANCCR. Here, we focus on Qian et al. (49) as a representative example of this broader debate, because their task was specifically designed to test between prospective and retrospective computations during contingency degradation. Qian et al. (49) reported that contingency degradation due to uncued rewards is stronger than that due to cued rewards, arguing for prospective computations over ANCCR's retrospective reward-driven updating (**fig. S9I**).

Though the task design is elegant, this interpretation is premature as Qian et al. do not consider the possibility of cross-cue generalization (**Supplementary Note 2**). When multiple cues precede the same reward, the cues can acquire a shared representation (cross-cue generalization or "acquired equivalence") (50–52). Cross-cue generalization can protect an established cue–reward association from degradation, because when every reward remains preceded by some cue (the "cued-reward" condition), a generalized cue←reward association persists and slows degradation of the original cue. In contrast, uncued rewards degrade both cue-specific and generalized associations, accelerating contingency degradation (**fig. S9J**). Incorporating cross-cue generalization into ANCCR reconciles the findings of Qian et al. (49). Evidence for such generalization is present in their own data—specifically, the rapid acquisition of responses to the second cue—and also in our experiments, where it manifests as slower extinction of CS-A behavior and dopamine responses in G2 relative to G1 (**fig. S3**, **fig. S9A–C**). Nevertheless, animals in G2 ultimately overcome this generalization, extinguishing conditioned behavior to CS-A while maintaining it to CS-B, producing distinct cue-specific reacquisition rates. Thus, cross-cue generalization transiently influences extinction but does not prevent the cue-specific memory erasure predicted by ANCCR.

Finally, RPE-based models including Qian et al. (49) advocated the use of an "unbiased" value RNN to infer latent states without hand-crafted features. In our dataset, however, the same value-RNN framework fails to capture the CS-A reacquisition-rate difference between G1 and G2 (**fig. S7F–H**) and their extinction-rate difference (**fig. S9F–H**), although incorporating cross-cue generalization could recover the latter effect. These results demonstrate that neither ANCCR nor RPE-based models naturally account for known cognitive processes such as cross-cue generalization (50–52). Once cross-cue generalization is incorporated, the set of results presented here and in Qian et al. (49) is better explained by retrospective learning.

Additionally, our framework addresses recent counter-intuitive findings by Burwell et al. (53) that elimination of dopamine omission dip accelerates extinction, contrary to RPE accounts (**fig. S9K**). Classical or latent cause models using RPE predict that large omission-evoked dopamine dips hasten extinction by driving either value reduction or context switching (**fig. S9L**, left/middle). In ANCCR, positive dopamine transients promote learning of an event's cause, whereas negative dopamine transients

suppress it. Therefore, ANCCR instead predicts that the omission dip suppresses formation of cue←omission (i.e., cue←frustration) association, acting as a compensatory brake on behavioral extinction. Removing this brake therefore accelerates behavioral extinction, aligning with Burwell et al.'s observations (**fig. S9L**, right; **fig. S9M**). Consistent with this, we observed that G2 showed slower extinction despite a persistent dopamine omission dip (**fig. S9D,E**). In addition, larger omission dips were associated with slower extinction across individual animals (**fig. S9N**). Moreover, animals with smaller dopamine dips showed weaker asymptotic behavior after reacquisition, consistent with competition between preserved cue←reward memory and a newly formed cue←frustration memory (**fig. S9O**). These results support ANCCR's interpretation that dopamine dips act as a protective mechanism, preventing formation of conflicting or excessive associations (**fig. S9P**). To summarize, within ANCCR, dopamine release at reward acts as a gate that determines when retrospective cue←reward associations are updated, whereas omission-evoked dopamine dips serve as a protective brake that suppresses formation of competing cue← frustration memories. This contrasts with RPE-based views in which both reward and omission signals drive value updating or context switching. Burwell et al.'s findings are consistent with ANCCR but not RPE or latent cause models. Overall, behavior and dopamine dynamics during and after extinction are best understood within a retrospective learning framework (**Supplementary Note 4**).

Retrospective extinction is effective across outcome modalities. To evaluate the generalizability of retrospective extinction, we tested whether cue-drug memories could also be retrospectively extinguished (**Fig. 4A**). Because the earlier task designs, developed to discriminate the models, strongly support ANCCR over latent cause models, here we used a simpler design to test whether retrospective extinction is effective across different outcome modalities. We used two auditory cues: a conditioned stimulus (CS+) paired with outcome (alcohol), and a control cue (CS-) never associated with the outcome [$\mathrm{CS+} \rightarrow 100\%\mathrm{alcohol}; \mathrm{CS-} \rightarrow \mathrm{nothing}$]. During extinction, G1 underwent standard extinction with omission of outcome after CS+ [$\mathrm{CS+} \rightarrow \mathrm{omit}; \mathrm{CS-} \rightarrow \mathrm{nothing}$], whereas G2 experienced retrospective extinction, receiving unpredicted outcomes during inter-trial intervals to trigger retrospective degradation of cue-outcome association [$\mathrm{CS+} \rightarrow \mathrm{omit}; \mathrm{CS-} \rightarrow \mathrm{nothing}; \mathrm{ITI} \rightarrow \mathrm{alcohol}$]. Conditioned responses were quantified using selectivity for CS+ over CS- while controlling for baseline differences across groups. G1 showed rapid reacquisition of anticipatory licking to CS+, whereas G2 displayed markedly delayed reacquisition (**Fig. 4B,C**; **fig. S10A-E**). Next, we tested whether retrospective extinction holds preliminary promise for reducing drug usage in humans. Preliminary human data indicate that cocaine exposure in treatment settings outside of drug-associated contexts reduced subsequent self-reported cocaine use measured in the following year (**Fig. 4D**). This suggests translational potential for retrospective

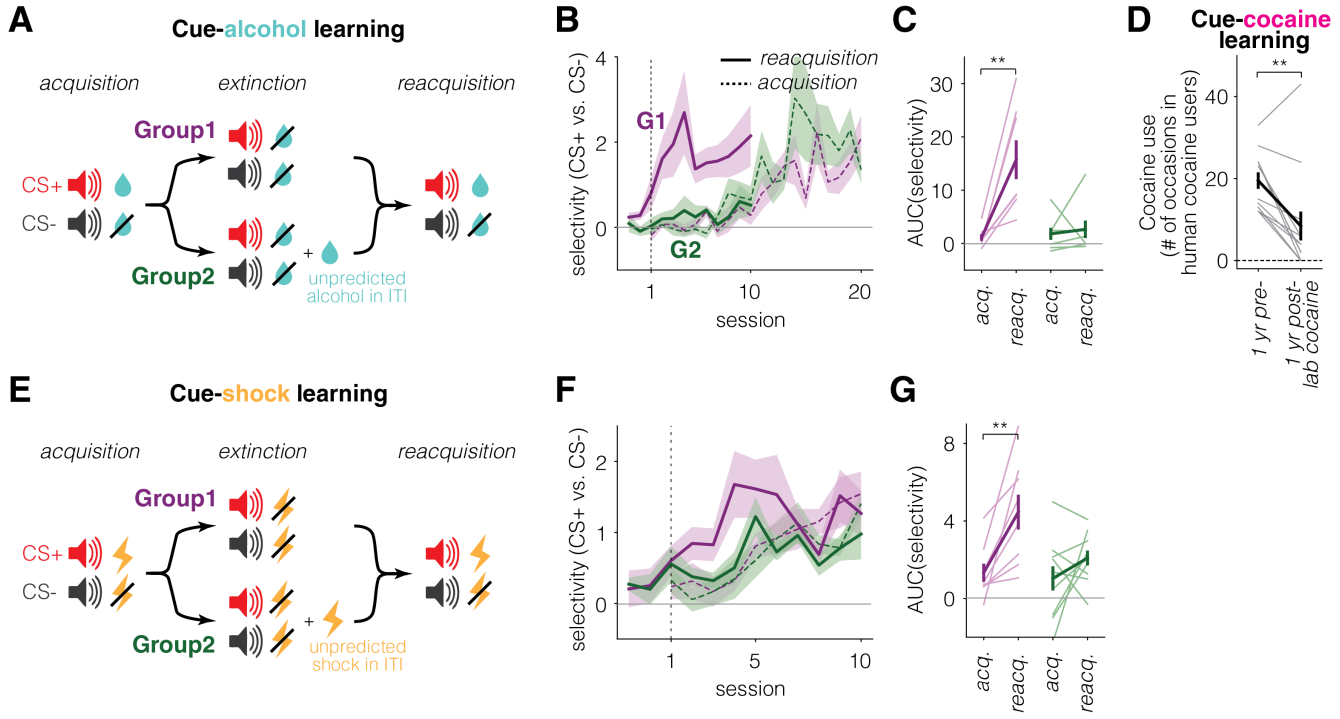

Fig. 4. Retrospective extinction erases memory across outcome modalities A–C. ANCCR simulations for G1, G2, and G2-opto. **A.** Schematic of cue-alcohol learning. Animals were trained to associate CS+ with alcohol and CS- with no alcohol. During extinction, alcohol following CS+ was omitted without (G1; standard extinction) or with (G2; retrospective extinction) additional uncued alcohol during ITI. This paradigm predicts that the original memory is preserved in G1, while unlearned in G2. **B.** Group-averaged selectivity (CS+ vs. CS-) based on anticipatory lick during acquisition (dotted) and reacquisition (solid). G1, but not G2, shows faster reacquisition than acquisition. Lines indicate mean across animals and shadings represent SEM. **C.** AUC of selectivity over first 10 days of acquisition or reacquisition. Light lines represent individual animals. Error bars indicate SEM across animals. **D.** Human recreational cocaine users ingested cocaine in an atypical environment, i.e., a hospital research setting such as a brain imaging unit. During a follow-up interview, they reported their cocaine use during the preceding 12 months. **E.** Schematic of cue-shock learning. Similar task structure as cue-alcohol learning, but with shocks instead of alcohol. On each trial, cue lasted for 30 s to allow a sufficient window for immobility analysis. **F.** Group averaged selectivity (CS+ vs. CS-) based on anticipatory immobility during acquisition (dotted) and reacquisition (solid). G1, but not G2, shows rapid reacquisition. **G.** AUC of selectivity over first 5 days of acquisition or reacquisition.

extinction approaches and calls for further investigation in controlled clinical trials. We finally tested whether these results also extend to cue-shock learning in mice with the same design as the cue-alcohol paradigm while measuring immobility as the conditioned response. G1 rapidly regained CS+ selective immobility, while G2 CS+ selectivity emerged only after extended training (**Fig. 4E-G**; **fig. S10F-L**). Together, these findings suggest that retrospective extinction induces unlearning across diverse outcomes and species, with implications for therapies targeting relapse in addiction and trauma-related disorders.

Disrupting frontal cortical activity suppresses rapid reacquisition. Although retrospective extinction provides a robust framework for memory erasure in experimental settings, it is constrained by a central limitation: its effectiveness depends on exposure to the outcome. In clinical contexts, however, there can be ethical and safety considerations that preclude re-exposing individuals to traumatic events or addictive substances in attempts to erase maladaptive associations. This practical barrier motivates the need for alternative approaches capable of limiting relapse without requiring outcome re-exposure as during retrospective extinction. In other words, this motivates the need for therapeutic approaches that limit reacquisition of conditioned behavior even after prospective extinction.

To address this, we investigated whether broad activity disruption in line with potential human neuromodulation therapies could prevent relapse. Specifically, we focused on the orbitofrontal cortex (OFC), a region driving maladaptive persistence of behavior (54, 55), particularly in addiction (56–60). As OFC is functionally heterogeneous along its medial-lateral axis (61–63), we focused on the ventral/medial OFC because a neuronal subpopulation in the ventral/medial OFC potentially encodes retrospective associations in its cue response (42, 64), and encodes reward salience in its reward response (65). A slowing of reacquisition could arise through two possible mechanisms during the Reacquisition phase, either by degrading the stored retrospective memory (**fig. S11A, B**), or by impairing rapid reinstatement of the estimated reward rate needed for behavioral expression (**Fig. 5A, B**). This latter prediction is consistent with the encoding of cognitive maps, outcomes, and value in OFC as a whole (65–73), and the encoding of reward salience in ventral/medial OFC (65).

To test whether suppressing of ventral/medial OFC during either of these periods can prevent rapid reacquisition, we trained animals with Acquisition of a single cue-reward association followed by standard prospective Extinction and then selectively inhibited OFC during Reacquisition either in the cue-reward period or a 5 s post-reward period. Optogenetic inhibition of OFC during the reward period markedly slowed

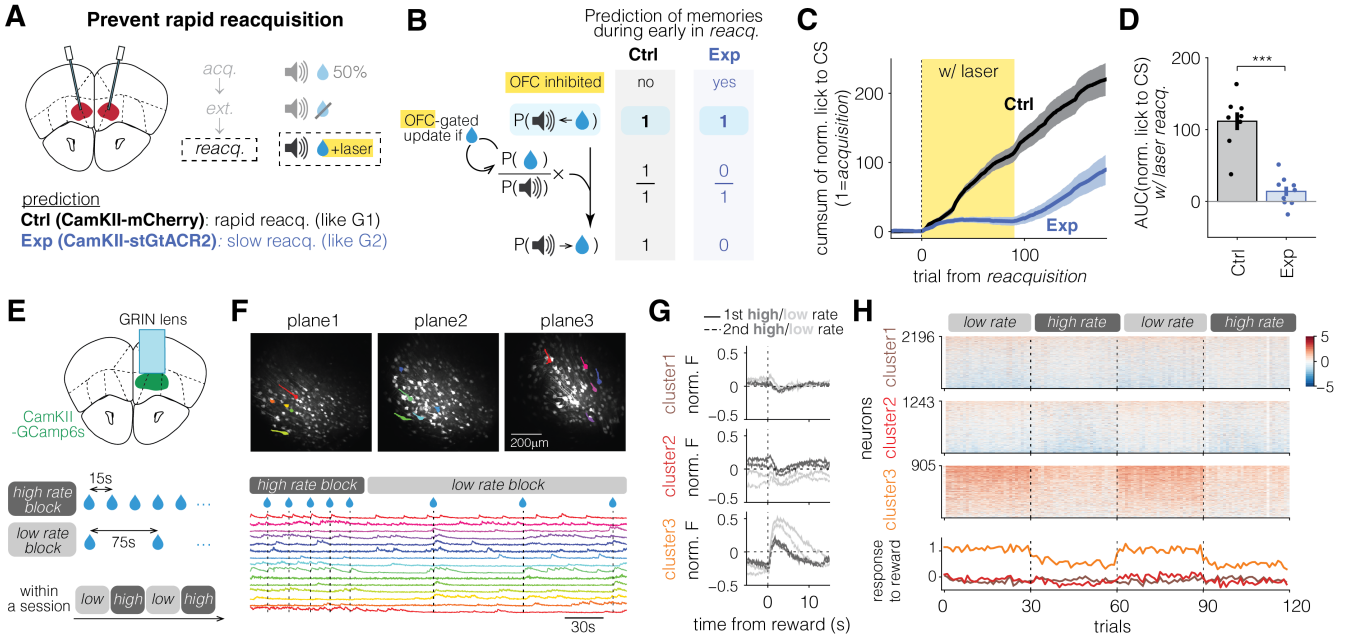

Fig. 5. Suppressing orbitofrontal cortex (OFC) can prevent rapid reacquisition. A–D. Suppressing orbitofrontal cortex (OFC) can prevent rapid reacquisition. **A.** Experimental schematic. After acquisition of single cue-reward association, standard extinction and reacquisition were followed in a sequence. During reacquisition following standard extinction, OFC CamKIIα-expressing neurons were optogenetically suppressed for 5 s following reward delivery in experimental group (Exp). The control group (Ctrl) underwent the same procedure without expression of inhibitory opsin. Note that behavioral outcome (anticipatory licks) is measured during a non-overlapping period with the laser (post-reward). **B.** Hypothesis. If OFC controls the update of P(reward) during reacquisition, its suppression in Exp should block the update and delay reacquisition, while Ctrl should reacquire rapidly. **C.** Group-averaged cumulative sum of normalized lick to CS across trials. Yellow shading marks the laser-on trials. **D.** AUC of normalized lick to CS during laser-on trials of reacquisition. Learning is suppressed in Exp. **E–H.** OFC encodes estimated reward rate. **E.** Experimental schematic. Uncued rewards were delivered at either short (mean 15 s, high rate block) or long (mean 75s, low rate block) inter-reward intervals, with reward rate switching three times within a session. Activity of OFC CamKIIα-expressing neurons were recorded through a GRIN lens. **F.** Example fields of view across three simultaneously recorded imaging planes (top) and fluorescence traces of representative neurons (bottom). Vertical dashed lines indicate reward delivery. **G.** PSTHs of clustered neurons (see Methods) averaged across the last three trials of each block. Each reward rate includes two blocks (1st block: solid, 2nd block: dashed). PSTHs show averaged normalized fluorescence across all neurons within each cluster. Cluster 2 primarily modulates baseline activity with reward rate, whereas cluster 3 modulates reward-evoked phasic response. Note that the higher baseline activity of cluster 2 during high reward rates cannot be trivially explained by slowly decaying reward responses measured by GCaMP6s fluorescence because phasic reward responses are negative in this cluster. **H.** Phasic reward responses of individual neurons across trials (top) and averaged response across neurons within each cluster (bottom). Phasic response to reward was quantified as the difference in area under curve of fluorescence traces 5 s before and after the first lick following reward delivery.

reacquisition, with behavioral suppression persisting until the intervention ended (**Fig. 5C,D**; **fig. S11L**). Importantly, consummatory licking was unaffected during laser stimulation (**fig. S11M**), indicating that slowed reacquisition reflected impaired learning, not reduced motivation or motor capacity. In contrast, suppressing OFC during cue-reward delay acted as a novel, reward-predictive cue, compensating for and preventing the ability to investigate any degraded cue-outcome memory (**fig. S11D-K**). In a separate cohort, inhibition of OFC during the outcome period of Acquisition slowed initial learning (**fig. S11N**), consistent with the idea that OFC activity at reward delivery contributes to estimating the current reward rate.

Because testing such reward rate encoding is difficult during reacquisition itself (e.g., G1 and G2 from **Fig. 1** have similar reward rates during reacquisition), we designed an experiment in which animals experienced different (uncued) reward rates. During this condition, we performed two-photon calcium imaging of OFC neurons to directly test whether OFC neurons encode reward rate immediately after reward delivery (i.e., the trial period during which OFC inhibition abolished behavioral reacquisition). Each session contained four alternating blocks of high rate (mean interval of 15 s) or a low rate (mean interval of 75 s) of rewards, with 30 rewards per block (**Fig. 5E**). The experiment was repeated across two sessions with counterbalanced block orders, yielding consistent results across days (**Fig 5G,H** and **fig. S12E-G** from Day 2; Day 1 shown in **fig. S12H-J**). Across three imaging planes in each of four animals, we recorded activity from ~4,000 OFC neurons (**Fig. 5F**; **fig. S12A**). Licking behavior was comparable across blocks (**fig. S12D**, left), yet many neurons differentiated reward rate either through changes in baseline activity or phasic reward response (**fig. S12D**, right). Spectral clustering (74) identified three distinct neuronal subpopulations (**fig. S12B,C**) including one whose phasic reward responses rapidly adjusted to the current reward rate (Cluster3; **Fig. 5G,H**; **fig. S12E-J**). This population may encode a direct estimate of reward rate or, alternatively, a memory of reward rate given reward (i.e., $P(\text{reward} \mid \text{reward})$)—a statistic capturing how likely rewards are to recur once one is encountered. Such a prior would allow animals to rapidly restore their internal estimate of reward rate, and thus behavior, when they return to the reward-rich environment after prolonged reward absence (e.g., extinction→reacquisition; home cage→training). Together, these results demonstrate that OFC activity during outcome delivery is essential for reinstating extinguished behaviors, suggesting that targeted neuromodulation of OFC activity could curb cue-induced relapse without requiring direct outcome re-exposure during extinction.

Discussion

Our findings identify a fundamental principle governing whether extinction suppresses behavior versus truly erases memory. By comparing standard, cue-based ("prospective") extinction with outcome-based ("retrospective") extinction, we show that unlearning occurs only when outcomes are experienced without their predictive cues. This retrospective extinction selectively degrades cue←outcome associations and is gated by dopamine release at outcome. Orbitofrontal cortex activity is necessary for rapid reacquisition after prospective extinction. Together, these results unify behavioral, circuit, and computational perspectives to reveal how dopamine-gated retrospective learning enables genuine memory erasure—reconceptualizing unlearning as a process governed by outcome-triggered, rather than cue-driven, updating of associative memories.

A reduction in conditioned behavior does not necessarily indicate loss of the underlying memory (12). Within the ANCCR framework, diminished responding can arise from degradation of three computational variables: cue←outcome memory, estimated outcome rate, and estimated outcome magnitude (**fig. S13A,B**). Only degradation of the former produces unlearning. Thus, findings traditionally interpreted as unlearning may instead reflect distinct mechanisms. For example, weakened relapse following gradual reduction of outcome frequency (14, 15, 22) (**fig. S13C,D**) or magnitude (16, 75) (**fig. S13E,F**) can be explained by a lowered prior of outcome rate or magnitude (Supplementary Note 5). Specifically, learning that outcome rate is lower or that its magnitude is reduced may be sufficient to suppress behavior even though the original memory of the cue-outcome association is still intact (**fig. S13C–F**). CS-retrieval extinction, in which the cue is presented by itself (and believed to open a reconsolidation window) before repeated extinction reduces relapse (17, 20, 21, 23, 24) (but see (25–31)). Though not a core prediction of ANCCR, this can be understood as altered inference of a physical context rather than memory degradation. A physical context is nothing more than a set of neutral cues: altering the distribution of neutral cues may lead to inference of a new physical context. Therefore, introducing a temporal gap between the retrieval cue and extinction (often much longer than acquisition inter-cue intervals) and thereby altering the inter-cue intervals expected for a physical context, may cause animals to infer a new context, thereby reducing relapse by limiting access to the original memory (**fig. S14A,B**). This view preserves the role of context-indexed learning and memory like the latent cause theory, though suggesting that statistics of neutral cues, instead of their outcome associations, dominate context inference (**Supplementary Note 5**). Reduced relapse in a paradigm with reversed retrieval and extinction trials (76, 77) and with varied ITI during extinction (78, 79) support this interpretation. By contrast, outcome-retrieval reconsolidation represents a rare case consistent with genuine memory degradation. An uncued outcome triggers retrospective updating that weakens all cue←shock (18) or cue←drug (19) associations, consistent with the experimentally observed reduction in relapse across all cues linked to that outcome (**fig. S14C–E**). Overall, relapses are best understood as shaped by interactions between associative memory persistence and modulation of behavioral output variables, a perspective further supported by our OFC findings (**Fig. 5**). This nuanced view broadens the implications for developing therapeutic strategies.

Beyond clarifying when extinction produces true unlearning, our results also identify a neural mechanism gating this computational outcome. In doing so, they refine current

views of mesolimbic dopamine function. Prior work showed that mesolimbic dopamine conveys causal associations (41, 80) but left open whether and how these signals control memory persistence versus erasure. By directly studying standard (prospective) extinction, retrospective extinction, and selective manipulations of dopamine at reward, we find that reward-evoked dopamine transients gate updates of retrospective cue←reward associations. By contrast, our findings, along with those of Burwell et al. (53), suggest that omission-evoked dips primarily act as a brake on forming competing cue←frustration memories. Thus, both positive and negative dopamine transients participate in the same gating process predicted by ANCCR: positive transients promote learning of an event's cause, whereas negative transients suppress it. This division of labor explains how dopamine can regulate whether extinction leads to true unlearning or mere behavioral suppression, providing a mechanistic link between computational processes and therapeutic opportunities.

Our results highlight a key distinction for therapeutic interventions in maladaptive memory-driven conditions such as addiction or PTSD: true degradation of the underlying cue–outcome memory versus suppression of its behavioral expression. Genuine erasure of a retrospective association requires exposure to the outcome itself, as in retrospective extinction. By contrast, although modulation of behavioral variables downstream of associative memory suppresses relapse, relapse risk remains so long as the underlying maladaptive association is preserved. For instance, while drug-free therapies leave drug-associations intact, supervised administration of drugs such as opioid agonists may degrade associations between opioid effects and drug-associated cues experienced in street settings (81–85). However, the efficacy of clinical implementations of retrospective extinction through drug exposure depends on overcoming one fundamental limitation. This is that treatment itself introduces novel contextual and social cues that are repeatedly experienced in the treatment setting. These novel cues, once paired with drug administration during treatment, can become potent relapse triggers, simply shifting the drug-related retrospective association from the original street cues to treatment-related cues. Thus, a retrospective association related to drug or trauma can only shift from one cue to another, thereby suggesting that shifting to cues rarely encountered in life might provide the best therapeutic approach to minimize relapse. Our findings therefore motivate systematic testing of strategies that deliberately shift maladaptive associations onto rarely encountered cues to minimize relapse risk, while carefully evaluating their practicality and safety in real-world clinical settings.

Though behavioral therapies to degrade maladaptive cue←outcome memories require outcome exposure, neuromodulation therapies may provide an alternate approach. Here we show that modulation of behavioral-expression variables such as estimated outcome rate encoding in OFC suppresses rate of reacquisition. This raises the possibility that if OFC is suppressed, the drug may still be consumed after prolonged abstinence without inducing drug-seeking driven by cues. Neuromodulation of outcome magnitude representations may be another approach. Finally, it may also be possible to simulate experience of uncued outcomes through neuromodulation. Thus, neuromodulation approaches provide complementary strategies to limit relapses when direct outcome re-exposure is not safe or feasible.

ACKNOWLEDGEMENTS

We thank M. Tadross, A. D. Redish, P. Dayan, G. Quirk, E. Harkin, B. Sabatini, S. Mihalas, L. Kusmierz, M. Kheirbek, M. Stryker, V. Sohal, K. Moussawi, J. Berke, H. Fields, D. Ron, and members of the Namboodiri laboratory for helpful discussions regarding the manuscript. We thank A. L. Cochran for discussions related to her latent cause model. This project was supported by NIH R01MH129582, R01AA029661, R01DA062018, the Scott Alan Myers Endowed Professorship, Alfred P Sloan Fellowship, Pew Biomedical Scholarship, Klingenstein-Simons Fellowship (VMKN), CIHR MOP–36429, PJT–175296 (ML), Klingenstein-Simons Fellowship, The David and Lucile Packard Foundation, and Shurl and Kay Curci Foundation (RC^2). We used LLMs to aid in editing for clarity. The authors have no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization: HJ, VMKN; Methodology: HJ, LZ, FF, ADS, SMLC, EG, PHJ, ML, RC2, VMKN; Investigation: HJ, LZ, FF, RC1, NB, MZ, ADS, SXW, AS, SMLC, EG, SB; Visualization: HJ, EG, SMLC, VMKN; Funding acquisition: ML, RC^2, VMKN; Project administration: HJ, VMKN; Supervision: VMKN; Writing – original draft: HJ, VMKN; Writing – review & editing: HJ, LZ, FF, RC^2, NB, MZ, ADS, SW, AS, SMLC, EG, SB, PHJ, ML, RC^2, VMKN. [1]Risha Chakraborty; [2]Ritchie Chen

Bibliography

1. A. T. McLellan, D. C. Lewis, C. P. O'Brien, and H. D. Kleber. Drug dependence, a chronic medical illness: implications for treatment, insurance, and outcomes evaluation. *JAMA*, 284 (13):1689–1695, October 2000. ISSN 0098-7484. doi: 10.1001/jama.284.13.1689.
2. Amy L. Milton and Barry J. Everitt. The persistence of maladaptive memory: Addiction, drug memories and anti-relapse treatments. *Neuroscience & Biobehavioral Reviews*, 36(4): 1119–1139, April 2012. ISSN 01497634. doi: 10.1016/j.neubiorev.2012.01.002.
3. Bram Vervliet, Michelle G. Craske, and Dirk Hermans. Fear Extinction and Relapse: State of the Art. *Annual Review of Clinical Psychology*, 9(Volume 9, 2013):215–248, March 2013. ISSN 1548-5943, 1548-5951. doi: 10.1146/annurev-clinpsy-050212-185542.
4. Mark E Bouton. Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biological Psychiatry*, 52(10):976–986, November 2002. ISSN 00063223. doi: 10.1016/S0006-3223(02)01546-9.
5. Andrew R. Delamater. Experimental extinction in Pavlovian conditioning: Behavioural and neuroscience perspectives. *The Quarterly Journal of Experimental Psychology Section B*, 57(2b):97–132, April 2004. ISSN 0272-4995, 1464-1321. doi: 10.1080/02724990344000097.
6. Mohammed R. Milad and Gregory J. Quirk. Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature*, 420(6911):70–74, November 2002. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature01138.
7. Karyn M. Myers and Michael Davis. Behavioral and Neural Analysis of Extinction. *Neuron*, 36(4):567–584, November 2002. ISSN 08966273. doi: 10.1016/S0896-6273(02)01064-4.
8. Renee M Napier, Michaela Macrae, and E James Kehoe. Rapid Reacquisition in Conditioning of the Rabbit's Nictitating Membrane Response.
9. Sean T. Ricker and Mark E. Bouton. Reacquisition following extinction in appetitive conditioning. *Animal Learning & Behavior*, 24(4):423–436, December 1996. ISSN 0090-4996, 1532-5830. doi: 10.3758/BF03199014.
10. Gabrielle Weidemann and E. James Kehoe. Savings in classical conditioning in the rabbit as a function of extended extinction. *Learning & Behavior*, 31(1):49–68, February 2003. ISSN 1543-4494, 1543-4508. doi: 10.3758/BF03195970.
11. Mark E. Bouton. Context and Behavioral Processes in Extinction: Table 1. *Learning & Memory*, 11(5):485–494, September 2004. ISSN 1072-0502, 1549-5485. doi: 10.1101/lm.78804.
12. Ivan P. Pavlov. Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Annals of neurosciences*, 17(3), June 2010. ISSN 09727531, 09763260. doi: 10.5214/ans.0972-7531.1017309.
13. Robert A. Rescorla. Spontaneous Recovery. *Learning & Memory*, 11(5):501–509, September 2004. ISSN 1072-0502, 1549-5485. doi: 10.1101/lm.77504.
14. Mark E. Bouton, Amanda M. Woods, and Oskar Pineño. Occasional reinforced trials during extinction can slow the rate of rapid reacquisition. *Learning and Motivation*, 35(4):371–390, November 2004. ISSN 00239690. doi: 10.1016/j.lmot.2004.05.001.
15. Samuel J. Gershman, Carolyn E. Jones, Kenneth A. Norman, Marie-H. Monfils, and Yael Niv. Gradual extinction prevents the return of fear: implications for the discovery of state. *Frontiers in Behavioral Neuroscience*, 7, 2013. ISSN 1662-5153. doi: 10.3389/fnbeh.2013.00164.
16. Nicholas G. W. Kennedy, Jessica C. Lee, Simon Killcross, Fred Westbrook, and Nathan M. Holmes. Prediction error determines how memories are organized in the brain: a study of Pavlovian fear extinction in rats, March 2024.
17. Jonathan L.C. Lee, Karim Nader, and Daniela Schiller. An Update on Memory Reconsolidation Updating. *Trends in Cognitive Sciences*, 21(7):531–545, July 2017. ISSN 13646613. doi: 10.1016/j.tics.2017.04.006.
18. Jianfeng Liu, Liyan Zhao, Yanxue Xue, Jie Shi, Lin Suo, Yixiao Luo, Baisheng Chai, Chang Yang, Qin Fang, Yan Zhang, Yanping Bao, Charles L. Pickens, and Lin Lu. An Unconditioned Stimulus Retrieval Extinction Procedure to Prevent the Return of Fear Memory. *Biological Psychiatry*, 76(11):895–901, December 2014. ISSN 00063223. doi: 10.1016/j.biopsych.2014.03.027.
19. Yi-xiao Luo, Yan-xue Xue, Jian-feng Liu, Hai-shui Shi, Min Jian, Ying Han, Wei-li Zhu, Yan-ping Bao, Ping Wu, Zeng-bo Ding, Hao-wei Shen, Jie Shi, Yavin Shaham, and Lin Lu. A

novel UCS memory retrieval-extinction procedure to inhibit relapse to drug seeking. *Nature Communications*, 6(1):7675, July 2015. ISSN 2041-1723. doi: 10.1038/ncomms8675.

20. Marie-H. Monfils, Kiriana K. Cowansage, Eric Klann, and Joseph E. LeDoux. Extinction-Reconsolidation Boundaries: Key to Persistent Attenuation of Fear Memories. *Science*, 324 (5929):951–955, May 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1167975.
21. Daniela Schiller, Marie-H. Monfils, Candace M. Raio, David C. Johnson, Joseph E. LeDoux, and Elizabeth A. Phelps. Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, 463(7277):49–53, January 2010. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature08637.
22. Amanda M. Woods and Mark E. Bouton. Occasional reinforced responses during extinction can slow the rate of reacquisition of an operant response. *Learning and Motivation*, 38(1): 56–74, February 2007. ISSN 00239690. doi: 10.1016/j.lmot.2006.07.003.
23. Yan-Xue Xue, Yi-Xiao Luo, Ping Wu, Hai-Shui Shi, Li-Fen Xue, Chen Chen, Wei-Li Zhu, Zeng-Bo Ding, Yan-ping Bao, Jie Shi, David H. Epstein, Yavin Shaham, and Lin Lu. A Memory Retrieval-Extinction Procedure to Prevent Drug Craving and Relapse. *Science*, 336 (6078):241–245, April 2012. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1215070.
24. Ahmet O. Ceceli, Sarah G. King, K. Rachel Drury, Natalie McClain, John Gray, Priyanthi S. Dassanayake, Jeffrey H. Newcorn, Daniela Schiller, Nelly Alia-Klein, and Rita Z. Goldstein. The synergy of methylphenidate- and reconsolidation-based extinction normalizes ventromedial prefrontal function in drug addiction. *Proceedings of the National Academy of Sciences*, 122(45):e2512310122, November 2025. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2512310122.
25. Alessia Auber, Vincenzo Tedesco, Carolyn E. Jones, Marie-H. Monfils, and Christian Chiamulera. Post-retrieval extinction as reconsolidation interference: methodological issues or boundary conditions? *Psychopharmacology*, 226(4):631–647, April 2013. ISSN 0033-3158, 1432-2072. doi: 10.1007/s00213-013-3004-1.
26. Wan Yee Macy Chan, Hiu T. Leung, R. Frederick Westbrook, and Gavan P. McNally. Effects of recent exposure to a conditioned stimulus on extinction of Pavlovian fear conditioning. *Learning & Memory*, 17(10):512–521, October 2010. ISSN 1549-5485. doi: 10.1101/lm.1912510.
27. Laura Luyten and Tom Beckers. A preregistered, direct replication attempt of the retrieval-extinction effect in cued fear conditioning in rats. *Neurobiology of Learning and Memory*, 144:208–215, October 2017. ISSN 10747427. doi: 10.1016/j.nlm.2017.07.014.
28. Marco Costanzi, Sara Cannas, Daniele Saraulli, Clelia Rossi-Arnaud, and Vincenzo Cestari. Extinction after retrieval: Effects on the associative and nonassociative components of remote contextual fear memory. *Learning & Memory*, 18(8):508–518, August 2011. ISSN 1549-5485. doi: 10.1101/lm.2175811.
29. Kate Hutton-Bedbrook and Gavan P. McNally. The Promises and Pitfalls of Retrieval-Extinction Procedures in Preventing Relapse to Drug Seeking. *Frontiers in Psychiatry*, 4, 2013. ISSN 1664-0640. doi: 10.3389/fpsyt.2013.00014.
30. M. Alexandra Kredlow, Leslie D. Unger, and Michael W. Otto. Harnessing reconsolidation to weaken fear and appetitive memories: A meta-analysis of post-retrieval extinction effects. *Psychological Bulletin*, 142(3):314–336, March 2016. ISSN 1939-1455, 0033-2909. doi: 10.1037/bul0000034.
31. Merel Kindt and Marieke Soeter. Reconsolidation in a human fear conditioning study: A test of extinction as updating mechanism. *Biological Psychology*, 92(1):43–50, January 2013. ISSN 03010511. doi: 10.1016/j.biopsycho.2011.09.016.
32. Youssef Shiban, Jasmin Wittmann, Mara Weißinger, and Andreas Mühlberger. Gradual extinction reduces reinstatement. *Frontiers in Behavioral Neuroscience*, 9, September 2015. ISSN 1662-5153. doi: 10.3389/fnbeh.2015.00254.
33. Yi Wang, Camilla C. Luck, Allison M. Waters, Luke J. Ney, and Ottmar V. Lipp. The effect of gradual extinction training on the renewal of electrodermal conditional responses. *Psychophysiology*, 61(12):e14681, December 2024. ISSN 0048-5772, 1469-8986. doi: 10.1111/psyp.14681.
34. Andrea L. Willcocks and Gavan P. McNally. An extinction retrieval cue attenuates renewal but not reacquisition of alcohol seeking. *Behavioral Neuroscience*, 128(1):83–91, 2014. ISSN 1939-0084, 0735-7044. doi: 10.1037/a0035595.
35. Gregory J Quirk and Devin Mueller. Neural Mechanisms of Extinction Learning and Retrieval. *Neuropsychopharmacology*, 33(1):56–72, January 2008. ISSN 0893-133X, 1740-634X. doi: 10.1038/sj.npp.1301555.
36. Amy L. Cochran and Josh M. Cisler. A flexible and generalizable model of online latent-state learning. *PLOS Computational Biology*, 15(9):e1007331, September 2019. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1007331.
37. Samuel J Gershman, Marie-H Monfils, Kenneth A Norman, and Yael Niv. The computational nature of memory modification. *eLife*, 6:e23763, March 2017. ISSN 2050-084X. doi: 10.7554/eLife.23763.
38. Samuel J. Gershman and Yael Niv. Exploring a latent cause theory of classical conditioning. *Learning & Behavior*, 40(3):255–268, September 2012. ISSN 1543-4494, 1543-4508. doi: 10.3758/s13420-012-0080-8.
39. A. David Redish, Steve Jensen, Adam Johnson, and Zeb Kurth-Nelson. Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychological Review*, 114(3):784–805, July 2007. ISSN 1939-1471, 0033-295X. doi: 10.1037/0033-295X.114.3.784.
40. Warren Woodrich Pettine, Dhruva Venkita Raman, A. David Redish, and John D. Murray. Human generalization of internal representations through prototype learning with goal-directed attention. *Nature Human Behaviour*, 7(3):442–463, March 2023. ISSN 2397-3374. doi: 10.1038/s41562-023-01543-7.
41. Huijeong Jeong, Annie Taylor, Joseph R Floeder, Martin Lohmann, Stefan Mihalas, Brenda Wu, Mingkang Zhou, Dennis A Burke, and Vijay Mohan K Namboodiri. Mesolimbic dopamine release conveys causal associations. *Science*, 378(6626):eabq6740, December 2022. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abq6740.
42. Vijay Mohan K Namboodiri and Garret D. Stuber. The learning of prospective and retrospective cognitive maps within neural circuits. *Neuron*, 109(22):3552–3575, November 2021. ISSN 08966273. doi: 10.1016/j.neuron.2021.09.034.
43. Vijay Mohan K Namboodiri. "But why?" Dopamine and causal learning. *Current Opinion in Behavioral Sciences*, 60:101443, December 2024. ISSN 23521546. doi: 10.1016/j.cobeha.2024.101443.
44. Luke T. Coddington, Sarah E. Lindo, and Joshua T. Dudman. Mesolimbic dopamine adapts the rate of learning from action. *Nature*, 614(7947):294–302, February 2023. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-022-05614-z.
45. Munir Gunes Kutlu, Jennifer E. Zachry, Patrick R. Melugin, Stephanie A. Cajigas, Maxime F. Chevee, Shannon J. Kelly, Banu Kutlu, Lin Tian, Cody A. Siciliano, and Erin S. Calipari. Dopamine release in the nucleus accumbens core signals perceived saliency. *Current Biology*, 31(21):4748–4761.e8, November 2021. ISSN 09609822. doi: 10.1016/j.cub.2021.08.052.
46. Jay A. Hennig, Sandra A. Romero Pinto, Takahiro Yamaguchi, Scott W. Linderman, Naoshige Uchida, and Samuel J. Gershman. Emergence of belief-like representations through reinforcement learning. *PLOS Computational Biology*, 19(9):e1011067, September 2023. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1011067.
47. L L Lindblom and H M Jenkins. Responses Eliminated by Noncontingent or Negatively Contingent Reinforcement Recover in Extinction. *Journal of Experimental Psychology: Animal Behavior Processes*, 7(2):175–190, 1981.
48. Robert A Rescorla. Redundant Treatments of Neutral and Excitatory Stimuli in Autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, 15(3):212–223, 1989.
49. Lechen Qian, Mark Burrell, Jay A. Hennig, Sara Matias, Venkatesh N. Murthy, Samuel J. Gershman, and Naoshige Uchida. Prospective contingency explains behavior and dopamine signals during associative learning. *Nature Neuroscience*, 28(6):1280–1292, June 2025. ISSN 1097-6256, 1546-1726. doi: 10.1038/s41593-025-01915-4.
50. Geoffrey Hall. Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. *Animal Learning & Behavior*, 24(3): 233–255, September 1996. ISSN 0090-4996, 1532-5830. doi: 10.3758/BF03198973.
51. R. C. Honey and Geoffrey Hall. Acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes*, 15(4):338–346, 1989. ISSN 1939-2184, 0097-7403. doi: 10.1037/0097-7403.15.4.338.
52. Jasper Ward-Robinson and Geoffrey Hall. The Role of Mediated Conditioning in Acquired Equivalence. *The Quarterly Journal of Experimental Psychology B*, 52(4):335–350, November 1999. ISSN 14641321, 02724995. doi: 10.1080/027249999393031.
53. Sasha C.V. Burwell, Haidun Yan, Shaun S.X. Lim, Brenda C. Shields, and Michael R. Tadross. Natural phasic inhibition of dopamine neurons signals cognitive rigidity, May 2024.
54. Alicia Izquierdo, Robin K. Suda, and Elisabeth A. Murray. Bilateral Orbital Prefrontal Cortex Lesions in Rhesus Monkeys Disrupt Choices Guided by Both Reward Value and Reward Contingency. *The Journal of Neuroscience*, 24(34):7540–7548, August 2004. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.1921-04.2004.
55. Geoffrey Schoenbaum, Summer L. Nugent, Michael P. Saddoris, and Barrry Setlow. Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations:. *Neuroreport*, 13(6):885–890, May 2002. ISSN 0959-4965. doi: 10.1097/00001756-200205070-00030.
56. Barry J. Everitt, Daniel M. Hutcheson, Karen D. Ersche, Yann Pelloux, Jeffrey W. Dalley, and Trevor W. Robbins. The Orbital Prefrontal Cortex and Drug Addiction in Laboratory Animals and Humans. *Annals of the New York Academy of Sciences*, 1121(1):576–597, December 2007. ISSN 0077-8923, 1749-6632. doi: 10.1196/annals.1401.022.
57. Rita A. Fuchs, K. Allison Evans, Macon P. Parker, and Ronald E. See. Differential involvement of orbitofrontal cortex subregions in conditioned cue-induced and cocaine-primed reinstatement of cocaine seeking in rats. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 24(29):6600–6610, July 2004. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.1924-04.2004.
58. E. D. London. Orbitofrontal Cortex and Human Drug Abuse: Functional Imaging. *Cerebral Cortex*, 10(3):334–342, March 2000. ISSN 14602199. doi: 10.1093/cercor/10.3.334.
59. Vincent Pascoli, Agnès Hiver, Ruud Van Zessen, Michaël Loureiro, Ridouane Achargui, Masaya Harada, Jérôme Flakowski, and Christian Lüscher. Stochastic synaptic plasticity underlying compulsion in a model of addiction. *Nature*, 564(7736):366–371, December 2018. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-018-0789-4.
60. Geoffrey Schoenbaum and Yavin Shaham. The Role of Orbitofrontal Cortex in Drug Addiction: A Review of Preclinical Studies. *Biological Psychiatry*, 63(3):256–262, February 2008. ISSN 00063223. doi: 10.1016/j.biopsych.2007.06.003.
61. Alicia Izquierdo. Functional Heterogeneity within Rat Orbitofrontal Cortex in Reward Learning and Decision Making. *The Journal of Neuroscience*, 37(44):10529–10540, November 2017. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.1678-17.2017.
62. Paul J. Cunningham and A. David Redish. Opposing, multiplexed information in lateral and ventral orbitofrontal cortex guides sequential foraging decisions in rats. *Neuron*, 113(18): 3031–3047.e6, September 2025. ISSN 08966273. doi: 10.1016/j.neuron.2025.06.009.
63. Melissa Malvaez, Christine Shieh, Michael D. Murphy, Venuz Y. Greenfield, and Kate M. Wassum. Distinct cortical–amygdala projections drive reward value encoding and retrieval. *Nature Neuroscience*, 22(5):762–769, May 2019. ISSN 1097-6256, 1546-1726. doi: 10.1038/s41593-019-0374-7.
64. Vijay Mohan K. Namboodiri, James M. Otis, Kay Van Heeswijk, Elisa S. Voets, Rizk A. Alghorazi, Jose Rodriguez-Romaguera, Stefan Mihalas, and Garret D. Stuber. Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. *Nature Neuroscience*, 22(7):1110–1121, July 2019. ISSN 1097-6256, 1546-1726. doi: 10.1038/s41593-019-0408-1.
65. Vijay Mohan K Namboodiri, Taylor Hobbs, Ivan Trujillo-Pisanty, Rhiana C. Simon, Madelyn M. Gray, and Garret D. Stuber. Relative salience signaling within a thalamo-orbitofrontal circuit governs learning rate. *Current Biology*, 31(23):5176–5191.e5, December 2021. ISSN 09609822. doi: 10.1016/j.cub.2021.09.037.
66. Sébastien Ballesta, Weikang Shi, Katherine E. Conen, and Camillo Padoa-Schioppa. Values encoded in orbitofrontal cortex are causally related to economic choices. *Nature*, 588(7838): 450–453, December 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-020-2880-x.
67. M. P. Noonan, M. E. Walton, T. E. J. Behrens, J. Sallet, M. J. Buckley, and M. F. S. Rushworth. Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, 107(47):20547–20552, November 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1012246107.
68. Camillo Padoa-Schioppa and John A. Assad. Neurons in the orbitofrontal cortex encode economic value. *Nature*, 441(7090):223–226, May 2006. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature04676.
69. Thomas A. Stalnaker, Nisha K. Cooch, Michael A. McDannald, Tzu-Lan Liu, Heather Wied, and Geoffrey Schoenbaum. Orbitofrontal neurons infer the value and identity of

predicted outcomes. *Nature Communications*, 5(1):3926, June 2014. ISSN 2041-1723. doi: 10.1038/ncomms4926.

70. Jung Hoon Sul, Hoseok Kim, Namjung Huh, Daeyeol Lee, and Min Whan Jung. Distinct Roles of Rodent Orbitofrontal and Medial Prefrontal Cortex in Decision Making. *Neuron*, 66 (3):449–460, May 2010. ISSN 08966273. doi: 10.1016/j.neuron.2010.03.033.
71. Jonathan D Wallis. Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature Neuroscience*, 15(1):13–19, January 2012. ISSN 1097-6256, 1546-1726. doi: 10.1038/nn.2956.
72. Robert C. Wilson, Yuji K. Takahashi, Geoffrey Schoenbaum, and Yael Niv. Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron*, 81(2):267–279, January 2014. ISSN 08966273. doi: 10.1016/j.neuron.2013.11.005.
73. Thomas W. Elston and Joni D. Wallis. Context-dependent decision-making in the primate hippocampal–prefrontal circuit. *Nature Neuroscience*, 28(2):374–382, February 2025. ISSN 1097-6256, 1546-1726. doi: 10.1038/s41593-024-01839-5.
74. Luke Pemberton, Huijeong Jeong, and Vijay Mohan K Namboodiri. Pre-existing and emergent cortical neuronal assembly sequences during learning. *bioRxiv*, November 2025. doi: https://doi.org/10.1101/2025.11.12.688154.
75. Bruno Popik, Felippe Espinelli Amorim, Olavo B Amaral, and Lucas De Oliveira Alvares. Shifting from fear to safety through deconditioning-update. *eLife*, 9:e51207, January 2020. ISSN 2050-084X. doi: 10.7554/eLife.51207.
76. Kathryn D. Baker, Gavan P. McNally, and Rick Richardson. Memory retrieval before or after extinction reduces recovery of fear in adolescent rats. *Learning & Memory*, 20(9):467–473, September 2013. ISSN 1549-5485. doi: 10.1101/lm.031989.113.
77. E.Z. Millan, Josie Milligan-Saville, and Gavan P. McNally. Memory retrieval, extinction, and reinstatement of alcohol seeking. *Neurobiology of Learning and Memory*, 101:26–32, March 2013. ISSN 10747427. doi: 10.1016/j.nlm.2012.12.010.
78. Allison Auchter, Lawrence K. Cormack, Yael Niv, Francisco Gonzalez-Lima, and Marie H. Monfils. Reconsolidation-Extinction Interactions in Fear Memory Attenuation: The Role of Inter-Trial Interval Variability. *Frontiers in Behavioral Neuroscience*, 11, January 2017. ISSN 1662-5153. doi: 10.3389/fnbeh.2017.00002.
79. G. P. Urcelay, D. S. Wheeler, and R. R. Miller. Spacing extinction trials alleviates renewal and spontaneous recovery. *Learning & Behavior*, 37(1):60–73, February 2009. ISSN 1543-4494, 1543-4508. doi: 10.3758/LB.37.1.60.
80. Eric Garr, Yifeng Cheng, Huijeong Jeong, Sara Brooke, Laia Castell, Aneesh Bal, Robin Magnard, Vijay Mohan K. Namboodiri, and Patricia H. Janak. Mesostriatal dopamine is sensitive to changes in specific cue-reward contingencies. *Science Advances*, 10(22): eadn4203, May 2024. ISSN 2375-2548. doi: 10.1126/sciadv.adn4203.
81. Laura Amato, Marina Davoli, Carlo A.Perucci, Marica Ferri, Fabrizio Faggiano, and Richard P. Mattick. An overview of systematic reviews of the effectiveness of opiate maintenance therapies: available evidence to inform clinical practice and research. *Journal of Substance Abuse Treatment*, 28(4):321–329, June 2005. ISSN 07405472. doi: 10.1016/j.jsat.2005.02.007.
82. Peter Blanken, Vincent M. Hendriks, Jan M. Van Ree, and Wim Van Den Brink. Outcome of long-term heroin-assisted treatment offered to chronic, treatment-resistant heroin addicts in the Netherlands. *Addiction*, 105(2):300–308, February 2010. ISSN 0965-2140, 1360-0443. doi: 10.1111/j.1360-0443.2009.02754.x.
83. European Monitoring Centre for Drugs and Drug Addiction. *New heroin-assisted treatment — recent evidence and current practices of supervised injectable heroin treatment in Europe and beyond :EMCDDA Insights*. Publications Office, LU, 2012.
84. Franziska Güttinger, Patrick Gschwend, Bernd Schulte, Jürgen Rehm, and Ambros Uchtenhagen. Evaluating Long-Term Effects of Heroin-Assisted Treatment: The Results of a 6-Year Follow-Up. *European Addiction Research*, 9(2):73–79, 2003. ISSN 1022-6877, 1421-9891. doi: 10.1159/000068811.
85. Christine Timko, Nicole R. Schultz, Michael A. Cucciare, Lisa Vittorio, and Christina Garrison-Diehn. Retention in medication-assisted treatment for opiate dependence: A systematic review. *Journal of Addictive Diseases*, 35(1):22–35, January 2016. ISSN 1055-0887, 1545-0848. doi: 10.1080/10550887.2016.1100960.
86. Roger L. Albin and Daniel K. Leventhal. Striatal Dopamine: The Cement of the Brain? Jeong H, Taylor A, Floeder JR, et al. Mesolimbic dopamine release conveys causal associations. Science 2022;378:eabq6740. *Movement disorders : official journal of the Movement Disorder Society*, 38(5):742, May 2023. ISSN 0885-3185. doi: 10.1002/mds.29379.
87. R. A. Rescorla and A. R. Wagner. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, 2:64–99, 1972.
88. Joseph R Floeder, Huijeong Jeong, Ali Mohebi, and Vijay Mohan K Namboodiri. Mesolimbic dopamine ramps reflect environmental timescales. *eLife*, 13:RP98666, August 2025. ISSN 2050-084X. doi: 10.7554/eLife.98666.
89. Arif A. Hamid, Jeffrey R. Pettibone, Omar S. Mabrouk, Vaughn L. Hetrick, Robert Schmidt, Caitlin M. Vander Weele, Robert T. Kennedy, Brandon J. Aragona, and Joshua D. Berke. Mesolimbic dopamine signals the value of work. *Nature Neuroscience*, 19(1):117–126, January 2016. ISSN 1546-1726. doi: 10.1038/nn.4173. Number: 1.
90. Mark W. Howe, Patrick L. Tierney, Stefan G. Sandberg, Paul E. M. Phillips, and Ann M. Graybiel. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature*, 500(7464):575–579, August 2013. ISSN 1476-4687. doi: 10.1038/nature12475.
91. HyungGoo R. Kim, Athar N. Malik, John G. Mikhael, Pol Bech, Iku Tsutsui-Kimura, Fangmiao Sun, Yajun Zhang, Yulong Li, Mitsuko Watabe-Uchida, Samuel J. Gershman, and Naoshige Uchida. A Unified Framework for Dopamine Signals across Timescales. *Cell*, 183(6): 1600–1616.e25, December 2020. ISSN 0092-8674. doi: 10.1016/j.cell.2020.11.013.
92. Ali Mohebi, Jeffrey R. Pettibone, Arif A. Hamid, Jenny-Marie T. Wong, Leah T. Vinson, Tommaso Patriarchi, Lin Tian, Robert T. Kennedy, and Joshua D. Berke. Dissociable dopamine dynamics for learning and motivation. *Nature*, 570(7759):65–70, June 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1235-y.
93. Yael Niv. Dopamine ramps up. *Nature*, 500(7464):533–535, August 2013. ISSN 1476-4687. doi: 10.1038/500533a.
94. John G. Mikhael, HyungGoo R. Kim, Naoshige Uchida, and Samuel J. Gershman. The role of state uncertainty in the dynamics of dopamine. *Current Biology*, 32(5):1077–1087.e9, March 2022. ISSN 0960-9822. doi: 10.1016/j.cub.2022.01.025.
95. Clara Kwon Starkweather, Benedicte M Babayan, Naoshige Uchida, and Samuel J Gershman. Dopamine reward prediction errors reflect hidden-state inference across time. *Nature Neuroscience*, 20(4):581–589, April 2017. ISSN 1097-6256, 1546-1726. doi: 10.1038/nn.4520.
96. Nathaniel D. Daw, Aaron C. Courville, and David S. Touretzky. Representation and Timing in Theories of the Dopamine System. *Neural Computation*, 18(7):1637–1677, July 2006. ISSN 0899-7667. doi: 10.1162/neco.2006.18.7.1637.
97. Samuel J. Gershman and Naoshige Uchida. Believing in dopamine. *Nature Reviews Neuroscience*, 20(11):703–714, November 2019. ISSN 1471-003X, 1471-0048. doi: 10.1038/s41583-019-0220-7.
98. Rachel S. Lee, Yotam Sagiv, Ben Engelhard, Ilana B. Witten, and Nathaniel D. Daw. A feature-specific prediction error model explains dopaminergic heterogeneity. *Nature Neuroscience*, 27(8):1574–1586, August 2024. ISSN 1097-6256, 1546-1726. doi: 10.1038/s41593-024-01689-1.
99. Elliot A. Ludvig, Richard S. Sutton, and E. James Kehoe. Stimulus Representation and the Timing of Reward-Prediction Errors in Models of the Dopamine System. *Neural Computation*, 20(12):3034–3054, December 2008. ISSN 0899-7667. doi: 10.1162/neco.2008.11-07-654.
100. Rajesh Rao. Decision Making Under Uncertainty: A Neural Model Based on Partially Observable Markov Decision Processes. *Frontiers in Computational Neuroscience*, 4, 2010. ISSN 1662-5188.
101. Yuji K. Takahashi, Thomas A. Stalnaker, Lauren E. Mueller, Sevan K. Harootonian, Angela J. Langdon, and Geoffrey Schoenbaum. Dopaminergic prediction errors in the ventral tegmental area reflect a multithreaded predictive model. *Nature Neuroscience*, 26(5): 830–839, May 2023. ISSN 1546-1726. doi: 10.1038/s41593-023-01310-x.
102. Zhewei Zhang, Kauê M. Costa, Angela J. Langdon, and Geoffrey Schoenbaum. The devilish details affecting TDRL models in dopamine research. *Trends in Cognitive Sciences*, 29(5): 434–447, May 2025. ISSN 1879-307X. doi: 10.1016/j.tics.2025.02.001.
103. Dennis A Burke, Annie Taylor, Huijeong Jeong, SeulAh Lee, Brenda Wu, Joseph R Floeder, and Vijay Mohan K Namboodiri. Reward timescale controls the rate of behavioral and dopaminergic learning, September 2024.
104. Ruud Van Zessen, Jacques P. Flores-Dourojeanni, Timon Eekel, Siem Van Den Reijen, Bart Lodder, Azar Omrani, Marten P. Smidt, Geert M.J. Ramakers, Geoffrey Van Der Plasse, Garret D. Stuber, and Roger A.H. Adan. Cue and Reward Evoked Dopamine Activity Is Necessary for Maintaining Learned Pavlovian Associations. *The Journal of Neuroscience*, 41(23):5004–5014, June 2021. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.2744-20.2021.
105. Kwang Lee, Leslie D. Claar, Ayaka Hachisuka, Konstantin I. Bakhurin, Jacquelyn Nguyen, Jeremy M. Trott, Jay L. Gill, and Sotiris C. Masmanidis. Temporally restricted dopaminergic control of reward-conditioned movements. *Nature Neuroscience*, 23(2):209–216, February 2020. ISSN 1097-6256, 1546-1726. doi: 10.1038/s41593-019-0567-0.
106. MB First, RL Spitzer, M Gibbon, and JB Williams. Structural Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition. (SCID-I/NP). *New York: Biometrics Research, New York State Psychiatric Institute*, November 2002.
107. Nancy S. Hemmes, Bruce L. Brown, James J. Jakubow, and Soledad Cabeza De Vaca. Determinants of Response Recovery in Extinction Following Response Elimination. *Learning and Motivation*, 28(4):542–557, November 1997. ISSN 00239690. doi: 10.1006/lmot.1997.0983.
108. Sylvia M.L. Cox, Chawki Benkelfat, Alain Dagher, J. Scott Delaney, France Durand, Samuel A. McKenzie, Theodore Kolivakis, Kevin F. Casey, and Marco Leyton. Striatal Dopamine Responses to Intranasal Cocaine Self-Administration in Humans. *Biological Psychiatry*, 65(10):846–850, May 2009. ISSN 00063223. doi: 10.1016/j.biopsych.2009.01.021.
109. Sylvia M. L. Cox, Yvonne Yau, Kevin Larcher, France Durand, Theodore Kolivakis, J. Scott Delaney, Alain Dagher, Chawki Benkelfat, and Marco Leyton. Cocaine Cue-Induced Dopamine Release in Recreational Cocaine Users. *Scientific Reports*, 7(1):46665, April 2017. ISSN 2045-2322. doi: 10.1038/srep46665.
110. Sylvia M. L. Cox, Chawki Benkelfat, Alain Dagher, J. Scott Delaney, France Durand, Theodore Kolivakis, Kevin F. Casey, and Marco Leyton. Effects of lowered serotonin transmission on cocaine-induced striatal dopamine response: PET [^{11} C]raclopride study in humans. *British Journal of Psychiatry*, 199(5):391–397, November 2011. ISSN 0007-1250, 1472-1465. doi: 10.1192/bjp.bp.110.084178.
111. Marco Leyton, Kevin F. Casey, J. Scott Delaney, Theodore Kolivakis, and Chawki Benkelfat. Cocaine craving, euphoria, and self-administration: A preliminary study of the effect of catecholamine precursor depletion. *Behavioral Neuroscience*, 119(6):1619–1627, 2005. ISSN 1939-0084, 0735-7044. doi: 10.1037/0735-7044.119.6.1619.
112. Linda C Sobell and Mark B Sobell. *A Technique for Assessing Self-Reported Alcohol Consumption*. Humana Press„ Totowa, NJ, 1992. ISBN 978-1-4612-6723-2.
113. Che-Hang Yu, Yiyi Yu, Liam M. Adsit, Jeremy T. Chang, Jad Barchini, Andrew H. Moberly, Hadas Benisty, Jinkyung Kim, Brent K. Young, Kathleen Heng, Deano M. Farinella, Austin Leikvoll, Rishaab Pavan, Rachel Vistein, Brandon R. Nanfito, David G. C. Hildebrand, Santiago Otero-Coronel, Alipasha Vaziri, Jeffrey L. Goldberg, Anthony J. Ricci, David Fitzpatrick, Jessica A. Cardin, Michael J. Higley, Gordon B. Smith, Prakash Kara, Kristina J. Nielsen, Ikuko T. Smith, and Spencer LaVere Smith. The Cousa objective: a long-working distance air objective for multiphoton imaging in vivo. *Nature Methods*, 21(1):132–141, January 2024. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-023-02098-1.
114. Chandrashekhar D. Borkar and Jonathan P. Fadok. A Novel Pavlovian Fear Conditioning Paradigm to Study Freezing and Flight Behavior. *Journal of Visualized Experiments*, (167): 61536, January 2021. ISSN 1940-087X. doi: 10.3791/61536.
115. Marius Pachitariu, Carsen Stringer, Sylvia Schröder, Mario Dipoppa, L. Federico Rossi, Matteo Carandini, and Kenneth D. Harris. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv*, page 061507, June 2016. doi: 10.1101/061507.

Supplementary Notes

Supplementary Note 1. Here, we simplify the calculation of prospective and retrospective associations by denoting them as $P(\text{cue} \rightarrow \text{reward})$ and $P(\text{cue} \leftarrow \text{reward})$ respectively. This involves two main simplifications. One is that the relevant association is better measured by the prospective and retrospective contingencies: specifically, $P(\text{cue} \rightarrow \text{reward}) - P(\text{reward})$ and $P(\text{cue} \leftarrow \text{reward}) - P(\text{cue})$ respectively. These conditional minus marginal probabilities measure how much more likely a reward is to follow a cue than its baseline probability, and how much more likely a cue is to precede a reward than its baseline probability. The second simplification is in describing the associations in terms of probability, which assumes a "trial period", i.e., a time period over which either pairing is calculated. In reality, ANCCR does not make such a trial-based assumption. It is instead based on the computation of event-based successor and predecessor representation contingencies from a continuous timeline of events. Because these considerations are not necessary to understand and motivate the experimental designs presented here, they are omitted in the presentation of the theory.

Supplementary Note 2. The results from **fig. S3** are consistent with cross-cue generalization providing a protective effect against contingency degradation. In the cue-alcohol experiment, which followed a similar paradigm as cue-sucrose experiment but employed only one CS+ (unlike the two CS+s in the cue-sucrose experiment), retrospective extinction impaired reacquisition, but without slowing extinction, in contrast to the results from Group2 (**fig. S3E-H**). This pattern suggests that, in the cue-sucrose experiment, cross-cue generalization contributed to slowed extinction in G2. Because both CS-A and CS-B were rewarded during acquisition, continued delivery of rewards with CS-B during retrospective extinction in G2 maintained a generalized cue-reward association, which slowed extinction of CS-A. Cross-cue generalization is a well-established phenomenon in associative learning (50–52), but for simplicity, many formal models including ANCCR and TDRPE have not yet explicitly modeled it, leaving its precise functional effects empirically uncertain. In contrast, as CS-A extinction in G1 was driven primarily by a decrease in estimated reward rate, it is less affected by cue generalization. This interpretation is supported by the absence of extinction-speed difference in the cue-alcohol experiment, where only a single cue was rewarded. Thus, the observed difference in CS-A extinction speed between groups 1 and 2 likely reflects cross-cue generalization.

Supplementary Note 3. We did not simulate the value RNN model (46) or latent cause models (36, 37) other than Redish et al. (39) on the dopamine optogenetic inhibition experiment in **Fig. 2** because they did not capture the behavioral results from groups 1 and 2 in **Fig. 1** (**fig. S6,7**). The Redish et al. 2007 latent cause model, while capturing the behavioral difference between groups 1 and 2, does not capture the results of the optogenetic inhibition experiment for Group 2 in **Fig. 2**. This is because the artificially induced negative RPE at reward delivery on CS-B trials makes the model attempt to improve its hidden context inference. This is achieved by making the model more sensitive to the external inputs, based on the reasoning that if the current inferred context is wrong, the external inputs likely contain the clues to better context inference. In control animals with no optogenetic inhibition, the model infers that the original acquisition context is active during the *CS-A extinction* phase. This is because the model identifies the partial similarity between *CS-A extinction* phase and the *Acquisition* phase in G2 due to the presence of CS-B trials in both phases. In experimental animals, however, the optogenetic inhibition artificially produces negative RPEs even though the external world has not otherwise changed. Although this makes the model more strongly attuned to the sensory inputs, the inferred context still remains the acquisition context, because the sensory evidence itself is unchanged. Further, as the assumed negative RPE induced by the optogenetic inhibition gets stronger, the value of both cues decreases in the acquisition context. Due to these combined effects, subsequent reacquisition becomes slower as the magnitude of simulated negative RPE increases (**fig. S6M**). Collectively, the observed results that optogenetic inhibition of dopamine reward responses on CS-B trials prevents CS-A memory erasure is inconsistent with this model.

Nevertheless, it is possible that this discrepancy is due to an implementation detail in Redish et al.'s model. This is because the intent of the model is to induce state splitting on encountering consistent negative RPEs. However, because of the particular implementation increasing the attention to external cues, it produces exactly the opposite effect as intended when applied to the optogenetic inhibition experiments performed here. Other implementations reducing the width of the radial basis function for stimulus categorization would be more likely to induce state splitting and capture our results. However, Redish et al. caution that this model is highly unstable. We did not therefore simulate this implementation.

Supplementary Note 4. ANCCR was proposed as an alternative model of associative learning and dopamine signaling. A common criticism is that ANCCR is "complex" (49, 86) whereas reward prediction error (RPE) models are "simple," because RPE is often introduced didactically as received minus predicted reward (87). In practice, however, the didactic RPE (essentially Rescorla–Wagner) cannot account for crucial phenomena such as persistent memory after extinction (4, 7, 11) or mesolimbic dopamine ramps (88–93), leading to the development of latent-cause (36–39) and hidden-state RPE models (94, 95). These model families are substantially more complex, and neither family individually captures both persistent memory and ramps. An RPE-based account of both would need to combine their complexities and further exploit flexible choices of state space and reward function (96–102). In contrast, ANCCR naturally captures both persistent memory and dopamine ramps. Thus, relative to the full RPE model family, ANCCR is not unusually complex or overparameterized. Notably, Qian et al. misstate in their main text that ANCCR has 12 free parameters instead of the 6 correctly listed in their Methods (49). Their value-RNN framework

does offer a principled way to discover latent states and eliminates the degrees of freedom afforded by hand-picked assumptions. The experiments here were designed specifically to test hidden-state inference, and the results disprove the model as it has no additional degrees of freedom that could account for the current findings. Overall, ANCCR is simpler than existing RPE-based models while explaining the same key findings.

In addition, the core computations in ANCCR yield several counterintuitive, easily falsifiable predictions. These include: (i) the persistence of memory after standard extinction and its loss after retrospective extinction; (ii) the exact quantitative dependence of cue–reward learning rate on the inter-reward interval (103); and (iii) the dependence of mesolimbic dopamine ramps on inter-trial intervals (88). Each prediction is empirically supported. Together, these results demonstrate that ANCCR provides a parsimonious account of well-established phenomena in associative learning and mesolimbic dopamine signaling.

Supplementary Note 5. ANCCR, like latent cause models, assumes that learned associations are context conditioned. The key difference is that ANCCR conditions on physical context, which is defined by sensory cues present in the environment, while latent cause models condition on latent contexts even when physical context remains unchanged. The context conditioning in ANCCR results in some nontrivial predictions. First, we have already shown both in terms of theoretical predictions and empirical observations that the rate of learning of a retrospective $A \leftarrow B$ association is proportional to the inter-B interval (103). Thus, ANCCR predicts that animals store the inter-event intervals for different types of stimuli experienced in a context (e.g., inter-reward interval, inter-CS-A interval, etc.). On transitioning from extinction to reacquisition, the original memory of inter-reward interval during conditioning provides a prior for the rapid restoration of estimated reward rate. Such rapid initialization of reward rate produces rapid reacquisition (during standard extinction when the original retrospective association is intact).

Supplementary Figures

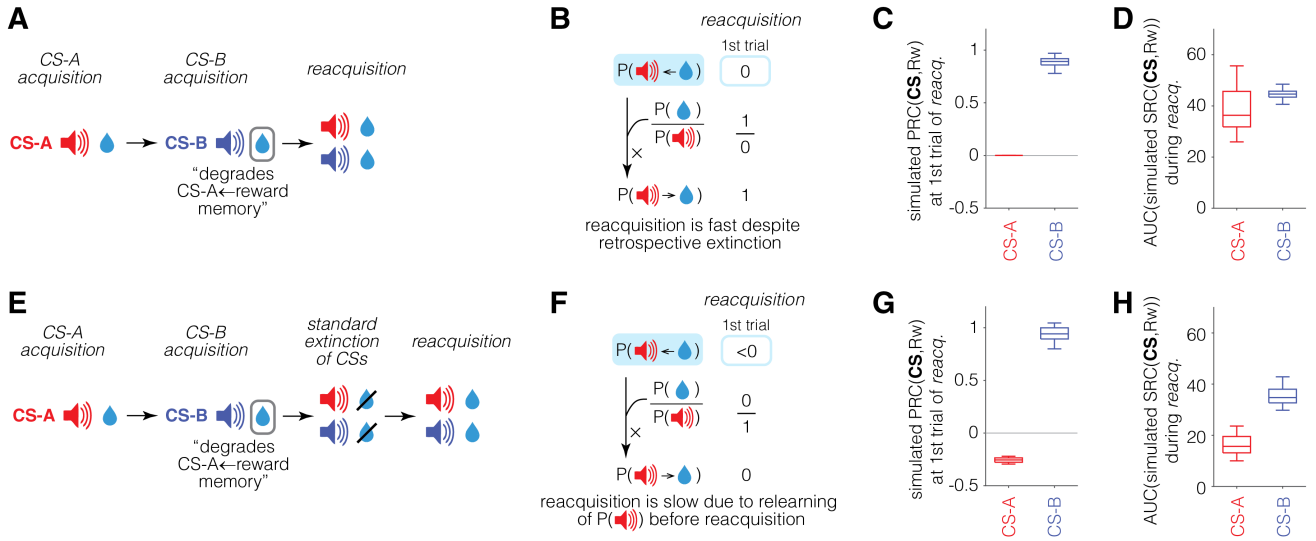

fig. S1. Reacquisition can remain fast despite retrospective extinction when the estimated cue rate is low at the time of reacquisition. Though the simplest retrospective extinction paradigm for CS-A is to present rewards while entirely omitting CS-A trials, such a paradigm would not alter the prospective $P(\text{CS-A} \rightarrow \text{reward})$ association. This is because updating this prospective conditional probability requires $P(\text{CS-A})$ to be non-zero, which can be achieved by presenting CS-A trials without reward following retrospective extinction. In other words, retrospective extinction should be followed by standard extinction to reveal the loss of memory in behavioral reacquisition. **A–D. A.** Separating CS-B acquisition from CS-A can degrade the CS-A←reward memory through retrospective extinction. **B.** However, when reacquisition follows CS-B acquisition immediately, the estimated rate of CS-A is near zero. Under ANCCR, this zero rate compensates for degraded CS-A←reward memory in the Bayes' rule like computation, as both numerator and denominator are near zero. Consequently, despite the erased retrospective memory, the model predicts that an animal would infer a high prospective memory (CS-A→reward memory) and acquire behavior rapidly. **C.** Simulation results confirm that the retrospective association (predecessor representation contingency, PRC) between CS-A and reward decreases to zero during CS-B acquisition, while that for CS-B remains high. **D.** Nonetheless, the prospective association (successor representation contingency, SRC) between CS-A and reward during reacquisition increases similarly fast as for CS-B, indicating fast reacquisition. **E–H. E, F.** Introducing standard extinction before reacquisition makes reacquisition speed depend on the degraded retrospective association. During standard extinction, ANCCR predicts that the estimated rate of CS-A recovers to its original level from CS-A acquisition. This recovery removes the zero-division compensation that previously enabled fast reacquisition despite degraded retrospective association, ensuring reacquisition speed reflects the strength of the remaining retrospective association. **G, H.** Simulation results show that the retrospective association between CS-A and reward is reduced during CS-B acquisition (G), resulting in slower recovery of the prospective association (H) compared with CS-B. Note that although this design could be the simplest approach to test retrospective extinction, we instead trained animals on CS-A and CS-B simultaneously and then introduced a separate retrospective extinction phase for CS-A (**Fig. 1E**). This was done to use CS-B as a within animal control and to avoid confounding effects of generalization, as sequential training (CS-A acquisition followed by CS-B acquisition) can accelerate CS-B acquisition.

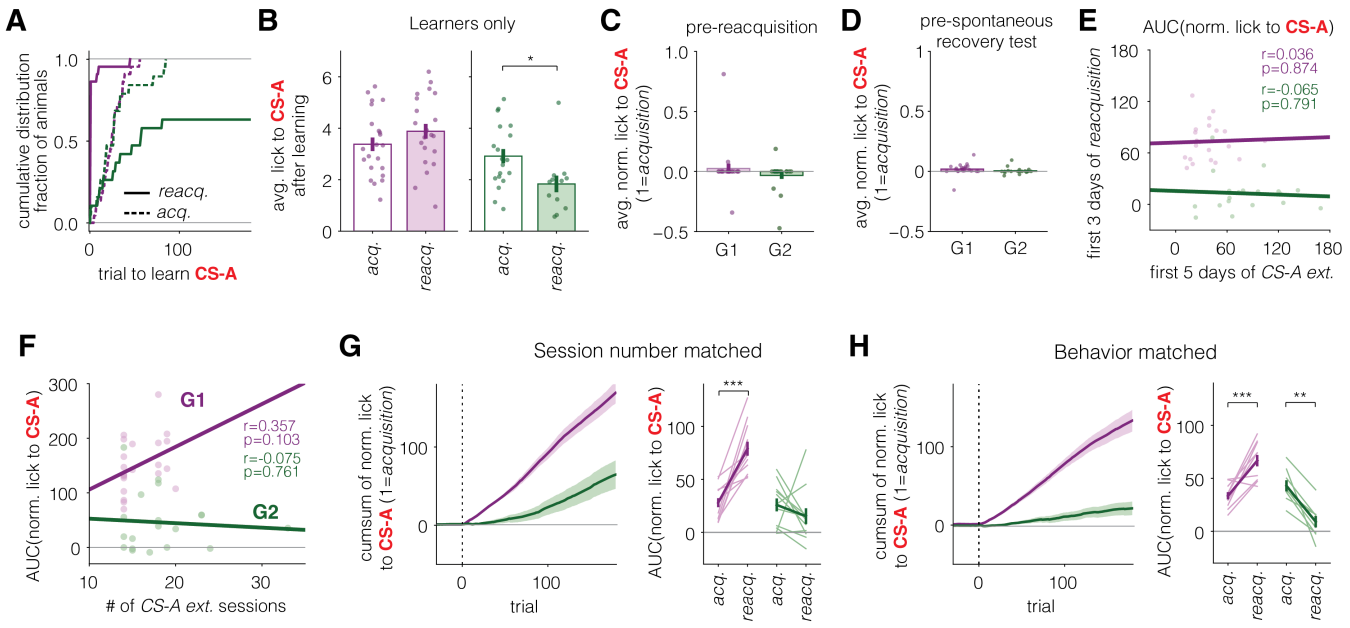

fig. S2. Differences in reacquisition or spontaneous recovery across groups are not due to training history or pre-test behavior. **A.** Cumulative distribution of animals based on trial to learn CS-A during acquisition and reacquisition. Mice were considered non-learners (trial to learn is infinity) if their post-learning average lick rate remained below 0.5 Hz and excluded from trial to learn analysis (7 of 19 G2 animals during reacquisition; see Methods). **B.** Lick to CS-A after acquisition (empty bar) or reacquisition (filled). Animals with post-learning lick rates below 0.5 Hz were excluded from this analysis (7 of 19 G2 animals during reacquisition; same exclusion as in **Fig. 1L**). **C.** No difference in pre-reacquisition licking behavior between groups. Individual animals are shown as circles. Error bar indicates SEM across animals. **D.** No difference in pre-test (average lick on a day before the test) licking behavior between groups in the spontaneous recovery analysis. Note that pre-test lick was averaged across trials within a given session, whereas spontaneous recovery measure in **Fig. 1N** uses the AUC of lick vs. trial plot over the first 5 trials where spontaneous recovery is primarily observed. **E.** No correlation between reacquisition speed and extinction speed to CS-A. **F.** No correlation between reacquisition speed and the number of CS-A extinction sessions. Each point represents an individual animal; the line indicates linear regression within each group **G.** Group-averaged cumulative sum of normalized lick to CS-A during acquisition (left) and AUC of normalized lick to CS-A over the first 6 days of acquisition and reacquisition (right), using animals matched for the number of CS-A extinction sessions across groups (n = 12 for G1, 11 for G2). Left, lines indicate mean across animals and shadings represent SEM. Right, light-colored lines indicate individual animals and error bar represents SEM. **H.** Same as G but using animals matched for extinction behavior (n = 10 for G1, 8 for G2). For behavior matching, CS-A extinction phase was extended 10 days after animal reached behavioral extinction threshold (see Methods).

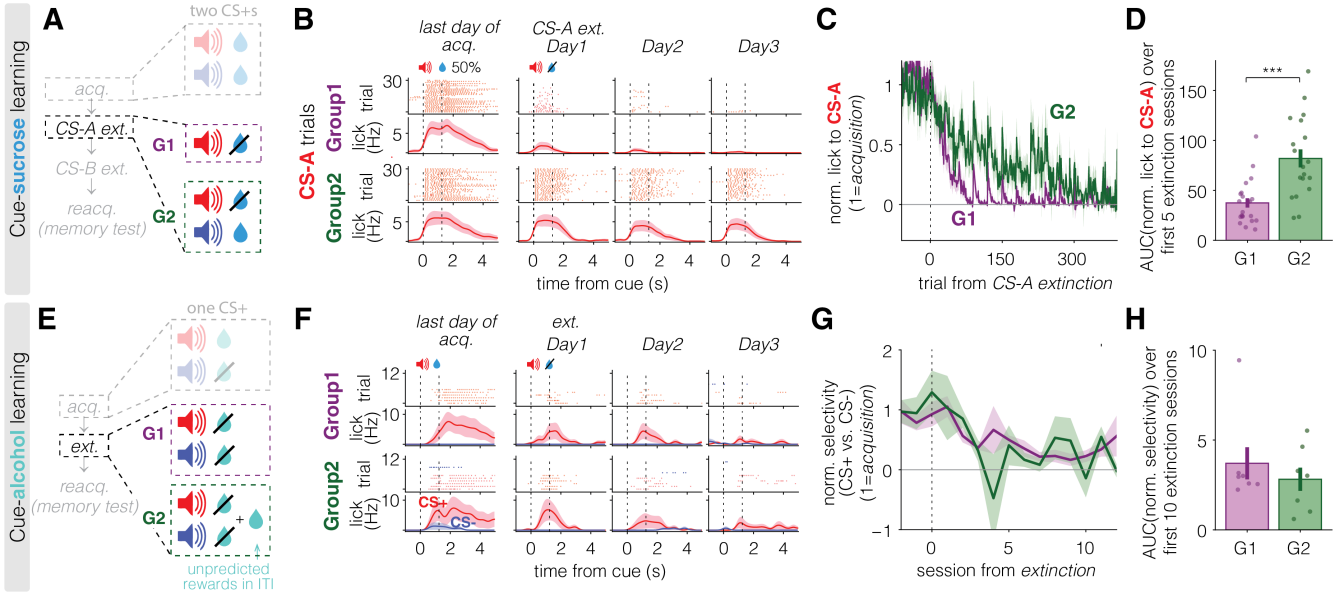

fig. S3. Slow CS-A extinction in Group2 may result from generalization across cues. A–D. Behavioral results from cue-sucrose learning. **A.** Task schematic for CS-A extinction in cue-sucrose learning (main task from **Fig. 1**). During CS-A extinction, G2 continued to experience CS-B reward trials, whereas G1 did not. Existence of CS-B rewarded trials in G2 may slow the extinction of CS-A due to generalization across cues acting as a protective mechanism against degradation of CS-A -reward association. **B.** Example lick raster plots (top row) and PSTHs (bottom row) during CS-A extinction for one representative mouse per group in cue-sucrose learning. The first column shows data from the last day of acquisition, and remaining columns (Day 1 to 3) show CS-A extinction. **C.** Group-averaged normalized lick to CS-A across trials during CS-A extinction. Trial 0 marks the first CS-A trial in CS-A extinction. Lines represent mean across animals and shadings represent SEM. **D.** AUC of normalized lick to CS-A over the first 5 days of CS-A extinction is significantly larger in G2 than G1. Individual animals are shown as circles. Error bar indicates SEM across animals. **E–H.** Behavioral results from cue-alcohol learning. **E.** Task schematic for extinction in cue-alcohol learning. Here, only one cue (CS+) was paired with reward during acquisition. In G2, unpredicted rewards were delivered in the inter-trial interval, rather than rewards paired with another CS+. This task paradigm avoids any slowdown in the extinction of CS+ due to generalization between cues. **F.** Same as B, but from mice in cue-alcohol learning. CS+ and CS- trials are shown in red and blue, respectively. **G.** Group-averaged normalized selectivity (see Methods) for CS+ over CS- across extinction sessions. Each animal's selectivity was normalized to its average during the last 5 days of acquisition. Session 0 indicates the first session of extinction. **H.** AUC of normalized selectivity for CS+ over CS- during the first 10 days of extinction.

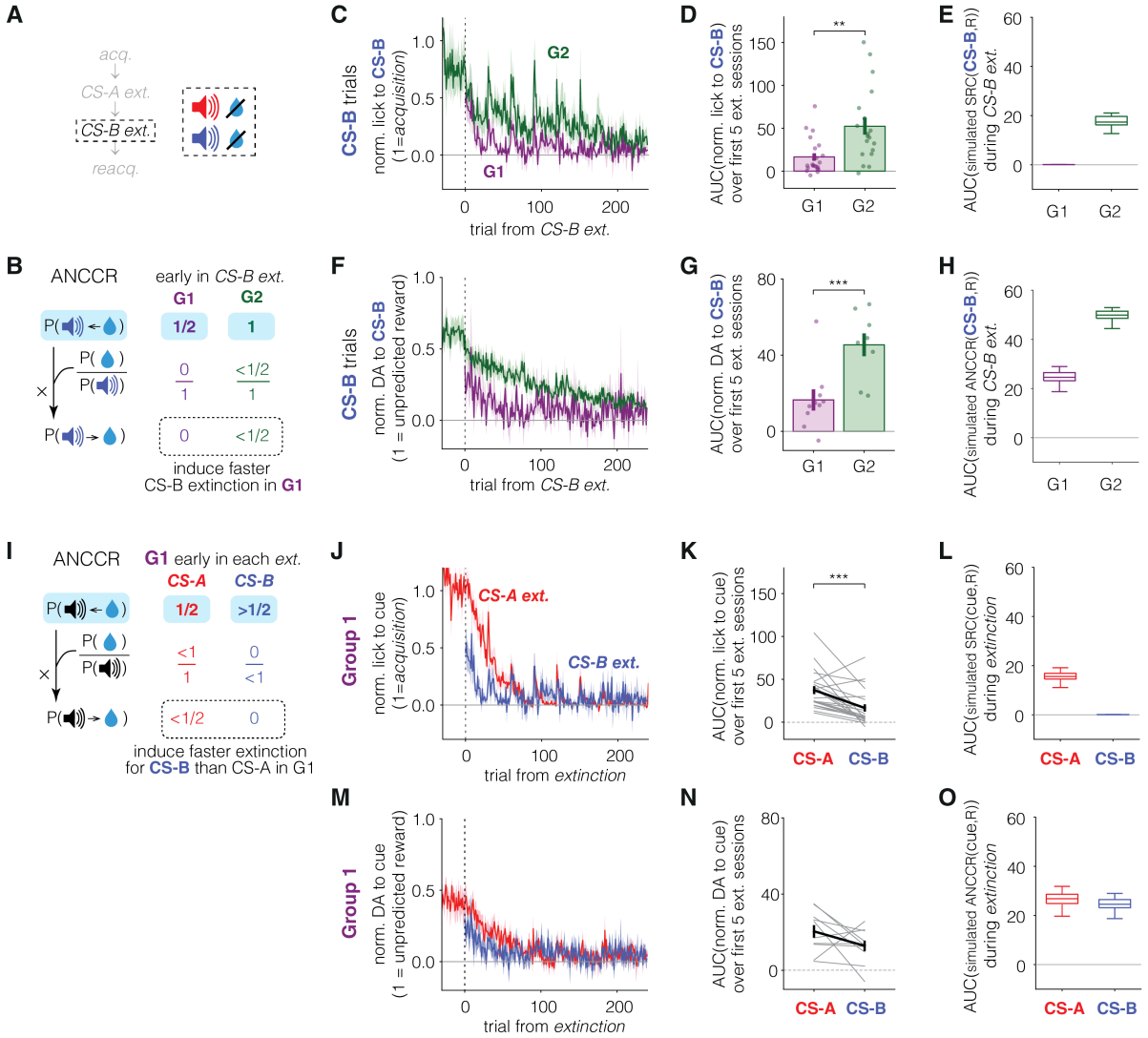

fig. S4. ANCCR predicts faster CS-B extinction in Group1 than in Group2. A–H. CS-B extinction is faster in G1 than in G2. **A.** Task schematic for CS-B extinction. **B.** Prediction from ANCCR. Because P(reward) had already decreased to zero during CS-A extinction in G1, CS-B extinction is predicted to occur faster in G1 than in G2. **C.** Group-averaged normalized lick to CS-B across trials during CS-B extinction. Trial 0 marks the first CS-B trial in CS-B extinction. Lines indicate mean and shadings represent SEM across animals. **D.** AUC of normalized lick to CS-B over the first 5 days of CS-B extinction. **E.** AUC of simulated prospective association (successor representation of contingency, SRC) between CS-B and reward, used to model behavior. **F.** Group-averaged normalized DA to CS-B across trials during CS-B extinction. **G.** AUC of normalized DA to CS-B over the first 5 days of CS-B extinction. **H.** AUC of simulated ANCCR between CS-B and reward during CS-B extinction. ANCCR reflects a weighted sum of prospective and retrospective association, hypothesized to represent dopamine response. Note that this weight may itself be dynamic depending on environmental statistics, but we have chosen not to include a fit of this dynamic weight in the ANCCR simulations. **I–O.** CS-B extinction is faster than CS-A extinction in G1. **I.** Prediction from ANCCR. In early CS-A extinction, P(rewards) decreases from its positive acquisition value to zero, driving extinction. In contrast, because P(reward) had already reached zero during CS-A extinction, subsequent CS-B extinction is predicted to occur more rapidly. **J.** Group-averaged normalized lick to CS-A and CS-B across extinction trials in G1. **K.** AUC of normalized lick to CS-A and CS-B over the first 5 days of each extinction phase in G1. Gray line indicates individual animals; error bar represents mean across animals. **L.** AUC of simulated SRC between cue (CS-A or CS-B) and reward during each extinction phase in G1. **M.** Group-averaged normalized DA to CS-A and CS-B across extinction trials in G1. **N.** AUC of normalized DA to CS-A and CS-B over the first 5 days of each extinction phase in G1. **O.** AUC of simulated ANCCR between cue (CS-A or CS-B) and reward during each extinction phase in G1. Note that there is no difference between CS-A and CS-B because higher retrospective association for CS-B (due to its lower baseline rate learned during preceding CS-A extinction) compensates for its weaker prospective association.

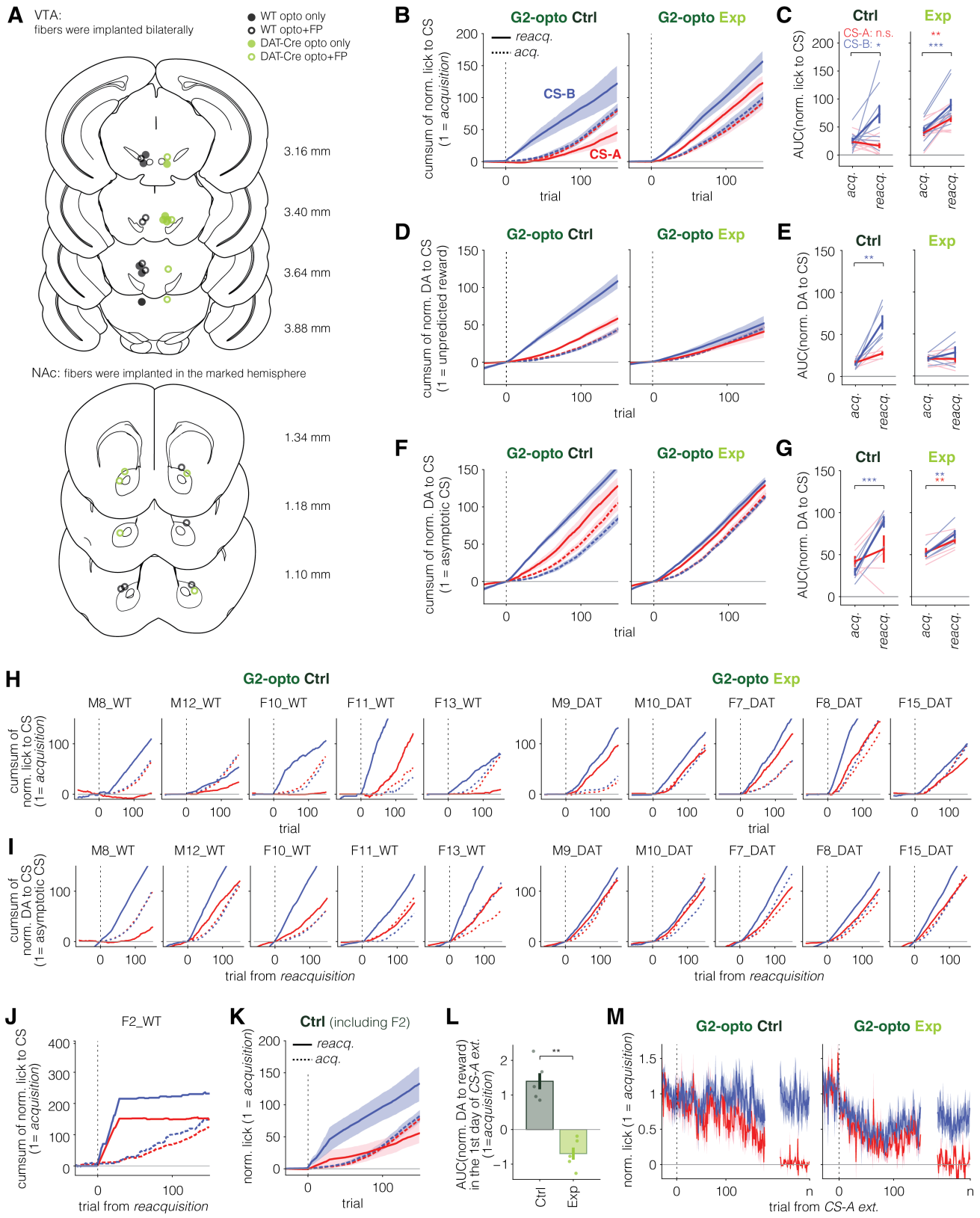

fig. S5. Reacquisition of dopamine response to cue in the VTA optogenetic experiment. A. Histological location of optic fibers for photometry recording of dopamine release in nucleus accumbens (top) and for optogenetic manipulation of VTA dopamine neurons (bottom). Dopamine release was simultaneously measured in a subset of animals for VTA optogenetic experiment (5 of 9 for G1, 5 of 9 for G2). The fiber was mislocated in one WT and one DAT-Cre animal (whose fibers are located at AP 3.88; fiber is located dorsal to VTA). but we included these animals: WT animal was in the control group with no opsin expression, and simultaneous fiber photometry recording in NAc in DAT-Cre animal confirmed strong inhibition of dopamine release, indicating sufficient light reached the VTA despite the dorsal placement of the fiber. The results did not change depending on the inclusion of these animals.

(fig. S5. Continued) **B–C.** Group-averaged normalized lick to CS (B) and AUC of normalized lick to CS (C) for all animals including those with and without dopamine measurements. These panels are identical to **Fig. 2D-G**, except that behavioral responses to both CS-A and CS-B are shown.**D–G.** Group-averaged normalized DA to CS (D, F) and AUC of normalized DA to CS (E, G) for animals in which dopamine release was measured. For D and E, DA to CS was normalized to maximum DA to reward across first three acquisition sessions. To dissociate the speed of learning (inflection point from zero) and the magnitude of response (asymptotic level of DA to CS), in F and G, DA to CS was normalized to asymptotic DA to CS in last 3 days of each phase. In both cases, reacquisition was rapid selectively to CS-B in WT animals. In DAT-Cre, the comparison between acquisition and reacquisition was dependent on normalization methods. Because dopamine inhibition during extinction can reduce estimated reward magnitude, overall cue responses (both CS-A and CS-B) was generally lower during reacquisition when compared to acquisition (see H and I). However, regardless of normalization method, no group difference in reacquisition speed was observed. **H–I.** Cumulative sum of normalized lick (H) and DA (I) to CS for individual animals using both normalization methods. **J–K. J.** One control animal (F2-WT) was excluded from analysis because its behavior on the first day of reacquisition was markedly different from all subsequent days, making its learning uninterpretable. **K.** Including this animal did not change the results (same analysis as in **Fig. 2E** but with F2-WT included). **L.** Normalized dopamine response to reward on the first day of CS-A extinction. Dopamine response to reward was significantly suppressed in Exp group, confirming effective optogenetic inhibition in Exp. Individual animals are shown as circles. Error bar indicates SEM across animals. **M.** Group-averaged normalized lick to CSs during CS-A extinction. Trial n indicates the last trial of CS-A extinction phase. Omission of rewards following CS-A induced a comparable reduction in anticipatory behavior to CS-A across groups. However, Exp group also showed a reduction in anticipatory behavior to CS-B, consistent with prior reports (104, 105). This generalized reduction in responding to reward-associated cues in Exp is consistent with the ANCCR prediction that below-baseline dopamine response at reward induces a reduction in the estimated reward magnitude.

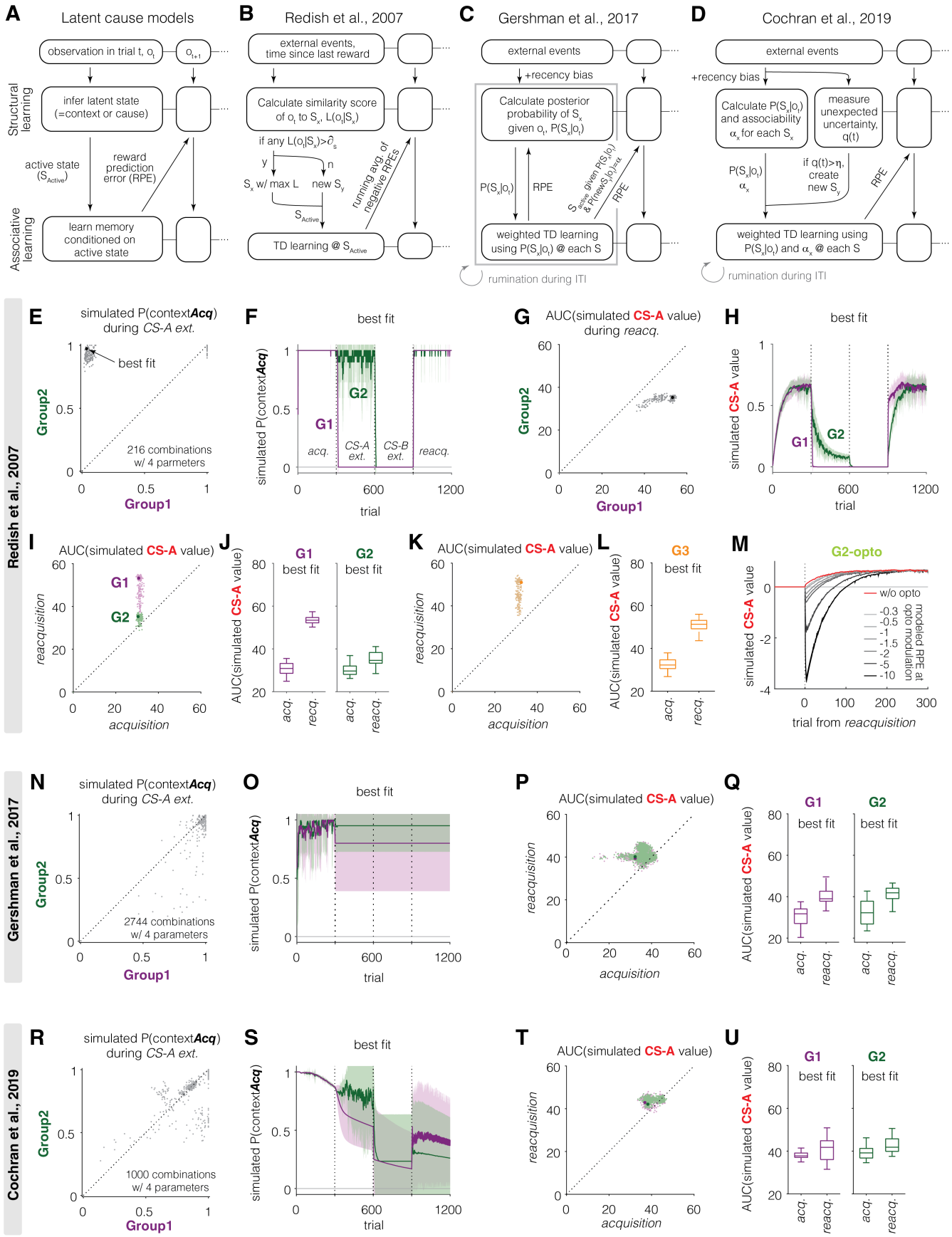

fig. S6. Variants of latent cause model do not capture behavioral results. A–D. Schematic of latent cause models. **A.** General framework. An agent infers a latent state (cause or context) from observations (o_t; structural learning), then learns cue-outcome memory conditioned on the active state (S_{active}; associative learning). Prediction error from associative learning influences latent state inference.

(fig. S6. Continued) **B.** Redish et al. (2007): Observations are compared to stored prototypes of existing latent states S_x. If similarity of any states exceeds a threshold ∂_s, the best-matching state is selected as active; otherwise, a new state S_y is created and selected as active. Sustained negative reward prediction errors (RPEs) across timesteps increase the likelihood of splitting into new states. **C.** Gershman et al. (2017): The agent infers a posterior distribution over latent states and updates memory for all states in proportion to this posterior. Structural learning (posterior update) and associative learning (memory update) alternate, with additional iterations ('rumination') occurring during inter-trial interval (ITI). Although associative updates are distributed across states, each trial is still assumed to be generated by a single latent cause, S_{active}. The state with the highest posterior at the end of each trial is selected and used as prior input for the next trial. Rumination due to replays does not change inferred state, S_{active}, but it updates the memory associated with that state. This assumption is critical for explaining extinction-retrieval results (**fig. S14**). **D.** Cochran et al. (2019): the agent infers posterior probability and estimates associability α_x for each state given the observation. If unexpected uncertainty $q(t)$ exceeds a threshold η, a new state is created. Memory is updated for all states, scaled by both the posterior and associability. These updates may iterate during ITI via rumination under the same posterior beliefs. **E–M.** Redish model captures results from G1 and G2, but not G3 or G2 with optogenetic manipulation. **E.** Simulated probability of inferring acquisition context, $P(\text{context}_{\text{Acq}})$, during CS-A extinction for each group across all parameter combinations (gray). Black dot indicates the best fit parameter combination (see Methods). **F.** Simulated $P(\text{context}_{\text{Acq}})$ over trials, across phases, using the best fit. Lines indicate mean across iterations and shadings represent STD. **G.** AUC of simulated CS-A value during reacquisition for each group across all parameter combinations. **H.** Simulated CS-A value over trials using the best fit. **I.** AUC of simulated CS-A value during acquisition vs. reacquisition for each group. Light dots indicate individual parameter combinations and darker dots mark the best fit combination. **J.** AUC of simulated CS-A value during acquisition and reacquisition for each group, using the best fit. In box plot, the central line indicates the median of iterations; box edges mark the 25th and 75th percentiles; whiskers extend to the most extreme data points within $1.5\times$ the interquartile range. **K–L.** Same as I–J, but for Group 3 (G3), which behaves like G1 and not G2. **M.** Simulated CS-A value over trials during reacquisition in G2-opto. Reacquisition was slower with optogenetic manipulation (greys) compared to without manipulation (red). Increasing the magnitude of modeled negative RPE during optogenetic manipulation (darker grey) progressively slowed reacquisition. **N–U.** Neither Gershman et al. (N–Q) nor Cochran et al. (R–U) model captures results from G1 and G2. Same panel structure as in E, F, I, J, but showing simulations from Gershman et al. model (N–Q) or Cochran et al. model (R–U).

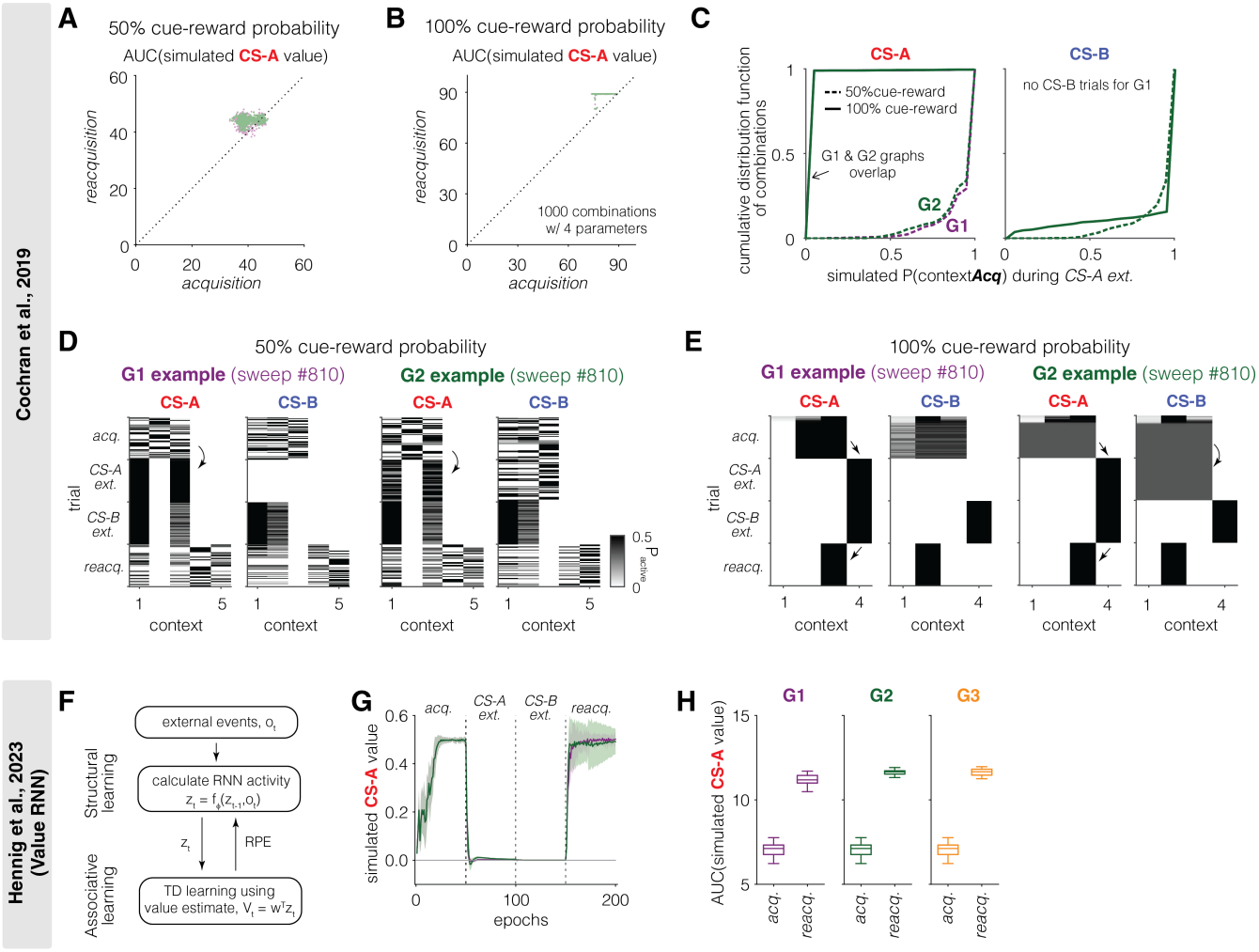

fig. S7. Additional simulation results from variants of latent cause model. A–E. Simulation results from Cochran et al. model. **A–B.** Simulated AUC of CS-A value during acquisition vs. reacquisition for all parameter combinations under 50% (A) or 100% (B) cue-reward probability. In both conditions, the model does not predict different speed of reacquisition between groups. Scatters indicate individual parameter combinations (Green, G1; Purple, G2). **C.** Cumulative distribution of simulated $P(\mathrm{context}_{\mathrm{Acq}})$ during CS-A extinction for CS-A (left) and CS-B (right) trials. Note that there are no CS-B trials for G1 during CS-A extinction. At 50% cue-reward probability (dotted), the model continues to infer $\mathrm{context}_{\mathrm{Acq}}$ in both groups. At 100% cue-reward probability (solid), the model infers a new context for CS-A trials (left) in both groups but maintains $\mathrm{context}_{\mathrm{Acq}}$ for CS-B trials in G2 (right). This shows that the presence of CS-B trials during CS-A extinction in G2 does not promote inference of $\mathrm{context}_{\mathrm{Acq}}$, even when cue-reward contingencies are deterministic. Instead, the model segregates CS-A and CS-B trials into distinct contexts. **D–E.** Trial-by-trial simulated context inference across phases for an example parameter combination (#810) under 50% (D) or 100% (E) cue-reward probability. Each row is a trial. Color scale shows inferred probability for each context. **F–H.** Simulation results from value RNN (Hennig et al., 2023). **F.** Schematic of value RNN. Activity of RNN units represents inferred state ('belief state') given trial. The weights for neural network (structural learning) and value calculation (associative learning) are trained to minimize RPE. **G.** Simulated CS-A value across trials. Value RNN predicts rapid reacquisition in both G1 and G2 and does not capture behavior results. **H.** AUC of simulated CS-A value during acquisition and reacquisition for each group. Value RNN fails to distinguish groups. In box plot, the central line indicates the median of iterations; box edges mark the 25th and 75th percentiles; whiskers extend to the most extreme data points within $1.5\times$ the interquartile range.

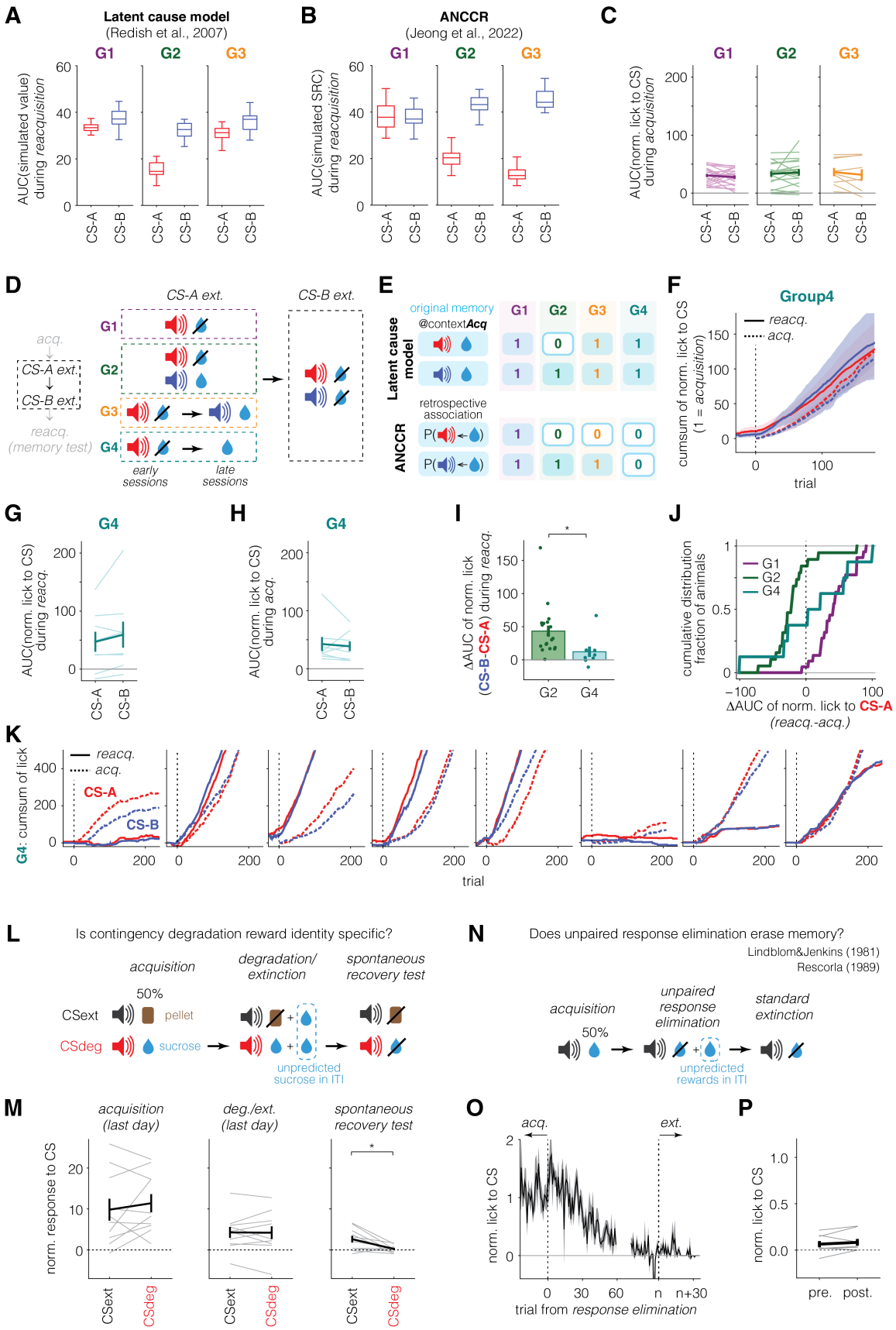

fig. S8. Presentation of uncued reward erases associated memories to the reward. A–B. Simulations. AUC of simulated CS-A and CS-B value (latent cause model, A) or SRC (ANCCR, B) during reacquisition, reflecting speed of reacquisition. **C.** AUC of normalized lick to CS-A and CS-B during acquisition. **D.** Task schematic for CS-A extinction. Group 4 (G4) underwent CS-A extinction like G3, but uncued rewards were delivered instead of CS-B reward trials in late extinction sessions.

(fig. S8. Continued) **E.** Model predictions. Latent cause model (top) predicts that original cue-reward memories are preserved in G4, as CS-A omission trials in early sessions drive context switch. In contrast, ANCCR (bottom) predicts unlearning of both CS-A and CS-B, as uncued rewards in late sessions degrade all retrospective associations paired with the reward. **F–K.** Behavioral results from G4 support ANCCR prediction. **F.** Group-averaged cumulative sum of normalized lick to CS-A (red) and CS-B (blue) during acquisition (dotted) and reacquisition (solid). On average, reacquisition was as slow as acquisition for both cues. Lines indicate mean across animals and shadings represent SEM. **G, H.** AUC of normalized lick to each cue during reacquisition (G) and acquisition (H). Individual animals are shown as light-colored lines. Error bar indicates SEM across animals. **I.** G4 shows significantly smaller difference in reacquisition speed between cues during reacquisition than G2. Individual animals are shown as circles. Error bar indicates SEM across animals. **J.** Cumulative distribution of animals based on reacquisition vs. acquisition difference in learning speed. G1 and G2 show opposing biases (faster or slower reacquisition), while G4 shows a wide range of distribution resulting in no bias on average. This variability may depend on how consistently animals associate rewards during unpaired reward-only sessions with the same context as the rest of experiment. Because a physical context is nothing but a set of sensory cues, some animals may infer the unpaired reward-only sessions lacking any neutral cues present elsewhere in the task as a distinct context, potentially preventing degradation of retrospective associations stored in the original context and resulting in little or no impairment in reacquisition. In contrast, G3 (textbfFig. 3) had auditory cues during these sessions and showed strong impairment that was consistent across animals. These results suggest that sensory cues support inference of a physical context, and that their absence can lead to misassignment of context (see **Supplementary Note 5** and the Discussion section in main text for more discussion about context dependency in ANCCR vs. latent cause models). **K.** Cumulative sum of lick to CS-A (red) and CS-B (blue) during acquisition and reacquisition for individual animals. **L–M.** Contingency degradation is reward-identity specific in freely moving rats. **L.** Task schematic. After the acquisition of two cue-reward pairs (CSdeg-sucrose and CSext-pellet), one cue (CSdeg) underwent contingency degradation while the other cue (CSext) underwent standard extinction. Memory was tested in a spontaneous recovery test. **M.** Normalized response to cues in each phase. More spontaneous recovery was observed to CSext than CSdeg (right), while no difference was found during acquisition (left) or degradation/extinction (middle). **N–P.** Unpredicted rewards eliminate memory in head-fixed mice. **N.** Task schematic. After cue-reward learning, response elimination was induced by omitting rewards after cue and delivering uncued rewards during ITI. Standard extinction followed to assess memory retention. **O.** Average normalized lick to CS over trials. Lick to CS declined during response elimination and did not recover in extinction as opposed to previous findings from pigeon auto shaping (47, 48). **P.** Normalized lick to CS during pre- and post-transition from response elimination to standard extinction.

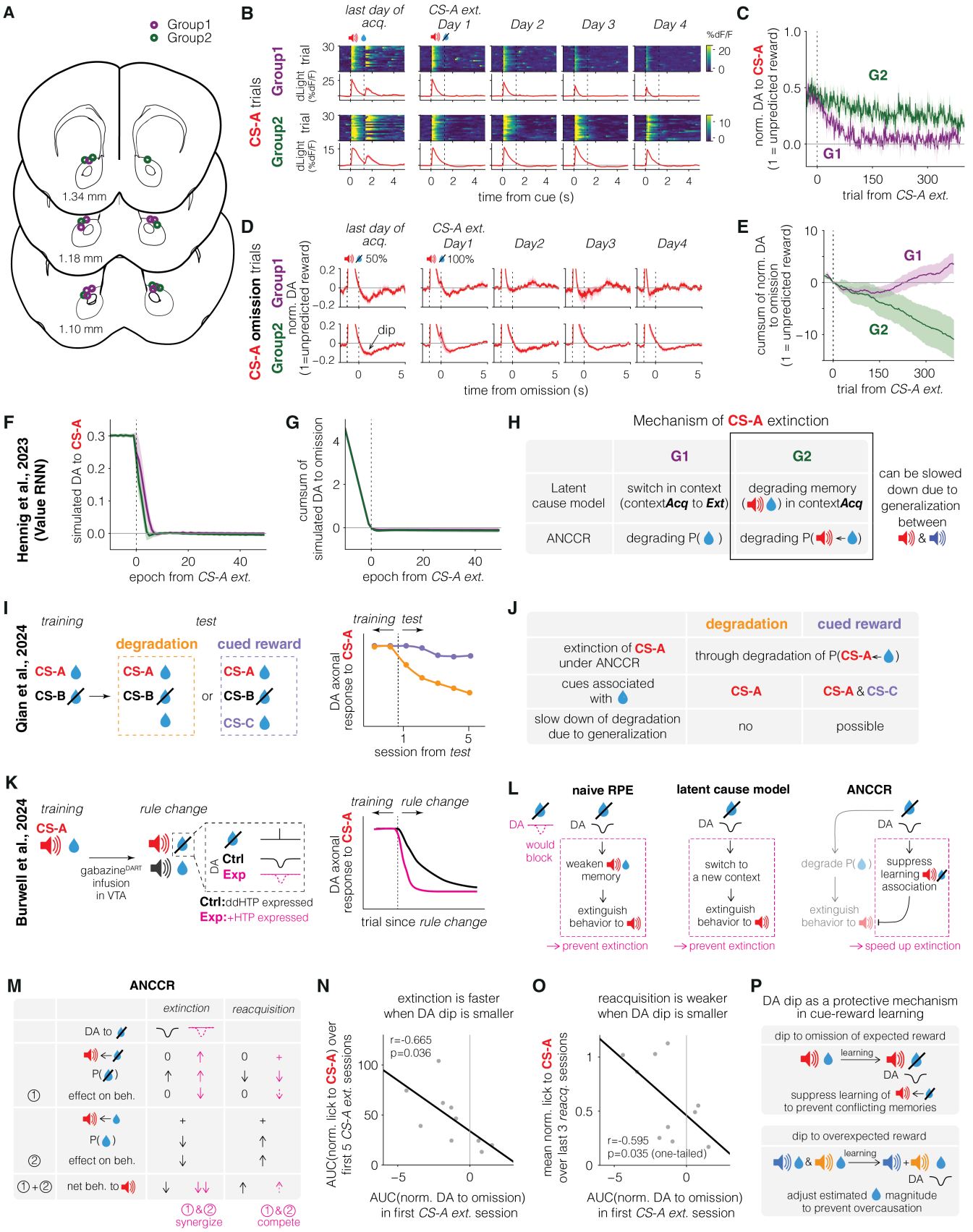

fig. S9. Potential mechanisms underlying dopamine dynamics during extinction. A–E. Changes in dopamine response associated with extinction are slower in G2 than G1 during CS-A extinction. **A.** Histological locations of optic fiber tips for animals from dopamine fiber photometry experiments. **B.** Example heatmaps (top) and PSTHs (bottom) of dopamine release (dF/F) in the nucleus accumbens in response to CS-A, from a representative mouse each in G1 and G2. The graphs are aligned to the cue onset.

(fig. S9. Continued) The first column shows data from the last day of acquisition, and the remaining columns show Days 1–4 of CS-A extinction. **C.** Group-averaged normalized DA to CS-A over trials during CS-A extinction. Lines indicate mean across animals and shadings represent SEM. **D.** Group-averaged PSTHs for CS-A omission trials over days. Time zero indicates the expected reward delivery time. During acquisition, omission occurred in 50% of CS-A trials; during extinction (Day 1 onward), all CS-A trials were unrewarded. G1 shows a pronounced DA dip that rapidly disappears, while G2 shows a prolonged dip across days. **E.** Group-averaged cumulative sum of normalized DA to omission across CS-A extinction trials. DA to omission was calculated as AUC of dF/F over a 1.25 s window after expected reward time. **F–G.** Value RNN, a model for unbiased and automatic discovery of state spaces, does not capture group differences in dopamine extinction. **F.** Simulated DA (RPE) to CS-A over epochs (see Methods) during CS-A extinction. Lines indicate mean across iterations and shading represent STD. The model predicts a similar dopamine extinction profile across groups. **G.** Cumulative sum of simulated DA to omission. Again, no group differences are observed. **H.** Summary of hypothesized mechanisms supporting group differences in CS-A extinction. Neither latent cause model nor ANCCR can account for slower extinction in G2 than G1 without additional assumptions. Assuming generalization between CS-A and CS-B can rescue both models. **I–J.** Findings from Qian et al. can be explained by ANCCR by assuming generalization. **I.** Task schematic and behavioral results. Animals that associated a single cue (CS-A) with reward during training received ITI reward either without any cue (degradation) or with a new cue (cued reward) during test. Dopamine response to CS-A decreased more slowly when ITI reward was cued. **J.** ANCCR with generalization explains the behavior results, as when a new cue is introduced in cued reward condition, generalization between the existing cue (CS-A) and the new cue provides a protective mechanism that slows degradation of CS-A responses. **K–L.** Findings from Burwell et al. could be explained by ANCCR. **K.** Task schematic and behavioral result. Animals were trained to associate CS-A with reward. During the rule change phase, CS-A was no longer rewarded while a new cue was paired with reward. Experimental group (Exp), in which dopamine dip at omission was blocked, exhibited faster extinction compared to control group (Ctrl). **L.** Summary of how dopamine dip at omission affects extinction in different models. In RPE-based theories (naïve RPE and latent cause model), extinction is triggered by a dip in dopamine at omission. In contrast, ANCCR posits that extinction arises from degradation of P(reward), independent of dopamine at omission. Instead, dopamine dip may reduce the estimated "meaningfulness" of omission event, weakening learning of a cue←omission (i.e., frustration) association. Thus, blocking the dopamine dip may paradoxically enhance learning of cue-omission association and speed up extinction. **M–O.** Preliminary results support the proposed role of dopamine dip in cue-omission learning in ANCCR. **M.** Schematic illustrating how dopamine dip at omission operates in ANCCR. A negative response at omission prevents learning of cue←omission (i.e., cue←frustration) association by reducing "meaningfulness" of omission event. This mechanism ensures that animals do not form a separate cue-omission memory on top of the existing cue-reward memory, as recognition of omission itself depends on prior learning of the cue-reward contingency and cannot be treated as an independent event. ANCCR predicts that abolishing the dopamine dip allows learning of a cue-omission association, which facilitates behavioral extinction by introducing additional cue-omission (i.e., frustration) memory alongside the decrease in estimated reward rate. However, this newly formed cue-omission memory competes with the preserved cue-reward memory during reacquisition, resulting in slower relearning. **N.** Correlation between magnitude of dopamine dip at omission on the first day of CS-A extinction and speed of CS-A extinction across animals. Across animals, a larger dopamine dip at omission on the first day of CS-A extinction correlated with slower extinction, consistent with ANCCR's prediction that cue-omission learning facilitates extinction. **O.** Correlation between magnitude of dopamine dip at omission on the first day of CS-A extinction and asymptotic level of reacquired behavior to CS-A. Animals with smaller dopamine dip at omission showed weaker asymptotic levels of reacquired behavior to CS-A, supporting the prediction that cue-omission memory competes with cue-reward memory during reacquisition. **P.** Dopamine dip acts as a protective mechanism in cue-reward learning. During omission of an expected reward, dopamine dip prevents maladaptive learning by suppressing cue←omission association that could compete with existing cue←reward memory. During overexpectation, dopamine dip adjusts the estimated reward magnitude associated with each cue to prevent overexpectation. Although the omission-evoked dip may also reduce the perceived magnitude of the omission (i.e., magnitude of frustration) itself, this has little impact because cue←omission learning is simultaneously suppressed. Together, these mechanisms stabilize learning by preventing conflicting or excessive associations.

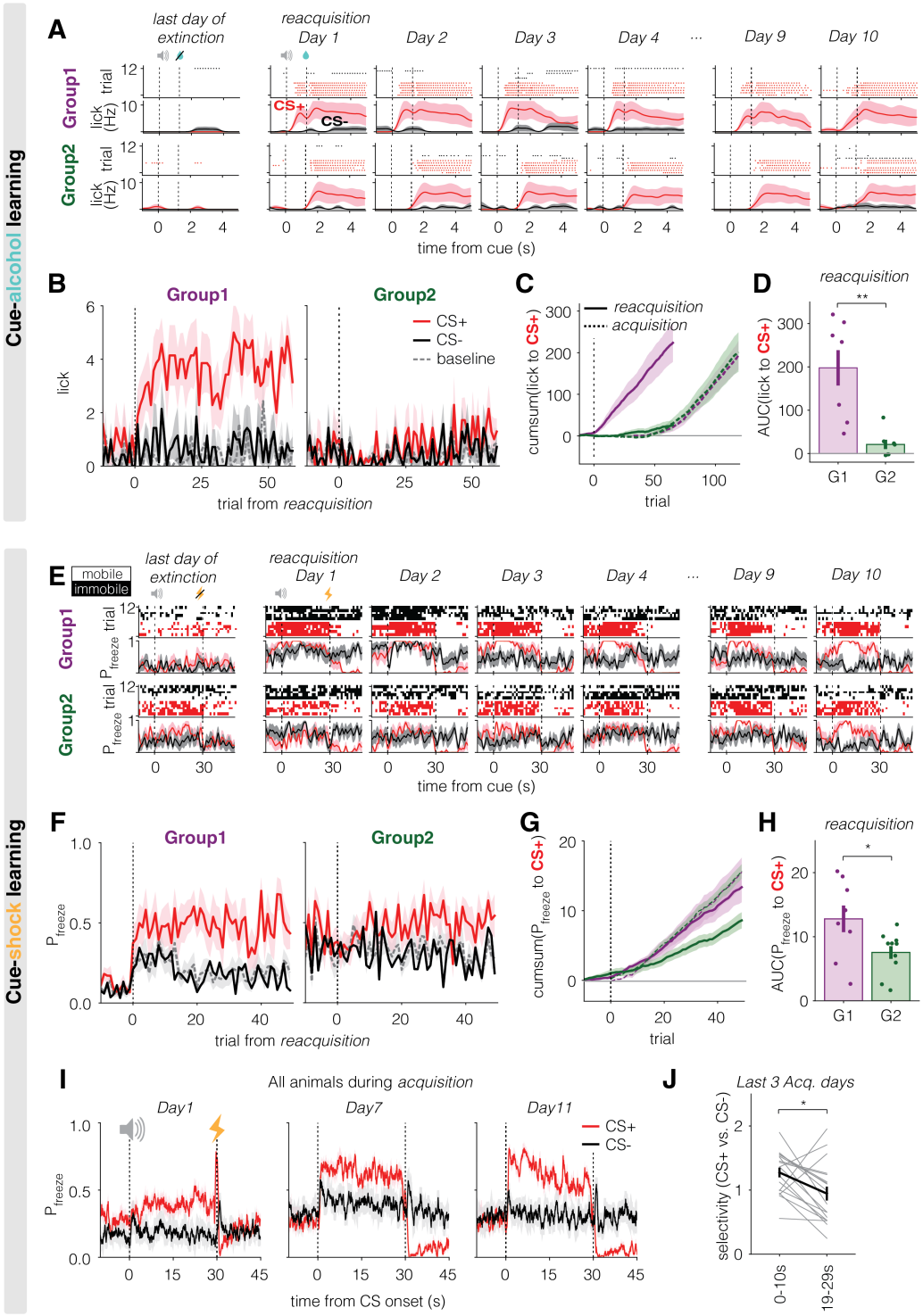

fig. S10. Trial-based analysis for cue-alcohol and cue-shock learning. A–D. Cue-alcohol learning. **A.** Example lick raster plots (top row) and PSTHs (bottom row) for a representative mouse for each group (top, G1; bottom, G2). The first column shows data from the last day (Day -1) of extinction, and the remaining columns show reacquisition. G1 reacquires anticipatory lick for CS+ (red) immediately on Day 1, while G2 did not show anticipatory lick until Day 10. Note that lick to CS− (black) remained for both groups throughout days. In PSTHs, lines indicate mean across trials and shadings represent SEM. **B.** Group-averaged lick count during a 1.25 s window after (red, CS+; black, CS−) or before (dashed gray, baseline) cue onset across trials. Trial zero is the first trial of reacquisition. G1, but not G2, rapidly reacquires anticipatory lick to CS+. Lines indicate mean across animals and shadings represent SEM. **C.** Cumulative sum of lick to CS+ across trials during acquisition (dotted) and reacquisition (solid). Trials < 0 in reacquisition indicate extinction trials. G1 shows rapid reacquisition, while reacquisition is as slow as acquisition in G2. **D.** AUC of lick to CS+ during reacquisition. Individual animals are shown as circles. Error bars indicate SEM across animals. A one-tailed t test was used for this comparison, based on a pre-specified directional hypothesis motivated by the results in **Fig. 4C** and the corresponding effects observed in cue-sucrose learning (**Fig. 1**).

(fig. S10. Continued) **E–J.** Cue-shock learning. Same format as A–D, using immobile probability (P_{freeze}) during a 10 s window as behavior measure. **E.** Example heatmap (top row) and PSTHs (bottom row) of immobility for representative mice. In heatmap, red and black indicate immobile states, while white indicates mobile states. In G1, selective immobility to CS+ (red) emerged by Day 3, with reduced immobility during CS− (black) and baseline. In contrast, G2 showed sustained immobility to both cue and baseline, with delayed selectivity emerging only by Day 10. **F.** G1 rapidly develops immobility to CS+ during reacquisition, while G2 shows nonselective immobility to both cues and baseline during extinction and a delayed emergence of CS+ selectivity during reacquisition. **G.** Cumulative sum of immobile probability to CS+ across trials during acquisition (dotted) and reacquisition (solid). Reacquisition of immobility to CS+ alone was not faster than acquisition, even in G1. However, selectivity to CS+ over CS− developed more rapidly during reacquisition in G1 (see **Fig. 3F**), suggesting that learning involves both response to CS+ and suppression of response to CS−. **H.** AUC of immobile probability to CS+ during reacquisition. A one-tailed t test was used based on the same pre-specified directional hypothesis as in D. **I.** Average PSTHs of immobility in CS+ (red) and CS− (black) trials across acquisition. As learning progresses, the difference between CS+ and CS− immobility becomes more pronounced during the early portion of the cue period. **J.** Selectivity index quantifying the difference in immobility between CS+ and CS− during early (0–10 s) vs. late (19–29 s) cue periods on the last 3 days of Acquisition (Day 9–11). Gray lines indicate individual animals. Error bar shows SEM across animals.

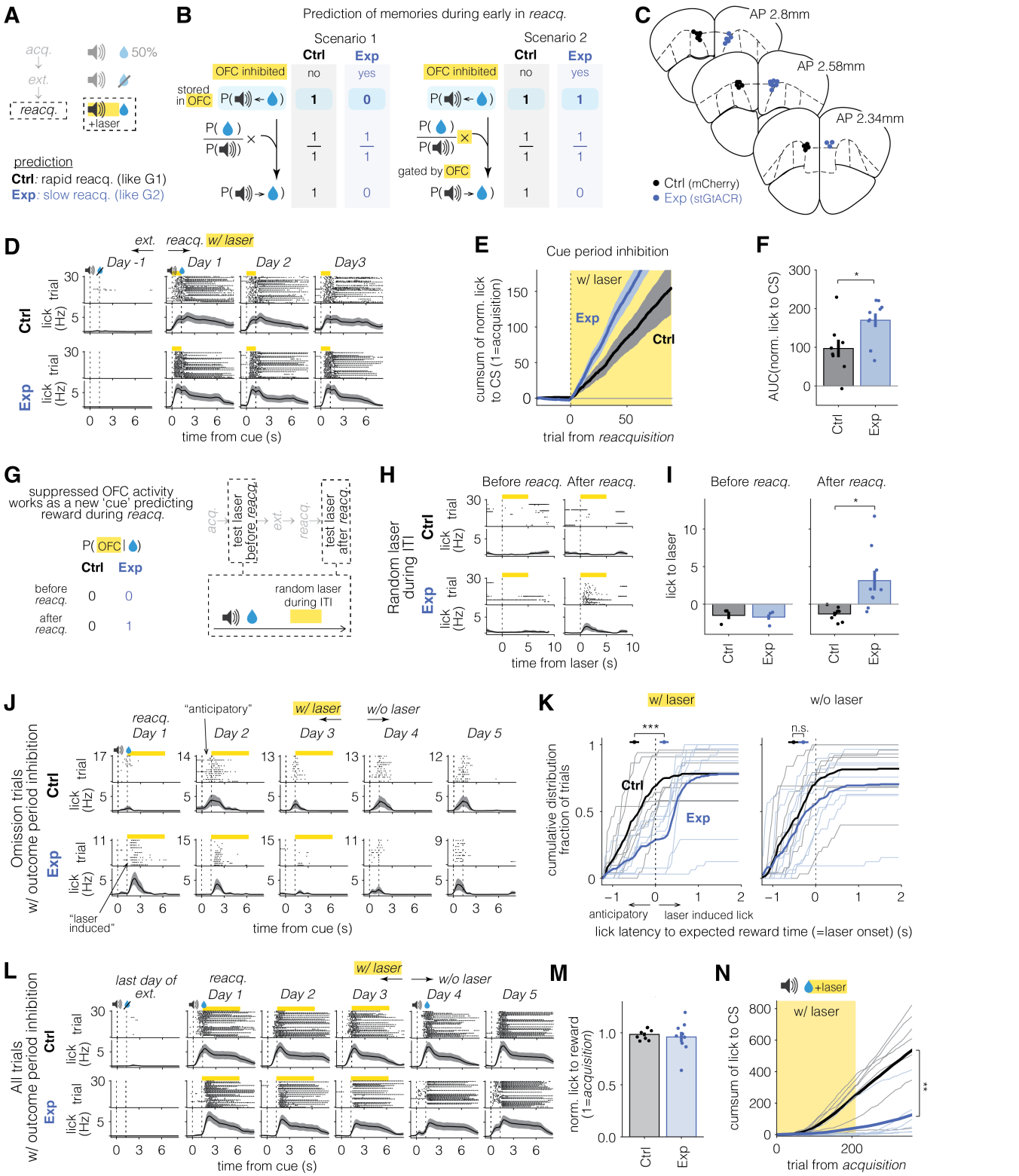

fig. S11. Suppressed OFC activity during cue works as a new cue predicting reward. A. Experimental schematic. During reacquisition following standard extinction, OFC CamKIIα-expressing neurons were optogenetically suppressed for 1.25 s from cue onset to reward in experimental group (Exp). The control group (Ctrl) underwent the same procedure without expression of inhibitory opsin. **B.** Two hypotheses that can slow down reacquisition in Exp. If retrospective association between cue and reward is stored in OFC, suppression of the memory can degrade reacquisition (scenario 1). Or if computation of prospective association is gated by OFC, preventing it would also result in a deficit in reacquisition (scenario 2). **C.** Histological location of optic fiber centers for OFC optogenetic experiments. **D.** Example lick raster plots and PSTHs for a representative mouse per group (top, Ctrl; bottom, Exp). Data are shown for the last day of extinction and the first three days of reacquisition (Day 1 to 3). Yellow bars indicate laser-on periods. Both Ctrl and Exp animals reacquire anticipatory lick on Day 1.

(fig. S11. Continued) **E.** Group-averaged cumulative sum of normalized lick to CS. Yellow shading indicates trials with laser on. Both groups show rapid reacquisition compared to acquisition, with more licking per trial in Exp. **F.** AUC of normalized lick to CS during reacquisition was larger in Exp than Ctrl. **G.** Experimental schematic to test if suppression of OFC works as a new cue. A 5 s long laser was randomly delivered during ITI before and after reacquisition to see if laser induces licking. During reacquisition, laser was delivered during cue-reward delay, allowing it to be associated with following reward. **H.** Example lick raster plots and PSTHs for example animal per group in random laser test before (left) and after (right) reacquisition. In Exp, delivery of laser induces licking only after reacquisition, where laser was followed by reward, indicating that the suppressed activity of OFC induced by the laser works as a new cue predicting reward. **I.** Comparison of lick to laser between groups before (left) and after (right) reacquisition. **J.** Example lick raster plots and PSTHs for one representative animal per group, showing omission trials during reacquisition with outcome period inhibition. Yellow bars indicate laser on periods. In Ctrl, anticipatory lick appears by Day 2. However, Exp animal shows lick immediately following laser onset on Day 1. This laser-induced lick diminishes from Day 1 to Day 3, consistent with extinction of laser-reward association. When laser was omitted on Day 4, Exp animal begins to show anticipatory lick. **K.** Cumulative distribution of first lick latency during omission trials, aligned to the expected reward time (and laser onset). Light traces represent individual animals. Bold traces show group averages. Ctrl shows anticipatory lick regardless of laser condition (left, with laser; right, without laser). Exp shows strong laser-induced lick in sessions with laser (left), which disappears when the laser is omitted (right). Horizontal error bar indicates group average of mean lick latency. **L.** Example lick raster plots (top row) and PSTHs (bottom row) for one mouse per group (top, Ctrl; bottom, Exp). Data are shown across days including the last day of extinction and the first five days of reacquisition (Day 1 to 5). Laser was delivered on the first three days of reacquisition (Day 1 to 3). Yellow bars indicate laser-on periods. Ctrl shows rapid reacquisition on Day 1, while Exp fails to reacquire for all three days. When laser was turned off on Day 4, anticipatory lick in Exp emerged, indicating the delay in reacquisition was due to suppression of OFC reward responses. **M.** Normalized lick to reward does not differ between groups, suggesting that OFC suppression does not impair motor ability or motivation to reward. **N.** Group-averaged cumulative sum of normalized lick to CS during acquisition when OFC reward response was inhibited (Exp in blue) or not (Ctrl in black). *, AUC of anticipatory licking across trials during acquisition with laser on was significantly lower when OFC reward response was inhibited than when it was not.

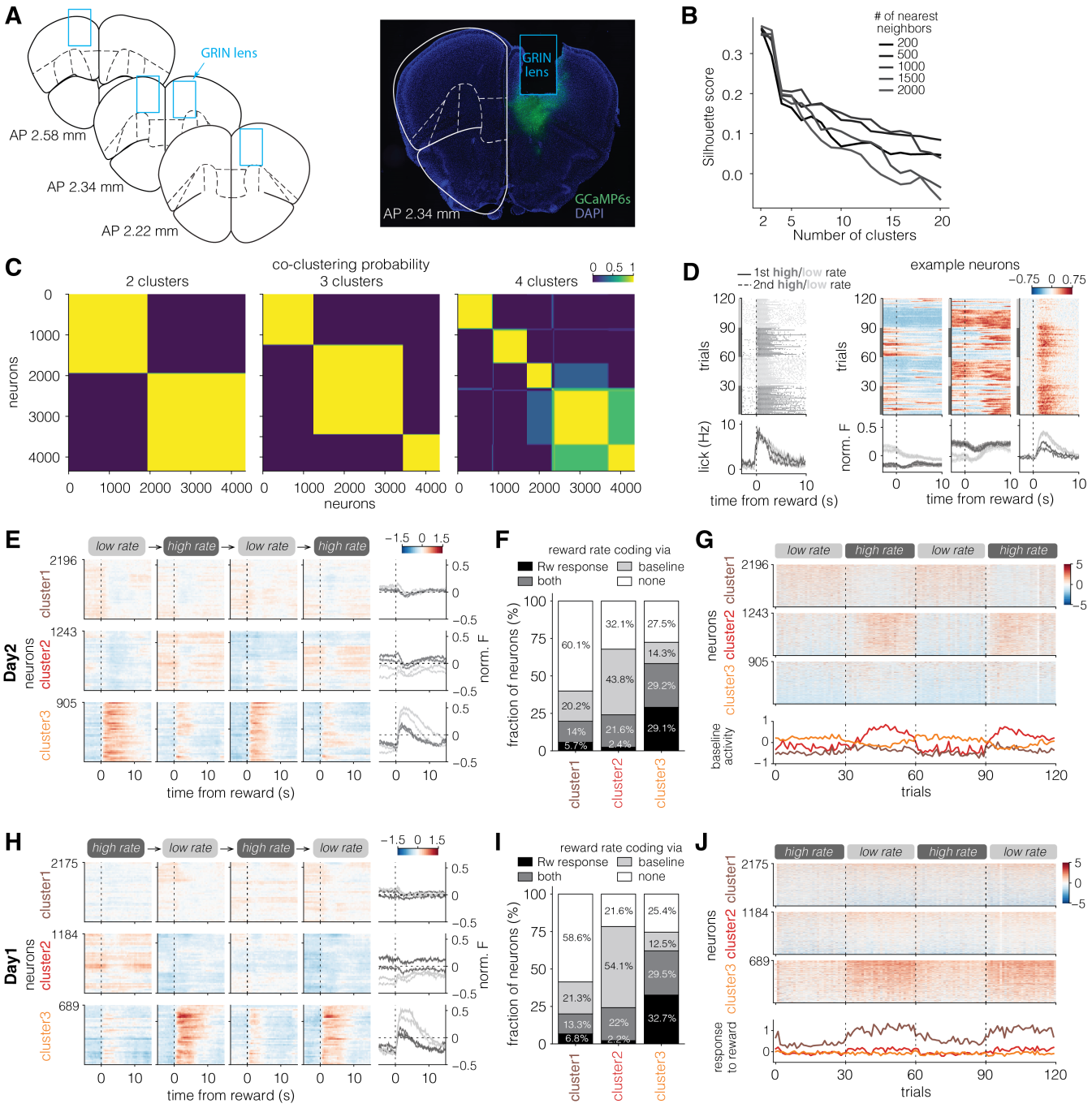

fig. S12. Reward rate coding by OFC neurons was consistently observed across days. A. Histological location of the GRIN lens (left) and representative histological images (right). The left panel shows the position of the lens bottom centered along the anterior-posterior axis. **B.** Silhouette score results from spectral clustering using varying number of clusters (k) and nearest neighbors. Both $k = 2$ or 3 produced similarly high silhouette scores, indicating that neurons within each cluster exhibit cohesive response patterns distinct from those in other clusters, regardless of number of nearest neighbors. **C.** Co-clustering probability across 100 independent clustering iterations with different random seeds, indicating the stability of clustering results. Again, both $k = 2$ or 3 yielded highly stable clustering results. As $k = 2$ and 3 did not differ substantially in clustering validity (silhouette score) or stability (co-clustering probability), we selected $k = 3$ to allow finer subdivision of neuronal response patterns. **D.** Lick rasters and PSTHs from a representative mouse (left) and heatmaps and PSTHs for five example neurons (right). Dark and light gray indicate low and high rate blocks, respectively. Solid line corresponds to the first low/high blocks (first and second blocks of the session), and dashed lines to the second low/high blocks (third and fourth blocks of the session). Time 0 marks the first lick after reward delivery. In case animal does not lick within 5 s after reward delivery, the trial was excluded from analysis. **E.** Heatmaps and PSTHs of clustered neurons (see Methods) averaged across the last three trials of each block. Neurons are sorted by average reward response across blocks. **F.** Fraction of neurons encoding reward rate via changes in phasic reward response, tonic baseline activity, or both, shown for each cluster.

(fig. S12. Continued) **G.** Same as in **Fig. 5H** but for baseline responses. **H–J.** Same analyses as in **fig. S12E,F** and **Fig. 5H**, but from the first recording day in which the order of reward-rate blocks was reversed. Clustering was performed separately for each day and yielded similar cluster structures. **H.** Heatmaps and PSTHs of clustered neurons averaged across the last three trials of each block. Note that the first block (high rate) shows more distinct activity pattern from the second high rate block compared with the nearly identical patterns observed across blocks of the same reward rate on Day 2 (**fig. S12E**), particularly for cluster 2, whose baseline activity encodes reward rate. Because animals had not experienced sucrose rewards or the experimental context (e.g., head fixation box) before Day 1, the distinct activity in the first block likely reflects novelty-related effects in addition to reward rate. **I.** Fraction of neurons encoding reward rate via changes in phasic reward response, tonic baseline activity, or both, shown for each cluster. **J.** Reward responses of individual neurons across trials (top) and averaged response across neurons within each cluster (bottom).

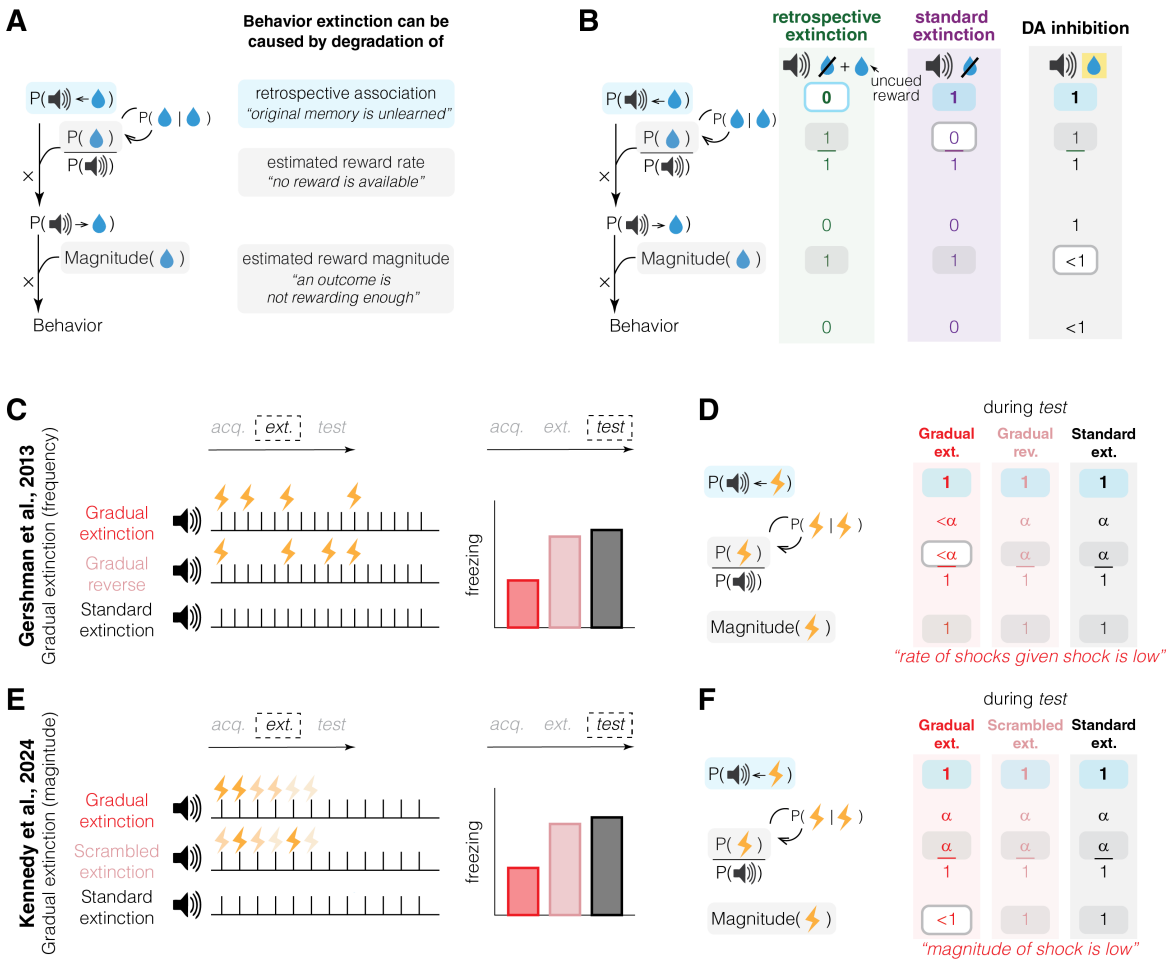

fig. S13. Low estimates of outcome rate or magnitude can induce behavior extinction despite preserved memory A–B. ANCCR predicts three distinct mechanisms for inducing behavior extinction. Extinction via degraded retrospective association (i.e., retrospective extinction; left in B) results in unlearning of the original memory and leads to slow reacquisition. In contrast, extinction caused by reduced estimated reward rate (e.g., standard extinction; middle in B) or reward magnitude (e.g., DA inhibition; right in B) occurs despite the original memory being intact. Degraded component is marked as dark edge around the shaded area in B. **C–D.** Extinction through gradual reduction in frequency of shock can be explained by ANCCR. **C.** Experimental design and finding from Gershman et al. (2013). When shock frequency gradually decreased during extinction (Gradual extinction), retention of behavior was weaker on following test than in Gradual reverse (increasing frequency) or Standard extinction (no shocks during extinction). **D.** ANCCR prediction. We hypothesized that animals estimate the rate of outcome given outcome (i.e., the expected rate of outcomes following a single outcome) and this prior can facilitate rapid reacquisition of behavior when the original memory is intact. This rate is only updated at outcome delivery. In Gradual extinction, the final shock occurs after a long gap since the previous one, reducing the estimated rate of shock given shock. In contrast, in Gradual reverse, the final shock is preceded by another shock after a shorter interval, preserving a higher rate. In Standard extinction, the last shock is experienced during acquisition, when shocks were frequent, so the rate of shock given shock remains high. Thus, only in Gradual extinction, animals show lower level of behavior. **E–F.** Extinction through gradual reduction in magnitude of shock can be explained by ANCCR. **E.** Experimental design and finding from Kennedy et al., (2024). In Gradual extinction, shock magnitude decreased gradually over trials during extinction. In Scrambled extinction, shock magnitude was presented in a randomized order during extinction. Standard extinction involved no shocks during extinction. Weaker freezing was observed in Gradual extinction compared to other conditions during test. **F.** ANCCR prediction. In Gradual extinction, estimated magnitude of shock decreased over time, leading animals to infer that the shock associated with cue is weak. In Scrambled extinction, stronger shocks may occur closer to the test, resulting in a higher estimated magnitude than Gradual extinction. In Standard extinction, the magnitude estimate remains as high as during acquisition.

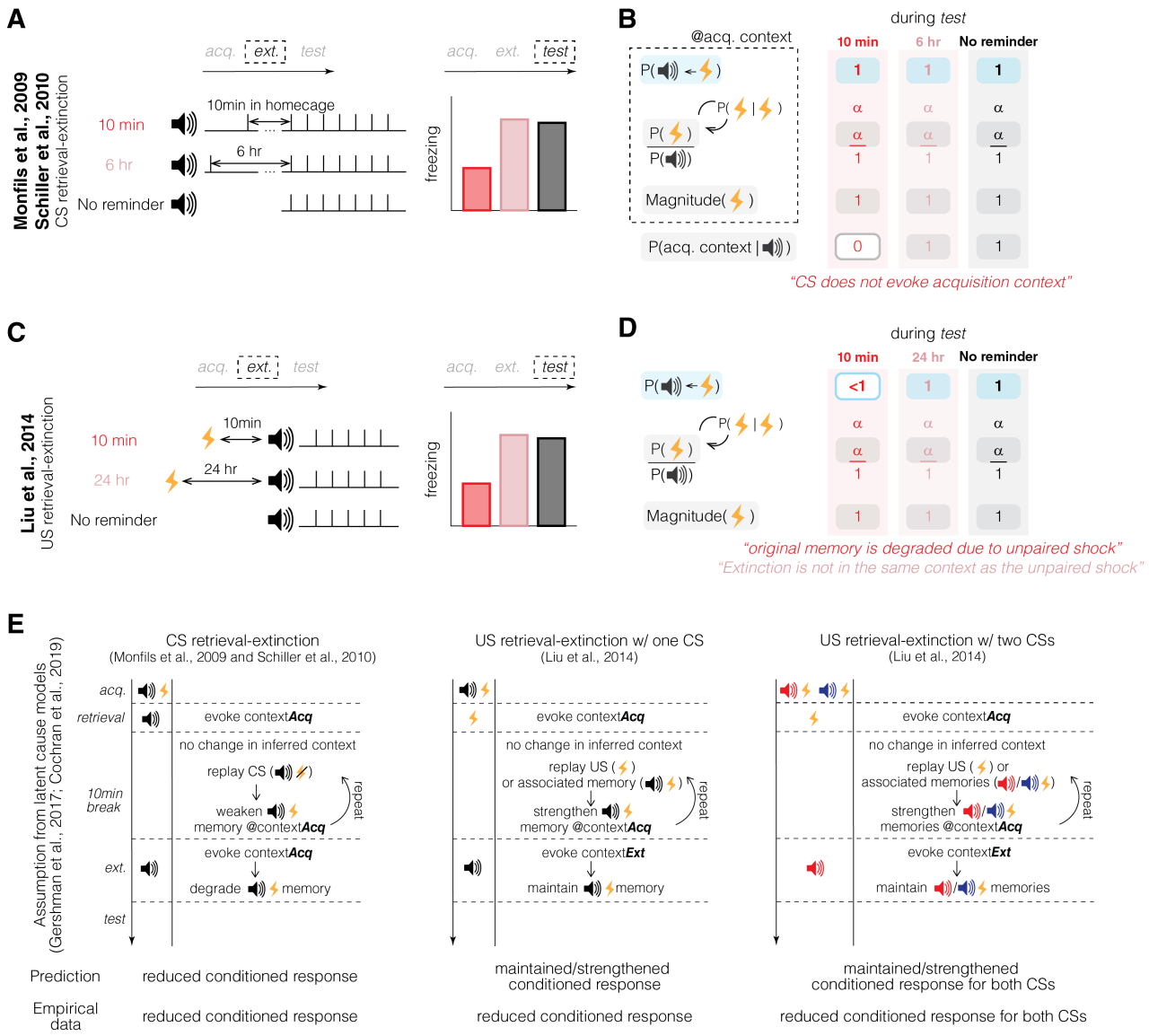

fig. S14. Potential mechanisms to explain retrieval-extinction reconsolidation results. A–B. Conditioned stimulus (CS) retrieval-extinction. **A.** Experimental schematic and findings from Monfils & Schiller. When a retrieval cue is presented shortly before standard extinction (10 min delay), recovery of conditioned behavior is suppressed. However, the effect disappears when the delay is extended to 6 hours. **B.** Potential explanation under ANCCR framework. A context is nothing more than a set of cues. Therefore, animals may infer a context change when the properties of CSs change. In the 10 min condition, the unusually long delay between the retrieval trial and the first extinction trial (compared to ITIs during acquisition) may lead animals to infer a new context. Note that the retrieval trial is not operationally different from extinction trials. Because this new context was never paired with cue-shock memory, conditioned response is low at test. In contrast, when the delay is extended to 6 hours, animals may treat the retrieval trial and extinction session as separate episodes (e.g., different experimental sessions), thus avoiding the context shift during extinction. **C–D.** Unconditioned stimulus (US) retrieval-extinction. **C.** Experimental schematic and findings from Liu et al. (2014). Similar schematic to CS retrieval-extinction, but use US as a retrieval instead of CS. Having retrieval US 10 min before extinction can reduce recovery of behavior at test compared to when delay is long (24 hr) or no US was given (No reminder). **D.** Potential explanation under ANCCR framework. In the 10 min condition, the retrieval US (unpaired shock) may degrade the retrospective association between cue and shock, weakening the original memory. This reduction could be facilitated by replay of unpaired US during 10 min delay. Supporting this idea, the same study showed that US retrieval followed by extinction of one cue reduced conditioned behavior to a different cue that had also been paired with the same shock. This suggests that memory degradation stems from the unpaired shock itself, rather than the extinction procedure. This is consistent with the ANCCR prediction that an unpaired outcome globally weakens retrospective association with its predictive cues. When the delay from US retrieval to extinction is long (24 hr), animals may fail to assign the US to the same context as the rest of experiment, preventing degradation of the retrospective association. A similar idea may explain inter-animal variability in G4 (**fig. S8**), where some animals showed no deficit in reacquisition after unpaired reward-only sessions. **E.** Prediction from latent cause models.

(fig. S14. Continued) In CS retrieval-extinction (left), latent cause models assume that the retrieval trial evokes inference of the acquisition context (contextAcq), and that rumination of retrieval trial (CS alone), which is assumed to not alter the inferred context but produce associative learning, during the 10 min delay weakens the cue-reward association at contextAcq. This reduces RPE during subsequent extinction trial, preventing context switching and resulting in unlearning. When the delay is long (6 hr), the model infers a new context due to a time-based prior favoring different contexts for temporally distant events, thereby preserving the original memory. However, latent cause models cannot account for the effects of US retrieval-extinction (middle and right), since replaying US alone or the CS-US association would not weaken cue-reward memory under the RPE framework (37). Thus, extinction trials following US retrieval should evoke a new context and preserve the original memory, contradicting the empirical findings.

Supplementary tables

Supplementary Table 1. Demographics and cocaine use patterns (mean±STD) of participants. *Three out of 12 participants had not used cocaine during the year post study. Mean number occasions used was based on all 12 participants.

	Participants (n = 12)	Non-respondents (n = 16)
Age	23.1 ± 2.4	24 ± 5.3
Sex		
M	11	15
F	1	1
Cocaine Use (# Occasions used)		
Prior year	19.5 ± 7.3	24.8 ± 10.9
1 year post study	$8.4 \pm 12.8^*$	–
Life time Use	49.8 ± 38.1	55.3 ± 41.4

Supplementary Table 2. Statistical results. *All two sample (independent) t tests were Welch's tests.

Figure	Description	Test	Statistic	p value	Number of samples
Fig. 1I	AUC(norm lick to CS-A)	Paired t test	t = G1, -7.562 ; G2, 2.865	Two-tailed p value = G1, 2.0×10^{-7}; G2, 0.010	n = G1, 22 animals; G2, 19 animals
Fig. 1L	trials to learn CS-A; learners only	Two sample t test*	t = G1, 6.500 ; G2, -0.114	Two-tailed p value = G1, 1.0×10^{-7}; G2, 0.910	n = G1, [22, 22] animals; G2, [19, 12] animals
Fig. 1N	AUC(norm lick to CS-A)	Two sample t test	t = 3.483	Two-tailed p value = 0.002	n = [22, 19] animals
Fig. 1Q	AUC(norm DA to CS-A)	Paired t test	t = G1, -5.154 ; G2, -0.408	Two-tailed p value = G1, 6.0×10^{-4}; G2, 0.696	n = G1, 22 animals; G2, 19 animals
Fig. 2E	AUC(norm lick to CS)	Paired t test	t = G1, CS-A, 1.182 ; G1, CS-B, -2.444 ; G2, CS-A, -3.371 ; G2, CS-B, -5.837	Two-tailed p value = G1, CS-A, 0.2710 ; G1, CS-B, 0.040; G2, CS-A, 9.8×10^{-4}; G2, CS-B, 3.9×10^{-4}	n = Ctrl, 9 animals; Exp, 9 animals
Fig. 2F	AUC(norm lick to CS-A)	Two sample t test with Bonferroni correction	t = G1 vs. G2, 8.864; G1 vs. Ctrl, 8.640; G1 vs. Exp, 1.085; G2 vs. Ctrl, -0.543; G2 vs. Exp, -6.438; Ctrl vs. Exp, -6.139	Two-tailed p value = G1 vs. G2, 5.2×10^{-10}; G1 vs. Ctrl, 7.0×10^{-8}; G1 vs. Exp, 1; G2 vs. Ctrl, 1; G2 vs. Exp, 2.7×10^{-5}; Ctrl vs. Exp, 1.2×10^{-4}	n = G1, 22 animals; G2, 19 animals; Ctrl, 9 animals; Exp, 9 animals
Fig. 2G	Norm consummatory lick	Two sample t test	t = -1.609	Two-tailed p value = 0.138	n = Ctrl, 9 animals; Exp, 9 animals
Fig. 3L	AUC(norm lick to CS); reacquisition	Paired t test	t = G1, -2.077; G2, -5.100; G3, -2.742	Two-tailed p value = G1, 0.050; G2, 7.5×10^{-5}; G3, 0.0254	n = G1, 22 animals; G2, 19 animals; G3, 9 animals
Fig. 4C	AUC(selectivity)	Paired t test	t = G1, -3.990; G2, -0.406	Two-tailed p value = G1, 0.007; G2, 0.699	n = G1, 7 animals; G2, 7 animals
Fig. 4D	Cocaine use	Paired t test	t = 4.001	Two-tailed p value = 0.002	n = 12 participants
Fig. 4G	AUC(selectivity)	Paired t test	t = G1, -3.702; G2, -1.378	Two-tailed p value = G1, 0.008; G2, 0.201	n = G1, 8 animals; G2, 10 animals
Fig. 5D	AUC(norm lick to CS)	Two sample t test	t = 6.576	Two-tailed p value = 4.6×10^{-5}	n = [8, 9] animals for [Ctrl, Exp]
fig. S2B	Averaged lick to CS-A after learning; learners only	Two sample t test	t = G1, -1.269 ; G2, 2.463	Two-tailed p value = G1, 0.211; G2, 0.021	n = G1, [22, 22] animals; G2, [19, 12] animals
fig. S2C	Averaged lick to CS-A	Two sample t test	t = 1.153	Two-tailed p value = 0.256	n = [22, 19] animals for [G1, G2]
fig. S2D	Averaged lick to CS-A	Two sample t test	t = 0.830	Two-tailed p value = 0.412	n = [22, 19] animals for [G1, G2]
fig. S2G, right	AUC(norm lick to CS-A); session number matched	Paired t test	t = G1, -6.130; G2, 0.990	Two-tailed p value = G1, 7.4×10^{-5}; G2, 0.345	n = G1, 12 animals; G2, 11 animals
fig. S2H, right	AUC(norm lick to CS-A); behavior matched	Paired t test	t = G1, -4.961; G2, 5.360	Two-tailed p value = G1, 7.8×10^{-4}; G2, 0.001	n = G1, 10 animals; G2, 8 animals
fig. S3D	AUC(norm lick to CS-A)	Two sample t test	t = -4.286	Two-tailed p value = 2.1×10^{-4}	n = [22, 19] animals for [G1, G2]
fig. S3H	AUC(norm selectivity)	Two sample t test	t = 0.748	Two-tailed p value = 0.470	n = [7, 7] animals for [G1, G2]
fig. S4D	AUC(norm lick to CS-B)	Two sample t test	t = -3.273	Two-tailed p value = 0.003	n = [22, 19] animals for [G1, G2]
fig. S4G	AUC(norm DA to CS-B)	Two sample t test	t = -4.829	Two-tailed p value = 7.1×10^{-4}	n = [10, 8] animals for [G1, G2]

(Continued on next page)

Figure	Description	Test	Statistic	p value	Number of samples
fig. S4K	AUC(norm lick to CS)	Paired t test	t = 4.074	Two-tailed p value = 5.4×10^{-4}	n = 22 animals
fig. S4N	AUC(norm DA to CS)	Paired t test	t = 1.586	Two-tailed p value = 0.147	n = 10 animals
fig. S5C	AUC(norm lick to CS)	Paired t test	t = Ctrl, CS-A, 0.226; Ctrl, CS-B, -2.201; Exp, CS-A, -2.240; Exp, CS-B; -4.618	Two-tailed p value = Ctrl, CS-A, 0.832; Ctrl, CS-B, 0.093; Exp, CS-A, 0.089; Exp, CS-B, 9.9×10^{-3}	n = Ctrl, 5 animals; Exp, 5 animals
fig. S5E	AUC(norm DA to CS)	Paired t test	t = Ctrl, CS-A, -2.562; Ctrl, CS-B, -4.973; Exp, CS-A, 0.044; Exp, CS-B, -1.077	Two-tailed p value = Ctrl, CS-A, 0.062; Ctrl, CS-B, 0.008; Exp, CS-A, 0.967; Exp, CS-B, 0.342	n = Ctrl, 5 animals; Exp, 5 animals
fig. S5G	AUC(norm DA to CS)	Paired t test	t = Ctrl, CS-A, -3.650; Ctrl, CS-B, -8.654; Exp, CS-A, -4.906; Exp, CS-B, -8.083	Two-tailed p value = Ctrl, CS-A, 0.022; Ctrl, CS-B, 9.8×10^{-4}; Exp, CS-A, 0.008; Exp, CS-B, 0.001	n = Ctrl, 5 animals; Exp, 5 animals
fig. S5L	AUC(norm DA to reward)	Two sample t test	t = 6.428	Two-tailed p value = 2.8×10^{-4}	n = [5, 5] animals for [Ctrl, Exp]
fig. S8C	AUC(norm lick to CS) ; acquisition	Paired t test	t = G1, 0.965; G2, 0.638; G3, 1.098	Two-tailed p value = G1, 0.345; G2, 0.531; G3, 0.304	n = G1, 22 animals; G2, 19 animals; G3, 9 animals
fig. S8G	AUC(norm lick to CS); reacquisition	Paired t test	t = -1.469	Two-tailed p value = 0.185	n = 8 animals
fig. S8H	AUC(norm lick to CS); acquisition	Paired t test	t = 0.465	Two-tailed p value = 0.656	n = 8 animals
fig. S8I	Delta AUC(norm lick)	Two sample t test	t = 2.614	Two-tailed p value = 0.016	n = [19, 8] animals for [G2, G4]
fig. S8M	Norm response to CS	Paired t test	t = acquisition, -0.840; deg/ext, 0104; test, 2.839	Two-tailed p value = acquisition, 0.422; deg/ext, 0.920; test, 0.019	n = 10 animals
fig. S8P	Norm lick to CS	Paired t test	t = -0.664	Two-tailed p value = 0.518	n = 8 animals
fig. S10D	AUC(lick to CS+)	Two sample t test	t = 3.877	One-tailed p value = 0.003	n = [7, 7] animals for [G1, G2]
fig. S10H	AUC(Pfreeze to CS+)	Two sample t test	t = 2.139	One-tailed p value = 0.029	n = [8, 10] animals for [G1, G2]
fig. S10J	selectivity	Paired t test	t = 2.629	Two-tailed p value = 0.018	n = 18 animals
fig. S11F	AUC(norm lick to CS)	Two sample t test	t = -2.401	Two-tailed p value = 0.031	n = [8, 9] animals for [Ctrl, Exp]
fig. S11I	Lick to laser	Two sample t test	t = before, 0.312; after, -3.095	Two-tailed p value = before, 0.772; after, 0.013	n = before, [3, 4] animals for [Ctrl, Exp]; after, [8, 9] animals
fig. S11K	Median lick latency	Two sample t test	t = w/ laser, -5.967; w/o laser, 0.249	Two-tailed p value = w/ laser, 3.6×10^{-5}; w/o laser, 0.807	n = [8, 9] animals for [Ctrl, Exp]
fig. S11M	Norm lick to reward	Two sample t test	t = -3.1×10^{-4}	Two-tailed p value = 1	n = [8, 9] animals for [Ctrl, Exp]
fig. S11N	AUC(norm lick to CS)	Two sample t test	t = 3.474	Two-tailed p value = 0.006	n = [7, 6] animals for [Ctrl, Exp]

Methods

Subjects.

Mice: A total 141 adult wild type mice (C57BL/6; #000664; Jackson Laboratory; 82 males and 60 females) were used across cue-sucrose (Group1: $n = 22$; Group2: $n = 19$; Group2 Control for optogenetic dopamine inhibition: $n = 10$; Group3: $n = 9$; Group4: $n = 8$; **Fig. 1–3**), cue-alcohol (Group1: $n = 7$ males; Group2: $n = 7$ males; **Fig. 4A–C**), and cue-shock (Group1: $n = 8$ males; Group2: $n = 10$ males; **Fig. 4E–G**), response elimination ($n = 8$; **fig. S8N–P**), standard extinction with optogenetic OFC inhibition (Ctrl: $n = 15$; Exp: $n = 15$; **Fig. 5A–D**), and random reward experiments ($n = 4$; **Fig. 5E–H**). In addition, 9 DAT-Cre heterozygous mice (B6.SJL-Slc6a3tm1.1(cre)Bkmn/J; #006660; Jackson Laboratory; 3 males, 6 females) were used as Group2 Experimental cohort for optogenetic dopamine inhibition. Animals that did not learn during acquisition phase were excluded from subsequent experiments (Group1/2: $n = 13$; Group3/4: $n = 2$; cue-alcohol: $n = 3$; optogenetic OFC inhibition: $n = 5$; optogenetic dopamine inhibition: $n = 3$; see Behavioral analysis for cue-sucrose task section in Analysis to find detailed criteria for non-learners). Note that a comparable fraction of non-learners ($\sim 20\%$) has been reported previously under similar 50% partial reinforcement condition (**Fig. 6B** in Burke et al. (103)). Additionally, animals that failed to extinguish behavior during extinction phase were excluded (Group2: $n = 1$; Group3: $n = 1$; optogenetic OFC inhibition: $n = 1$). Animals implanted with optic fibers for fiber photometry or optogenetic manipulation or a lens for two-photon imaging were single-housed following surgery, whereas behavior-only animals were group housed. All mice were maintained on a 12 h light/dark cycle. For all experiments except for cue-alcohol and cue-shock conditioning, daily water intake was adjusted to maintain animals at ~ 85–88% baseline weight. Body weight was recorded daily and maintained above 82% pre-deprivation weight. All experimental procedures were performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by the UCSF Institutional Animal Care and Use Committee.

Rats: Ten Long-Evans rats (5 males, 5 females) were used. Rats were food deprived to 85% of their *ad libitum* weight throughout the experiment except for the first week between Extinction/Degradation and Recovery tests (See Behavioral tasks). Food deprivation resumed beginning one week prior to Recovery tests. All experimental procedures were performed in strict accordance with protocols approved by the Animal Care and Use Committee at Johns Hopkins University.

Humans: Twenty-eight non-dependent, non-treatment seeking stable cocaine users (**Supplementary Table 1**) were recruited from the community through advertisements in local newspapers. For all individuals, the primary and preferred route of cocaine self-administration was intranasal. Exclusion criteria included current or past Diagnostic and Statistical Manual of Mental Disorder, Fourth Edition (DSM-IV) (106) dependence (a moderate to severe substance use disorder) on any substance, a current non-substance related axis I disorder, and any cardiovascular, neurological or other medical condition that might be aggravated by study participation. Four participants met DSM-IV criteria for current ($n = 2$) or past ($n = 2$) cocaine abuse (a mild substance use disorder), whereas 6 met criteria for past or current substance abuse of another substance (alcohol: $n = 4$; cannabis: $n = 3$; ketamine: $n = 1$; heroin: $n = 1$). Participants were physically healthy as determined by a medical exam, electrocardiogram and standard laboratory tests. All participants gave written informed consent, and all procedures were carried out in accordance with the Declaration of Helsinki and were approved by the Research and Ethics Board of the Montreal Neurological Institute.

Surgeries. All surgeries were performed under aseptic conditions following procedures similar to those described previously (41). Mice were anesthetized with isoflurane (3–5% for induction and 1–2% for maintenance). A custom-designed head ring was implanted on the skull for head fixation in all mice except for animals used for cue-shock experiments. Mice used for cue-shock experiments or rats did not undergo surgeries. For optogenetic inhibition of VTA dopaminergic neurons (**Fig. 2**), 500 nL of Cre-dependent stGtACR2 (AAV1-hSyn1-SIO-stGtACR2-FusionRed, 6.3×10^{12} GC/ml) was injected bilaterally in VTA (AP -3.2, ML ± 0.99, DV -4.51 with 5° angle). To prevent backflow, the injection needle was left in place for additional 10 min after injection. Optic fibers (NA 0.39, 200 μm, RWD) were implanted 100–200 μm above the virus injection sites. In a subset of animals (cue-sucrose experiment, Group 1: $n = 10$; Group 2: $n = 8$; Group 2 optogenetic dopamine inhibition, control (WT): $n = 5$; experimental (DAT-Cre): $n = 5$), 500 nL of dLight1.3b (AAVDJ-CAG-dLight1.3b, 2.4×10^{12} GC/ml) was injected into the nucleus accumbens core (AP 1.3, ML ± 1.4, DV -4.55) at 100 nL/min to allow simultaneous measurement of dopamine release. An optic fiber (NA 0.66, 400 μm, Doric Lenses) was implanted 350 μm above the virus injection site. For optogenetic inhibition of OFC (**Fig. 5**), 500 nL of either stGtACR2 (AAV1-CamKIIα-stGtACR2-FusionRed, 6.5×10^{12} GC/ml, for experimental group) or mCherry (AAV2-CamKIIα-mCherry, 3×10^{12} GC/ml, for control group) was injected bilaterally in OFC (AP 2.6, ML ± 0.83, DV -2.39 with 10° angle), and optic fibers were implanted 500 μm above the injection sites. For two-photon imaging of OFC neurons, a total of 1 mL of GCaMP6s (AAV9-CamKIIα-GCaMP6s-WPRE-SV40, 3.1×10^{12} GC/ml) was injected unilaterally across two sites in OFC (AP 2.5, ML ± 1.1, DV -2.3; AP 2.9, ML ± 1, DV -2.3). A 1 mm-diameter lens (gradient refractive index lens, GLP1040 Inscopix) was implanted dorsal to the injection sites (centered at AP 2.5, ML ± 0.75, DV -2.2). Animals used in the cue-sucrose, response elimination, standard extinction with optogenetic OFC inhibition, and random reward experiments were water-deprived starting one week before behavioral training and remained water-restricted throughout the experiments. Animals recovered from surgery for at least one week prior to the onset of water deprivation. After completion of experiments, viral expression and optic fiber placements were confirmed using a Keyence microscope.

Behavioral tasks.

Cue-sucrose task (Fig. 1–3, fig. S2–9): To compare two distinct types of extinctions (standard extinction and retrospective extinction) predicted from ANCCR in a way that dissociates predictions of the latent cause model, we designed a Pavlovian conditioning paradigm using two auditory cues. Mice were first trained to associate two 0.25 s tones (12 kHz and 3 kHz) with a delivery of a sucrose drop ($\sim$5 μl) during the *Acquisition* phase. We refer to these reward-associated cues as CS-A and CS-B throughout the study; the cue identities of CS-A and CS-B were counterbalanced across animals. Sucrose was delivered 1s after cue offset with 50% probability. After a 3 s fixed delay following the outcome (reward or omission), the inter-trial interval (ITI) commenced. The ITI was randomly drawn from a truncated exponential distribution (mean = 60 s; maximum = 180 s). Each session consisted of 60 trials (30 CS-A and 30 CS-B). Once animals showed significantly greater anticipatory licking during the 1.25 s cue-reward delay compared to the 1.25 s pre-cue baseline ("lick to CS"), they received at least 3 further days of training to ensure stable asymptotic behavior prior to transitioning to the *CS-A extinction* phase. Animals were then assigned to one of four groups that underwent distinct *CS-A extinction* procedures: 1) Group1 (Standard extinction of CS-A; CS-A omission only daily sessions). Thirty CS-A trials were presented without subsequent reward,

and no CS-B trials were delivered. To match total session duration with *Acquisition*, ITI was extended to 120 s (twice the *Acquisition* ITI). This allows animals to extinguish behavior to CS-A through degradation in estimated reward rate, while leaving CS-A←reward retrospective association intact. 2) Group2 (Retrospective extinction of CS-A; CS-A omission trials intermixed with CS-B trials over many daily sessions). Thirty CS-A omission trials were intermixed with 30 CS-B trials that retained their 50%reward probability. Because rewards delivered on CS-B trials drive degradation of the CS-A←reward retrospective association, this paradigm enables retrospective extinction. ITI duration was kept identical to *Acquisition* (60 s). For a direct comparison between Group1 and Group2, we ran the first two cohorts (Group1: n = 12; Group2: n = 11) in which the number of CS-A extinction sessions was matched (14-16 days for the cohort 1, 18-20 days for the cohort 2). The trial at which the fitted sigmoid reached 25% of the (max-min) range was defined as the moment of extinction, and the session containing this trial was designated as the day of behavioral extinction. Extinction proceeded substantially more slowly in Group2 (6.64±2.9 days, mean±STD) than in Group 1 (1.42±0.49 days) (**fig. S3A–D**).). As a result, Group1 experienced a larger number of *CS-A extinction* sessions after their behavior had already reached asymptotic extinction than Group2. This difference in behavioral history could confound later memory test analyses by making animals with longer behavioral extinction produce slower reacquisition (although we did not find any evidence supporting that Group1 with longer behavioral extinction than Group2 produced faster reacquisition in our post-hoc analysis; **fig. S2**). To eliminate this confound, we ran a third cohort (Group 1: $n = 10$; Group 2: $n = 8$) in which extinction was matched by behavior rather than by number of sessions. Using the first two cohorts, we quantified the number of days from the moment of CS-A behavior extinction to the end of the *CS-A extinction* phase (15.67±2.36 days for Group 1; 10.09±3.15 days for Group 2). To identify the moment of behavioral extinction, we concatenated lick to CS values across extinction sessions and fit the trial-by-trial data with a sigmoid function. For the second cohort, we again fixed the number of *CS-A extinction* sessions for Group 1 (14 days). Group2, however, was trained until each animal experienced the same number of post-behavior extinction *CS-A extinction* sessions as the average number observed in Group2 (∼10 days) of the first cohort. To achieve this in real time while conducting the experiments, for each day, we tested whether lick to CS for both cues remained significantly above zero. Once licking to neither cue was significant, training continued for an additional 10 sessions, corresponding to the average number of post-extinction sessions in Group1 from the first cohort. When we later analyzed the second cohort using the same sigmoid-based curve-fitting method applied to the first cohort, we confirmed that there was no significant difference in number of additional days of *CS-A extinction* following their respective extinction days (Group 1: 12.2±0.87 days; Group 2: 14.63±5.19 days; two-sampled t-test, $p = 0.190$). In the first cohort animals, we tested whether increasing the reward amount on CS-B trials would accelerate extinction in Group2, given the hypothesis that retrospective updating depends on the magnitude of dopamine release at reward. However, extinction speed did not differ between animals that received larger reward (7.17±2.85 days) and those that did not (5.92±2.81 days; two-sampled t-test, $p = 0.695$), thus these data were combined. Subsequent dopamine release measurement from the NAc showed that CS-B rewards elicited large dopamine responses (75.06±23.36% of the unpredicted reward response), likely reflecting partial reinforcement effect, suggesting that the standard sucrose reward already produced sufficient dopamine release for retrospective updating. Finally, comparisons with cue-alcohol experiments suggest that the slower extinction observed in cue-sucrose Group2 may reflect cross-cue generalization (see **Supplementary Note 2**). For optogenetic inhibition of dopamine neurons (**Fig. 2**), a 473 nm laser (8-11 mW, Shanghai Laser & Optics Century Co.) pulse was delivered for 5 s following reward delivery. All optogenetic manipulations were initiated more than three weeks after surgery to ensure sufficient viral expression. 3) Group3 (Retrospective extinction of CS-A; CS-A omission only daily sessions followed by CS-B trials only daily sessions). To dissociate predictions of ANCCR from those of the latent cause model, we separated CS-A omission trials and CS-B trials into different daily sessions. Animals were first trained on CS-A omission only sessions (30 trials per session; identical to Group 1 *CS-A extinction*) for 5 days ("early sessions"), followed by CS-B only session with 50% reward probability for 13 additional days (30 trials per session; "late sessions"). For one animal that did not extinguish lick to CS-A within the initial 5 early sessions, the early session phase was extended to 11 days to ensure at least two complete days of behavior extinction before transitioning to the CS-B only late sessions. In both early and late sessions, the ITI was maintained at 120 s (twice the *Acquisition* ITI) to match overall session duration to *Acquisition*. 4) Group4 (Retrospective extinction of both CS-A and CS-B; CS-A omission only sessions followed by reward only sessions). To further test retrospective extinction under conditions that degrade memories in a non-cue selective manner, the CS-B trials only late sessions in Group3 were replaced with reward only sessions (30 rewards per session). Average inter-reward interval in reward only sessions was 120 s. Under the assumption that animals treat these sessions without CS-A or CS-B as belonging to the same physical context as acquisition, these reward only sessions degrade all retrospective associations linked to the reward, thereby producing retrospective extinction for both CS-A and CS-B. After the *CS-A extinction*, animals in all groups underwent the same *CS-B extinction* phase for 8-10 days (but extended for 5 animals who were resistant to extinction), where 30 CS-A and 30 CS-B trials were intermixed and no rewards were delivered with 60 s mean ITI. This procedure does not modify the CS-B←reward retrospective association, while suppressing behavior through reduction in estimated reward rate; thus, CS-B←reward memory remains intact in Group1-3, whereas it remains degraded in Group 4 due to non-selective retrospective extinction applied during *CS-A extinction* phase. After *CS-B extinction* phase, CS←reward memory was assessed in all groups during the *Reacquisition* phase, which was procedurally identical to the *Acquisition* phase. *Reacquisition* lasted for 7-9 days.

Reward identity specific contingency degradation task (fig. S8L,M): We wanted to test whether retrospective contingency degradation is reward identity specific. To this end, we wanted to conduct experiments with highly sensorily separate rewards (e.g., sweet liquid vs non-sweet solid grain pellets). Because this was challenging in a head-fixed preparation, we performed these experiments in freely moving rats. During the Acquisition phase, two auditory cues were paired with different food rewards over eight daily sessions. The two food rewards (45 mg grain pellets and 0.1 ml of 20% sucrose) were preceded by a pure tone or white noise. Each auditory cue lasted for 20 seconds and was followed by a 50% probability of its associated reward. The ITIs were exponentially distributed with a mean of 4 minutes and a range of 21 s and 14 min. Each session lasted 70 minutes, and within a session, each trial type was presented eight times in pseudorandom order. Rats were guaranteed to experience four rewarded trials and four nonrewarded trials of each type. Cue-reward assignments were counterbalanced with sex. During the Extinction/Degradation phase, one cue underwent extinction while the other underwent contingency degradation over eight daily sessions. The extinguished cue terminated with a reward probability of 0, while the degraded cue offset was still associated with 0.5 reward probability. During the inter-trial intervals, the reward associated with the degraded cue was delivered every 20 seconds with a 50% probability except when the upcoming trial was less than 20 seconds away. This was to avoid noncontingent rewards being delivered during the 20 second pre-cue baseline. Cue-reward assignments, noncontingent reward identity, and sex were counterbalanced. Following the final session

of Extinction/Degradation, rats stayed in their home cage for two weeks and were then placed in two spontaneous recovery tests. During each test, the auditory cues were presented with the same frequency and pseudorandom order as the prior phases, but no rewards were delivered.

Response elimination task (fig. S8N–P): Classic autoshaping studies in pigeons have shown that presenting uncued rewards while omitting rewards on CS+ trials ('negative contingency' or contingency degradation in general) eliminates the conditioned response but does not erase the underlying cue-reward memory (47, 48, 107). Evidence comes from the observation that, after response elimination, subsequent standard extinction sessions often produce a recovery of conditioned responding, indicating that the original memory remains intact. This behavioral pattern directly contradicts a key prediction of ANCCR, which proposes that uncued outcomes should trigger retrospective updating that weakens the original association and produces true memory erasure. To test whether uncued rewards can erase memory in rodents under a head-fixed paradigm, we adapted the classic pigeon 'negative contingency' design (47). We used partial reinforcement conditions, because prior study (107) reported stronger recovery following response elimination when acquisition involves partial reinforcement. During Acquisition, mice learned that 0.25 s auditory cue (12 kHz tone) predicted a 5 ul sucrose reward with 50% probability, delivered after 1 s trace interval. The ITI was uniformly distributed between 240 s to 360 s, and each session contained 12 trials. Once lick to CS was significantly greater than zero for 5 days, animals advanced to Response elimination phase. In this phase, all cue-paired rewards were omitted, and instead six uncued rewards were delivered at random ITI times. This design directly mimics the 'negative contingency' paradigm used in the pigeon autoshaping literature (47). Response elimination continued until mice showed no significant CS-evoked licking for more than 5 consecutive days, or until a total of seven days without significant CS licking was reached. After response elimination, mice underwent three standard extinction sessions, where only the cue was presented without any associated or unassociated rewards. These sessions were used to determine whether behavioral recovery occurred.

Cue-alcohol task (Fig. 4A–C, fig. S10A–E): To familiarize mice with alcohol, all animals underwent an intermittent access two bottle choice procedure prior to cue-alcohol conditioning task. Animals received three 24 h drinking sessions per week with continuous access to a 20% (v/v) ethanol bottle and a water bottle, separated by 24 h and 48 h withdrawal periods during weekdays and weekends, respectively. During withdrawal periods, animals had access to water only. The positions of alcohol and water bottles were alternated at the start of each drinking session to control for side preferences. A 'spill cage' without an animal was used to measure evaporation or spillage, and consumption amounts were corrected accordingly. Alcohol and water intake were measured for each 24 h drinking session. After 14 drinking sessions, animals consumed an average of 10.88±2.51 g/kg (mean±STD) of alcohol and displayed an average alcohol preference (over water) of 0.50±0.11, with no difference across groups (consumption, G1: 10.65±2.43, G2: 11.11±2.75, two-sampled t-test, $p = 0.746$; preference, G1: 0.51±0.10, G2: 0.49±0.13, two-sampled t-test, $p = 0.744$). Because this study was not intended to model severe alcohol dependence but rather to use alcohol as an appetitive outcome for cue-outcome conditioning, we included all animals regardless of their final alcohol intake or preference. Importantly, all mice exhibited robust consummatory licking for alcohol during the subsequent head-fixed cue-alcohol conditioning task.

To test standard vs. retrospective extinction of cue-alcohol memory, we used Pavlovian conditioning paradigm similarly to the cue-sucrose task. During the *Acquisition* phase, animals were trained to associate one auditory cue (CS+) with alcohol delivery and a second cue (CS-) with no outcome. Cue identities (3 kHz and 12 kHz) were counterbalanced across animals. Each cue was presented for 0.25 s, followed by a 1 s trace period, after which a drop of alcohol ($\sim$9 μl) was delivered on CS+ trials. Following the outcome, the ITI commenced. The ITI was drawn from a uniform distribution between 240 s and 360 s. Each session consisted of 6 CS+ trials and 6 CS- trials. Behavior was quantified using a selectivity index, defined as selectivity index= (mean(lick to CS+)−mean(lick to CS-))/$\sqrt{\mathrm{var(lick\ to\ CS+)} + \mathrm{var(lick\ to\ CS\text{-})}}$, which measures the separation between CS+ and CS- licking responses. Animals were trained until they exhibited > 6 days with a selectivity index greater than 1 before transitioning to the *Extinction* phase. Average selectivity index was not different across groups for the last 6 days of *Acquisition* (Group1: 1.77±0.57; Group2: 1.12±0.75; two-sample t-test, $p = 0.121$). For *Extinction*, animals were divided into two groups. 1) Group1 (Standard extinction). All rewards in CS+ trials were omitted. Animals therefore experienced no alcohol during extinction in this group. This procedure suppresses behavior by reducing the estimated reward rate but leaves the original CS+←alcohol memory intact. 2) Group2 (retrospective extinction). Rewards in CS+ trials were omitted as in Group1, but animals additionally received 6 uncued alcohol rewards delivered at random times during ITIs (half following CS+ trials and half following CS- trials). To maximize separation between uncued rewards and cue presentations while preserving a sufficiently large window for random timing, the first and last 10 s of each ITI were excluded from the possible reward delivery windows. These uncued rewards drive retrospective degradation of CS+←alcohol memory. Extinction training continued until each animal exhibited > 6 days with a selectivity index smaller than 0.2. Average selectivity index was not different across groups for the last 6 days of Extinction (Group1: 0.07±0.11; Group2: -0.04 ± 0.24; two-sample t-test, $p = 0.327$). After extinction, memory was tested during Reacquisition phase, which was trained for 17 days and was procedurally identical to the *Acquisition* phase.

Cue-cocaine task (Fig. 4D): Human participants were recruited following a three-stage screening procedure. First, a semi-structured telephone screening was conducted. Second, those who were interested and tentatively met the entry criteria were invited for a face-to-face psychiatric interview, using the Structural Clinical interview for DSM-IV (SCID) (106). Third, volunteers underwent a medical examination, conducted by a physician, to ensure that it was medically safe for volunteers to participate in the study. Participants took part in one of three different studies involving intranasal cocaine self-administration (108–111). For each study, in-depth information was obtained from participants about their cocaine use histories using a timeline follow-back (TLFB) procedure (112), including number of occasions used per month during the past year, overall lifetime use, estimated amount (milligrams) used per occasion and routes of administration (see **Supplementary Table 1**).

Detailed characteristics of the cocaine self-administration procedures are described in previous reports (108–111). In each study participants self-administered between 1 mg/kg and 5.1 mg/kg cocaine, taken intranasally, on two to four separate occasions, in a hospital setting. Throughout testing, a project-dedicated nurse and emergency physician were on-site. Following testing, participants were kept overnight for observation with a nurse onsite and an emergency physician on call. There were no serious adverse effects. Since these two participants indicated no further interest in study participation, their information was excluded from this report.

For follow-up recruitment, participants were re-contacted via telephone or email. They were informed that additional information was required for a follow-up project of the initial study they participated in. They were not told that these self-reports would be compared

to information obtained prior to the initial study. The sample consisted of 12 respondents (**Supplementary Table 1**). Sixteen participants could not be re-contacted, having either changed their telephone number or email address without notifying us, or did not respond to multiple re-contact attempts. Thus, the present findings need to be interpreted in light of the following considerations. First, the sample size of 12 (out of 28) respondents is modest. Although no differences were observed between respondents and non-respondents a sample bias cannot be excluded. The time interval between participation in the initial study and follow-up evaluation was on average 21.5±9.4 months (mean±STD). Two of the 12 respondents met criteria for at least 1 past substance use disorder (Subject 1: cocaine and cannabis abuse and Subject 2: alcohol and ketamine abuse) prior to study participation. Respondents were questioned about their cocaine use frequency and route of administration during the 12-month period immediately after study participation using a TLFB procedure (112) (**Supplementary Table 1**). In addition, the incidence of potential cocaine use disorders was assessed, using the same diagnostics (DSM-IV) as those used during the SCID interview prior to participation of the initial study. Eight participants were re-interviewed on the telephone, four in person. Since there were no differences in cocaine use between subjects from each interview medium, their data were combined.

Cue-shock task (Fig. 4E–G, fig. S10G–L): Mice were handled for 2-5 minutes per day for the 5 days preceding cue-shock conditioning. Animals were placed into a sound-isolating chamber and received auditory cues and/or electric footshock (outcome) controlled by MED-PC (MED Associates). During *Acquisition*, animals were trained to associate one auditory cue (CS+) with a 0.35 mA footshock and a second auditory cue (CS-) with no outcome. Cues were 30 s tones (5 and 10 kHz), with CS+ and CS- identities counterbalanced across animals; on CS+ trials, the cue co-terminated with a 1-s shock. Each session consisted of 5 CS+ and 5 CS- trials. Trials were separated by an ITI drawn from a uniform distribution between 108 s and 134 s (mean of 120 s with a 10% jitter). Learning was assessed using the fraction of immobility during first 10 s of CS period. The fraction of immobility during the 10 s pre-cue baseline was also quantified. Animals were trained for 10 days, during which the population-averaged fraction of immobility during CS+ was significantly larger than that during the CS- and baseline periods for last 7 days. To quantify learning at the individual-animal level, we used the same selectivity index applied in the cue-alcohol task, substituting the average fraction of immobility during the CS for the lick to CS measure. Animals were then divided into two groups such that the average selectivity index on the last 3 days of *Acquisition* was not different across groups (Group1: 1.36±0.50; Group2: 1.39±0.45; two-sample t-test, p = 0.897) before transitioning to *Extinction*. During the *Extinction* phase, two groups went through different procedures. 1) Group1 (Standard extinction). All shocks in CS+ trials were omitted. 2) Group 2 (Retrospective extinction). Shocks in CS+ trials were omitted as in Group1, and animals additionally received 4 uncued shocks at random times during ITIs (2 following CS+ trials and 2 following CS- trials). The first and last 12 s of each ITI (10% of mean ITI) were excluded from the possible shock delivery windows. During the first two extinctions sessions, 10 uncued shocks (5 following CS+ trials and 5 following CS- trials) were mistakenly delivered. Because animals completed 15 total Extinction sessions (i.e., 13 correctly configured sessions following the first two), it is unlikely that the effects of these initial sessions persisted into the later stages of Extinction. After 15 days of Extinction with the comparable immobility fraction across cues for the last 2 days (Group1: 0.20±0.68; Group2: 0.22±0.46; two-sample t-test, p = 0.96), memory was tested during 14 days of *Reacquisition*.

Note that we used ABB design, in which *Acquisition* occurred in context A, and *Extinction* and *Reacquisition* occurred in context B. This design was adopted to dissociate the contributions of retrospective memory and reward-history-based priors to the speed of reacquisition. In ANCCR, animals estimate event rate (e.g., reward/shock rate) in a context-specific manner, and these estimates are restored quickly at the beginning of each session (or during reacquisition after long absence of outcomes) once any outcome is experienced in that context possibly using memory of probability of outcomes given an outcome (e.g., $P(\text{reward} \mid \text{reward})$ or $P(\text{shock} \mid \text{shock})$). Thus, rapid reacquisition in Group1 during cue-sucrose/alcohol experiments could, in principle, be attributed to a high prior of $P(\text{reward} \mid \text{reward})$ as it does not update during extinction due to absence of rewards and remains as high as in *Acquisition*. By switching to a novel context B for extinction and reacquisition, Group1 enters *Reacquisition* with $P(\text{shock} \mid \text{shock}) = 0$ because no shocks were ever experienced in context B, eliminating any inflated prior of estimated shock rate. In contrast, Group2 receives uncued shocks in context B, resulting a nonzero $P(\text{shock} \mid \text{shock})$ at the start of *Reacquisition*. Therefore, if reacquisition were driven by shock-history priors (higher in Group2 than Group1) rather than retrospective memory (erased in Group1, while preserved in Group2), Group2 should reacquire faster than Group1. Using ABB design thus isolates retrospective-memory effects from confounding priors based on outcome statistics. Context 1 consisted of entirely grey siding, a rectangle arena shape, purple light, and vanilla scent, while context 2 consisted of vertical, high contrast stripes, a hexagonal arena, green light, and lemon scent. Both contexts have a shock grid floor. Context identities were counterbalanced across animals.

Simple extinction task with OFC optogenetic inhibition (Fig. 5A–D, fig. S11): During *Acquisition*, 0.25 s auditory cue predicted delivery of a 5 μl sucrose reward with 50% probability following a 1 s trace interval. The ITIs were drawn from a truncated exponential distribution (mean = 120 s, maximum = 360 s) and began after a 3 s fixed post outcome delay. Once animals showed significantly elevated cue-evoked licking for more than 5 days, they advanced to the *Extinction* phase, in which rewards were omitted. After >2 days in which lick to CS was no longer significant, animals proceeded to *Reacquisition*. OFC activity was optogenetically inhibited during one of two task epochs: 1) cue-reward delay period (1.25 s; 473 nm, 8–11 mW, Shanghai Laser & Optics Century Co.), or 2) outcome period (5 s). Each animal experienced both inhibition conditions, achieved by running two sequential extinction-reacquisition cycles (e.g., *Acquisition* $\rightarrow$ *Extinction* $\rightarrow$ *Reacquisition* with OFC inhibition #1 $\rightarrow$ *Extinction* $\rightarrow$ *Reacquisition* with OFC inhibition #2). The order of inhibition periods was counterbalanced across animals. During the outcome period inhibition condition, laser delivery was terminated on Day 4 of *Reacquisition* to test whether animals rapidly reacquired the conditioned response once OFC inhibition was removed. In a subset of animals ($n = 3/15$ for Ctrl, $4/15$ for Exp), thirty 5 s laser pulses were delivered at random times during ITI to test whether laser alone could elicit licking. This test was performed twice: once after animals had learned the task but before *Extinction* ('test laser before reacq.' in fig. S11G), and once after *Reacquisition* ('test laser after reacq.'). Following the 'test laser before reacq.' session, animals underwent one additional acquisition session without laser before proceeding to extinction. To test the effect of OFC outcome period inhibition on initial acquisition, a separate cohort underwent *Acquisition* with 5 s OFC inhibition during outcome period (fig. S11N). These animals did not undergo *Extinction* or *Reacquisition*.

Random reward task with OFC imaging (Fig. 5E–H, fig. S12): Low (75 s inter-reward interval) and high (15 s) reward rate blocks were alternated twice within each session. Each block consisted of 30 rewards ($\sim$3 μl sucrose). The experiment was conducted over two days with reversed block order across days (Day 1: high$\rightarrow$low$\rightarrow$high$\rightarrow$low; Day 2: low$\rightarrow$high$\rightarrow$low$\rightarrow$high). Day 2 was the primary data analyzed due to the animals already being habituated to reward delivery on the previous day.

Photometry recording. After three weeks from the virus injection, fiber photometry recordings were performed using an open-source system (PyPhotometry). In PyPhotometry, two LED wavelengths were emitted alternatively at 130 Hz through a fluorescence mini cube (Doric Lenses) using time-division sequence. Fluorescence signals were collected by a separate photodetector (Doric Lenses). Light intensity at the tip of the patch cord was maintained at 40 µW (470 nm) and 20 µW (405 nm) across sessions. Raw fluorescence signals were bandpass filtered (0.01–20 Hz) to remove low-frequency drift caused by photobleaching or motion artifacts and to suppress high-frequency electrical noise. The 405 nm signal was then linearly regressed onto the 470 nm signal to obtain a fitted isosbestic reference. Fractional fluorescence signal ($\Delta F/F$) was computed as (470 nm signal – fitted 405 nm signal)/(fitted 405 nm signal).

Two-photon microscopy. Calcium activity of OFC neurons was imaged using two-photon microscopy after expressing the calcium indicator (GCaMP6s) under CamKIIα promoter. A minimum of 6 weeks was allowed for sufficient viral expression before imaging began. Imaging was performed using Bruker Ultima 2pplus two-photon microscope equipped with a Cousa 10x long working distance objective lens (113) and a 920 nm excitation laser (SpectraPhysics). A resonant scanner combined with an electrically tunable lens was used so simultaneously image three planes at 8 Hz per plane. Imaging planes were spaced >50 µm apart to minimize overlap of cells across adjacent planes. Laser power at the sample was maintained at 25–35 mW throughout sessions. Animals were positioned on a tip, tilt, and rotation stage (Thorlabs) to precisely align the surface of lens perpendicular to the excitation path. Goniometer readings were recorded on each day, and imaging planes were compared

Analysis.

Behavioral analysis for cue-sucrose task: For the cue-sucrose task, memory was assessed by quantifying the speed and strength of reacquisition using two complementary measures. First, we computed the area under curve (AUC) of lick to CS vs. trial plot (e.g., **Fig. 1I**; used throughout the paper). Lick to CS was defined as the difference in lick counts between post-cue and pre-cue 1.25 s windows. Lick to CS values were normalized to the average lick to CS measured during last 3 days of acquisition for each animal to account for variability in baseline lick rates across animals. AUC was calculated over the last day of extinction and the following first 3 days of reacquisition. This metric was used as the primary measure because it allows inclusion of all animals, including those that did not show clear evidence of behavioral reacquisition. We used the same AUC metric to measure the speed of extinction (**fig. S3,4**). However, because AUC reflects both the speed and strength (i.e., asymptotic level of licking) of reacquisition, we additionally used cumulative sum-based analysis to dissociate these components (e.g., **Fig. 1J–L**, **fig. S2A,B**).

For cumulative-sum analysis, we plotted the cumulative sum of lick to CS across trials and identified the learned trial as the first trial at which the distance between the cumulative-sum curve and the unity (diagonal) line exceeded 80% of the maximum distance. Because this threshold is relative, it can identify a learned trial even when the cumulative-sum trace remains close to zero. To prevent classifying such cases as learners, we computed the average lick to CS using trials after the detected learned trial ('average lick to CS after learning; **fig. S2B**) and excluded any animals with < 0.5 licks on average as non-learners. We included 2 extinction sessions and 6 reacquisition sessions for this analysis. To directly compare acquisition and reacquisition learned trials directly, we prefixed the cumulative-sum vector with an arbitrary list of 30 zeros (30 trials/session $\times$ 2 sessions) corresponding to extinction trials, followed by 6 acquisition days when performing cumulative-sum analysis during acquisition.

Memory was additionally assessed by measuring spontaneous recovery during extinction (**Fig. 1M,N**). Because the number of extinction days experienced after behavior had extinguished differed across groups—and spontaneous recovery is known to weaken over the course of extinction—we selected extinction sessions in a way that matched behavioral history across groups. In Group2, extinction of CS-A occurred in the presence of a non-zero reward rate (due to ongoing rewards in CS-B trials), which could contribute spontaneous recovery and confound the interpretation of memory strength. Therefore, for Group2, we used the 8 days of CS-B extinction phase, during which no rewards were delivered, as the comparison window. We first computed, across Group2 animals, the average number of days with no significant lick to CS-A (13$\pm$4.7 days) during the *CS-A extinction* phase. For each Group1 animal, we then identified the extinction day on which it had experienced the same number of days without significant CS-A licking as the Group2 average. The eight subsequent extinction days in each animal were used for comparison with the corresponding *CS-B extinction* window in Group2. Because spontaneous recovery is mostly prominent at the beginning of each session, we quantified it by computing the AUC of lick to CS vs. trial plot using only the first 5 trials of each session.

Analysis of dopamine response for cue-sucrose task: Dopamine learning was similarly assessed as behavioral learning by substituting 'lick to CS' with 'DA to CS'—defined as the difference in area under the dLight fluorescence trace between post-cue and pre-cue 1.25 s window (**Fig. 1Q**). DA to CS value was normalized to the maximum of the mean dopamine responses to rewards during the first 3 days in learning, measured as the difference in fluorescence AUC between the post-reward and post-cue 1.25 s windows. This normalization enables comparison across animals by accounting for differences in overall fluorescence that may arise from variability in viral expression or in the alignment between fiber tip and the expression site. For **fig. S5**, we showed the DA to CS normalized to asymptotic DA to CS for each phase (acquisition or reacquisition) as well. To quantify the effectiveness of optogenetic inhibition of dopaminergic neurons in **Fig. 2D**, DA to reward with optogenetic inhibition was compared to that without inhibition. For this specific analysis, we used 5 s analysis window as inhibition lasted for 5 s. Dopamine dip at reward omission (DA to omission) was similarly measured as difference in fluorescence AUC between post-omission (when reward was expected) and post-cue 3 s windows (**fig. S9E,N,O**).

Behavioral analysis for cue-alcohol and cue-shock task: Learning was quantified using two complementary approaches. First, conditioned behavior was measured on trial-by-trial basis as in cue-sucrose task. For cue-alcohol task, lick to CS was defined as the difference in lick counts between post-cue and pre-cue 1.25 s windows. For cue-shock task, Pfreeze to CS was defined as the difference in immobility fraction between post-cue and pre-cue 10 s windows. We used a 10 s window in the early cue period because immobility differences were largest early in the cue (**fig. S10K,L**), consistent with a previous report that showed a transition from freezing to early cue to flight behavior to late cue when two cues serially precede a foot shock (114). Second, because trial-by-trial analysis does not account for baseline behavioral differences across group (e.g., during CS- and pre-cue periods; this difference is strongest for cue-shock task; see **fig. S10G**), we computed a session-by-session selectivity index (equation below; **Fig. 4**). Behavioral readouts were licks to CS in cue-alcohol task and Pfreeze to CS in

cue-shock task.

$$\text{selectivity index} = \frac{\text{mean(behavior to CS+)} - \text{mean(behavior to CS-)}}{\sqrt{\text{var(behavior to CS+)} + \text{var(behavior to CS-)}}}$$

On a few sessions, recording files were corrupted and could not be accessed, and therefore excluded from analysis.

Behavioral analysis for reward identity specific contingency degradation task: Throughout the experiment, conditioned responding was measured as the mean rate of anticipatory reward port entries during the 20 second cues minus the entry rate during a 20 second pre-cue baseline (**fig. S8M**). Conditioned responding during the two recovery tests were averaged together.

Behavioral analysis for response elimination task: Lick to CS was computed in the same manner as in cue-sucrose task. To test whether behavior recovered during standard extinction following unpaired response elimination, we compared the average lick to CS during the last three days of unpaired response elimination phase with the average lick to CS on the first day of standard extinction (**fig. S8P**).

Two-photon imaging data analysis: Raw imaging files acquired using Bruker Prairie View were motion-corrected, and regions of interests (ROIs) were detected using Suite2p (115). All detected ROIs were visually inspected by experimenters to confirm that their size and shape were consistent with neuronal soma and not dendritic segments or any other noise. Raw fluorescence signals (F) from each ROI were corrected for neuropil signal (F_neu) as $F' = F - 0.7 \times F_\text{neu}$. Then we calculated normalized fluorescence signal as the ratio $(F' - F'_{\text{median}})/(F'_{\max} - F'_{\min})$ and normalized fluorescence was used for all subsequent analyses. If two ROIs recorded from adjacent planes (> 50 μm apart) exhibited a cross correlation greater than 0.8 in their normalized fluorescence traces and were separated by less than 50 μm in the x–y plane, the ROI with the lower variance in normalized fluorescence (presumed to have lower signal-to-noise ratio) was excluded.

Neurons recorded on each day were pooled across four animals (4048 on Day1, 4344 on Day2) for clustering analysis. For each neuron, pre-reward and post-reward activity were quantified as AUC of the normalized fluorescence signal within a 5 s window before and after the first lick following each reward delivery, respectively. If an animal did not lick within the first 10 s after reward delivery, that reward was excluded from analysis. Pre-reward and post-reward activity across trials (30 trials $\times$ 4 blocks) for all neurons were then concatenated and subjected to principal component analysis (PCA) for dimensionality reduction. The number of principal components retained was determined using an elbow detection method on the scree plot, which identifies the point of maximum deviation from a straight line connecting the first and last components. Neuronal activity was then projected onto the retained principal component axes, and neurons were clustered using the spectral clustering algorithm. To determine the optimal clustering parameters (number of clusters (k) and possible n nearest neighbors), we performed clustering across all parameter combinations for 10 iterations each, sweeping k from 2 to 20 and the number of nearest neighbors from [200, 500, 1000, 1500, 2000]. Clustering quality was evaluated using two independent metrics: 1) silhouette score and 2) co-clustering probability across iterations. Both measures were comparably high at $k = 2$ and $k = 3$ (**fig. S12B,C**); we therefore selected $k = 3$ to obtain a finer classification of neuronal populations. At $k = 3$, the silhouette score was maximized when number of nearest neighbors was 2000. Using these optimized parameters ($k = 3$ and number of nearest neighbors=2000), we performed a final spectral clustering to obtain the clustering result. This clustering pipeline followed the procedure previously reported in (74).

To test whether individual neurons encoded reward rate through their phasic reward response, we compared the difference between post-reward and pre-reward activity across two distinct reward rate blocks (low vs. high) using a t-test. Neurons with a p value < 0.01 were classified as reward rate coding via reward response. Similarly, to identify neurons encoding reward rate through baseline activity, we compared pre-reward activity alone between the low- and high-reward rate blocks using the same statistical criterion (**fig. S12H,I**). Response to reward was defined as the difference between post-reward and pre-reward activity (**Fig. 5H**, **fig. S12J**) and baseline activity was defined as pre-reward activity (**fig. S12G**).

Simulations.

Tasks: Each experiment was simulated for 20 iterations. For ANCCR and latent cause models, we simulated 300 trials per CS type in each phase. For value RNN, each iteration consisted of 50 training epochs, with 5000 trials per CS type in each phase and a batch size of 12 episodes.

ANCCR: We previously proposed a learning model, ANCCR, based on learning of retrospective associations. ANCCR identifies cues that consistently precede meaningful events such as rewards, thereby learning a retrospective cue←reward association. This retrospective association can then be converted into a prospective cue→reward association through a Bayes' rule like computation:

$$M_{\to cr} = M_{\leftarrow cr}\, M_{\leftarrow r-} / M_{\leftarrow c-},$$

where $M_{\leftarrow r-}$ is proportional to the baseline rate of a reward in the environment, and $M_{\leftarrow c-}$ is proportional to the baseline rate of a cue. Here, $M_{\to cr}$ corresponds to the successor representation (a measure of prospective association that is not baseline-corrected), while $M_{\leftarrow cr}$ represents the predecessor representation (a measure of non-baseline-corrected retrospective association). For intuitive clarity throughout the paper, we denoted these as $P(\text{cue} \to \text{reward})$ and $P(\text{cue} \leftarrow \text{reward})$, respectively. Because ANCCR is fundamentally driven by retrospective association, it predicts true learning and unlearning occur only when rewards are experienced. Consequently, the retrospective association functions as the original memory, providing an explanation for the rapid reacquisition observed after standard extinction.

We used the same ANCCR model described previously (103) but introduced one modification: the meaningfulness of an event was re-evaluated each time the event occurred, rather than treated as permanently meaningful once it first crosses a threshold. In the original implementation, meaningfulness became permanent when the sum of an event's innate meaningfulness and its learned meaningfulness (reflected in the dopamine response) exceeded a threshold. This simplification was appropriate for scenarios in which meaningfulness does not change after learning—for example, rewards are innately meaningful, and cues become permanently meaningful after cue-reward association is formed. However, optogenetic inhibition of dopamine neurons at the time of reward suppressed the dopamine response below baseline, which in turn should reduce the total meaningfulness (innate+learned) of the reward and thereby suppress retrospective update. To capture this effect, we modified the model so that the meaningfulness of an event is computed dynamically on every occurrence, allowing meaningfulness of events (even the innately meaningful events like reward) to reduce when dopamine responses are arbitrarily reduced. In addition, we assumed that animals use their prior $P(\text{reward} \mid \text{reward})$ to rapidly restore the estimated reward rate when they returned to the reward-rich environment

(e.g., reacquisition) after prolonged absence of rewards (e.g., extinction). For simplicity, we implemented this prior-based inference only on the first trial of the reacquisition phase for all groups, although the actual biological implementation may differ. For all simulations, we used the following parameters: $w = 0.5$, $\alpha_0 = 4 \times 10^{-5}$, $M_{\text{opt}} = 0.5$, $dt = 0.2$ s. T was defined as

$$T = M_{\text{opt}} \sum_x \lambda_x \Big/ \sum_x \lambda_x^2,$$

where λ_x is an inverse of average inter-event interval for each event x. This ensures that the eligibility timescale matches the timescales operating in the environment. α was defined as

$$\alpha = 1 - (1 - \alpha_0)^{(\text{IRI}/dt)},$$

where IRI is inter-reward interval. This scaling ensures that learning of baseline event rates (e.g., $M_{\leftarrow r-}$ and $M_{\leftarrow c-}$), which are updated using the learning rate α_0, and learning of retrospective association ($M_{\leftarrow cr}$), which is updated using learning rate α, proceed on comparable timescales. For additional details, see Methods and Supplementary Note of Burke et al. (103). Dopamine response during optogenetic inhibition of dopaminergic neurons was modeled as -1.5 (1 simulated dopamine response to an unexpected reward).

Learning speed was quantified using simulated baseline-corrected prospective association (successor representation contingency, SRC). Specifically, we computed the area under the SRC vs. trial curve for the first 90 trials of acquisition or reacquisition.

Latent cause models: We simulated three different latent cause models—Redish (2007) et al. (39), Gershman (2017) et al. (37), and Cochran (2019) et al. (36)—as illustrated in textbffig. S6A-D. For each model, we systematically swept the model parameters most relevant to late state inference, as described below. All simulations for latent cause models were performed by modifying the scripts provided by Cochran et al. on Github (https://github.com/cochran4/OnlineLatentStateLearning).

Redish (2007) et al. (39)

- ϑ_P (state-creation threshold): determines when a new state is created based on the activation of existing states; lower values increase the likelihood of creating new states; default, 10^{-8}; swept, [10^{-10}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}]
- ξ_0 (decay rate of $\bar{\delta}$, the running average of negative RPEs): controls how long past negative RPEs influence $\bar{\delta}$; higher values integrate history over longer timescales, increasing evidence for new states; default, 0.9999; swept, [0.9999, 0.99, 0.9, 0.8]
- ξ_1 (scaling of $\bar{\delta}$ updates): sets how strongly each negative RPE contributes to $\bar{\delta}$; higher values increase the impact of each negative RPE, promoting state splitting; default, 1.5; swept, [1.5, 1, 0.5]
- ξ_{DB} (attention squashing parameter): controls the steepness of the tanh transformation linking $\bar{\delta}$ to cue attention; lower values make cue-attention changes more sensitive to $\bar{\delta}$, increasing splitting; default, 1; swept, [1, 0.5, 0.25]

Gershman (2017) et al. (37)

- α (concentration parameter): governs the prior probability of assigning the current trial to a new latent cause; higher values increase the likelihood of inferring new causes; default, 0.1; swept, [0.01, 0.1, 0.5, 1, 1.5, 2]
- κ (temporal kernel exponent): determines how strongly recent trials dominate the prior; higher values make the model favor the most recent latent cause, reducing splitting; default, 1; swept, [0.1, 0.5, 1, 1.5, 2, 5, 10]
- σ_x^2 (cue variance under each latent cause): prior uncertainty in cue observations; lower values increase sensitivity to small cue deviation, promoting an inference of new cause; default, 1; swept, [0.001, 0.05, 0.01, 0.1, 0.2, 0.5, 1]
- σ_r^2 (reward variance under each latent cause): prior uncertainty in reward outcomes; lower values make reward prediction errors more diagnostic, increasing state splitting; default, 1; swept, [0.001, 0.05, 0.01, 0.1, 0.2, 0.5, 1]

Cochran (2019) et al. (36)

- γ (latent state transition probability): controls the chance of switching states between consecutive trials; higher values increase trial-to-trial state transitions and new state inference; default, 0.05; swept, [10^{-10}, 0.001, 0.05, 0.01, 0.1, 0.2, 0.5, 1]
- υ (threshold for creating new state): determines how much mismatch is required to trigger a new state; lower values result in more frequent creation of new states; default, 0.2; swept, [0.1, 0.2, 0.5, 10, 1000]
- δ (unexpected-uncertainty threshold): controls how large prediction errors must be to start accumulating evidence for a state change; lower values make the model highly sensitive to deviations, increasing splitting; default, 0.6; swept, [0.05, 0.3, 0.6, 1.5, 5]
- σ_0 (initial expected uncertainty): initial standard deviation of prediction error; lower values make early prediction errors more informative, facilitating new-state inference; default, 0.5; swept, [0.01, 0.1, 0.5, 1, 5]

The best-fitting parameter set was defined as the sweep combination that produced 1) the largest difference between groups in the probability of inferring the acquisition context, $P(\text{contextAcq})$, during CS-A trials in the CS-A extinction phase, and 2) the largest $P(\text{contextAcq})$ during CS-A trials in reacquisition for Group1. The acquisition context was defined by ranking all inferred contexts based on their total inferred probability during Acquisition and selecting the smallest set of top-ranked contexts whose cumulative inferred probability exceeded 90% of acquisition trials. Learning speed was quantified using the same procedure as for ANCCR, but substituting SRC with the value estimates generated by each model. For the optogenetic dopamine inhibition experiment, we simulated only Redish et al. model, as it was the only latent cause model that captured the behavioral results of Group1 and Group2. Optogenetic inhibition was modeled as a negative RPE with varying magnitudes $[-10, -5, -2, -1.5, -1, -0.5, -0.3]$, and results with RPE $= -1.5$ (same magnitude as used in ANCCR simulation) were used in the main figure ($1 =$ modeled positive RPE evoked by an unpredicted reward).

For the Cochran et al. (36) model, we additionally simulated a 100% cue–reward probability version of the task to test whether the model could capture the behavioral patterns of Groups 1 and 2 under deterministic conditions. We focused on simulating the Cochran et al.

model because it is the most recent latent-cause framework and is conceptually aligned with the Gershman et al. (37) model (both rely on Bayesian inference over latent causes). We did not re-simulate the Redish et al. (39) model for this condition, as we had already confirmed that it successfully captures the Group1 and Group2 results under the 50% probabilistic condition (although it does not capture either the Group3 pattern or Group2 with optogenetic manipulation). The same parameter sweep combinations were used as in the probabilistic (50%) condition. For **fig. S7D,E**, we selected the representative parameter set that captured the differential behavior between the 50% and 100% conditions (but see **fig. S7A-C** for full sweep results): the sweep in which no new state was inferred for CS-A trials in either Group1 or Group2 under the 50% condition, whereas a new state was inferred for CS-A trials in both groups under the 100% condition.