Aversive learning retroactively prioritizes neutral episodic memories structured by prior
knowledge of predictive sequences

Blazej M. Baczkowski[1,2,3,4] & co-authors

[1] Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences,
Leipzig, Germany

[2] IMPRS NeuroCom, Leipzig, Germany

[3] Institute of Psychology, University of Leipzig, Leipzig, Germany

[4] Hector Research Institute of Education Sciences and Psychology, University of Tübingen,
Tübingen, Germany

Author Note

Abstract

Individuals can infer danger by linking memories of neutral experiences with those of later threatening encounters. Yet it remains unclear how seemingly inconsequential neutral events – encoded in safe contexts and therefore prone to forgetting – are nevertheless preserved in memory well enough to support threat predictions. Here, we tested whether Pavlovian threat conditioning retroactively prioritizes long-term episodic memory for previously encoded neutral events by leveraging prior knowledge of their predictive relations. Young adults first learned two deterministic temporal sequences of six semantic categories (A->B->C and D->E->F), and then incidentally encoded trial-unique exemplars from each category while viewing these sequences. In a subsequent conditioning phase, the final categories (C vs. F) became differentially predictive of mild electric shocks. After 24 hours, recognition memory for categories encoded before conditioning varied by sequence and position: when a sequence was inferred to end with a shock-predictive category, memory for its starting category (A) was prioritized over the following category (B) – an effect absent in the control categories (D and E). These results suggest that, when neutral events are scaffolded by prior sequence knowledge, aversive learning can retroactively reshape their episodic memories despite a temporal gap to support early threat anticipation.

Keywords: episodic memory | Pavlovian conditioning | sequence learning | fear | inference

Word count: intro: <1200; methods: <1600; results: <1600; discussion: <1300

Aversive learning retroactively prioritizes neutral episodic memories structured by prior knowledge of predictive sequences

Introduction

Dangerous situations often leave enduring memories that are critical for recognizing threats in the future. Yet detecting threats entails exposure to potential harm – a risk that can be mitigated with timely avoidance enabled by the recognition of early warning signals. These signals might be neutral events that precede threatening encounters but initially appear inconsequential. For instance, consider a rustling in the bushes that may seem trivial at first and hence be quickly forgotten. But if moments later an aggressive dog jumps out and then attacks, the sound retrospectively becomes significant for future threat prediction. How are neutral experiences retroactively strengthened in memory when their relevance for guiding adaptive behavior is revealed only by later aversive learning?

One prominent mechanism for learning about threats is Pavlovian conditioning, where a neutral stimulus comes to signal potential harm through association with an aversive outcome (Rescorla, 1988; Rescorla & Holland, 1982). This learning supports rapid defensive reactions, such as freezing to a tone paired with electric shock (LeDoux, 2000; Mobbs, Headley, Ding, & Dayan, 2020). However, conditioning is limited when familiar or similar cues are absent and danger must instead be inferred – potentially by drawing on associations acquired under safe conditions (Baczkowski, Haaker, & Schwabe, 2023). In the sensory preconditioning paradigm (Brogden, 1939), individuals first learn that cue A predicts cue B, and later that B predicts shock. At test, A alone elicits defensive responses, despite never being paired with the aversive outcome (Wong, Westbrook, & Holmes, 2019). This demonstrates that danger can be inferred from associations acquired across separate safe and aversive episodes. Yet whether preconditioned cues are preserved strongly enough in memory to later serve as reliable warning signals may depend on processes beyond associative learning.

From an episodic memory perspective, information deemed useful for guiding future behavior is prioritized for long-term storage, as proposed by the adaptive memory framework

(Nairne, Thompson, & Pandeirada, 2007; Shohamy & Adcock, 2010). Conditioning illustrates this principle: events directly linked to aversive outcomes acquire emotional salience, rendering their survival relevance immediately apparent. Consequently, they are well remembered through mechanisms of prediction error learning (Kalbe & Schwabe, 2020, 2021b) and arousal-related modulation of amygdala–hippocampus circuits (McGaugh, 2018; Roozendaal & McGaugh, 2011). By comparison, in preconditioning, neutral events are separated from aversive outcomes by temporal gaps, leaving their threat relevance initially ambiguous and therefore rendering them more susceptible to forgetting. Because these cues lack salience at encoding, their long-term retention cannot rely on encoding strength alone. Instead, their mnemonic status appears to depend on post-encoding mechanisms that stabilize initially fragile memories during consolidation, once later aversive learning retroactively signals their biological significance (Clewett, Dunsmoor, Bachman, Phelps, & Davachi, 2022; Holmes, Raipuria, Qureshi, Killcross, & Westbrook, 2018). Processes such as memory reactivation (Hermans et al., 2017; Staresina, Alink, Kriegeskorte, & Henson, 2013; Tambini & Davachi, 2013), offline replay (Lewis & Durrant, 2011; Mattar & Daw, 2018), or schema-based integration – in which prior knowledge scaffolds new learning (Ghosh & Gilboa, 2014; Tompary & Davachi, 2024; Tse et al., 2007) – could strengthen memories of events that were initially deemed insignificant. Collectively, these mechanisms suggest that neutral episodic memories encoded outside an aversive context can nonetheless be prioritized for long-term storage once their importance for future threat prediction becomes apparent in hindsight.

Here, we examined whether pre-existing knowledge about temporal event sequences can support retroactive memory prioritization. Prior knowledge of predictive relations – for example, that A predicts B, which in turn predicts C – can serve as a relational schema, enabling individuals to recognize distinct events as part of a common underlying pattern despite being encoded in separate episodes. To illustrate, imagine someone visiting a neighbor: in the backyard they notice a kennel (A), then hear barking (B), and finally see a dog (C). Next, during a separate encounter inside the house, the same dog becomes aggressive and attacks (C -> outcome). While the attack itself is likely to be remembered,
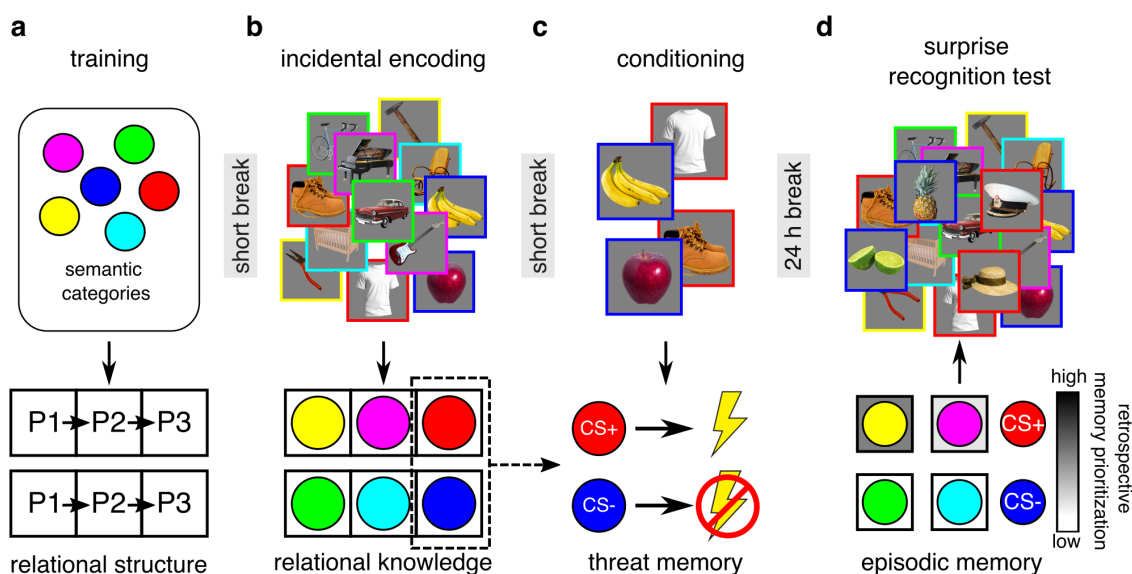
Figure 1. Study rationale and design. We investigated whether neutral information is selectively prioritized in episodic memory when its relevance for predicting future threats is retroactively inferred from pre-existing knowledge of temporal event sequences. **a)** Participants learned two temporal sequences of semantic categories (three per sequence), where category order predicted the image exemplars. **b)** Next, participants incidentally encoded trial-unique images embedded in the learned sequences while making semantic judgments. **c)** To assign aversive value to one sequence, exemplars from the final category of each sequence were repeated during Pavlovian threat conditioning: one category predicted a mild electric shock (CS+), while the other did not (CS−). **d)** Long-term memory was assessed 24 hours later with a surprise old–new recognition task including previously seen exemplars from all six categories and category-matched novel foils. We hypothesized that memory for preconditioned categories in the CS+ sequence (positions A and B) would differ from memory for categories in the CS− sequence (positions D and E), with prioritization selectively modulated by sequence position in the CS+ sequence. Colored borders are for illustrative purposes. Images are reproduced under Pixabay CC0 license.

the kennel may also become memorable – not for its immediate importance, but because it fits a familiar predictive pattern: kennels house dogs, barking signals their presence (A -> B -> C). In this way, once C becomes threatening, A or B can acquire relevance because their predictive relation to C is embedded in a pre-existing schema.

Guided by this reasoning, we asked whether aversive learning retroactively modulates neutral episodic memories encoded under safe conditions, whose relevance for predicting threats is inferred from prior knowledge of temporal sequences. To this end, we adapted the sensory preconditioning paradigm with episodic learning (Figure 1). We established predictive relations at the level of semantic categories (e.g., furniture -> tools) rather than individual items (e.g., chair -> hammer). This category-level design enabled generalization across exemplars (Hsieh, Gruber, Jenkins, & Ranganath, 2014; Sherman & Turk-Browne, 2020) and, by requiring participants to encode many trial-unique exemplars, placed greater demands on episodic memory. Supporting this rationale, prior work shows that Pavlovian associations formed at the category level enhance recognition memory for exemplars of a

shock-predictive category irrespective of direct reinforcement, suggesting within-category mnemonic generalization (Dunsmoor, Martin, & LaBar, 2012; Dunsmoor, Murty, Davachi, & Phelps, 2015). Hence, our design ensured that any retroactive memory effects would reflect category-level rather than item-level learning.

Our pre-registered study comprised three tasks presented in separate phases (Figure 2). In the preconditioning phase, participants first learned two sequences with fixed temporal order (A->B->C and D->E->F), each comprising three semantic categories, while making semantic judgments. Once the sequence structure was learned, participants incidentally encoded non-repeating exemplars from each category while viewing them in these sequences. In the conditioning phase, exemplars of the third-position categories were repeated: one category (C) predicted mild electric shock (CS+), while the other (F) was never reinforced (CS−). This manipulation differentiated the sequences by relevance to aversive outcome and created overlap across phases, which is known to facilitate retrospective inference (Kurth-Nelson, Barnes, Sejdinovic, Dolan, & Dayan, 2015; Wimmer & Shohamy, 2012). In the final phase, after approximately 24 hours to allow for memory stabilization during consolidation (Dudai, Karni, & Born, 2015), we assessed recognition for exemplars of all six categories, along with unseen category-matched foils.

We hypothesized that retroactive prioritization would manifest selectively in recognition memory for preconditioned categories: those in the CS+ sequence (A and B), which gain relevance through later aversive learning, should differ as a function of sequence position, whereas those in the CS− sequence (D and E) should remain unaffected. Critically, our three-step sequence design enabled us to dissociate two properties of the preconditioned categories A and B: temporal proximity and predictive utility. Although B is temporally closer to C, A initiates the predictive chain – A leads to B, which leads to C – rendering B redundant once A is known. Episodic memory reflecting temporal proximity would yield better recognition for B, whereas memory reflecting predictive utility would prioritize A over B. Collectively, our design tests whether aversive learning can retroactively prioritize neutral episodic memories encoded outside an aversive context – a process that, we propose,

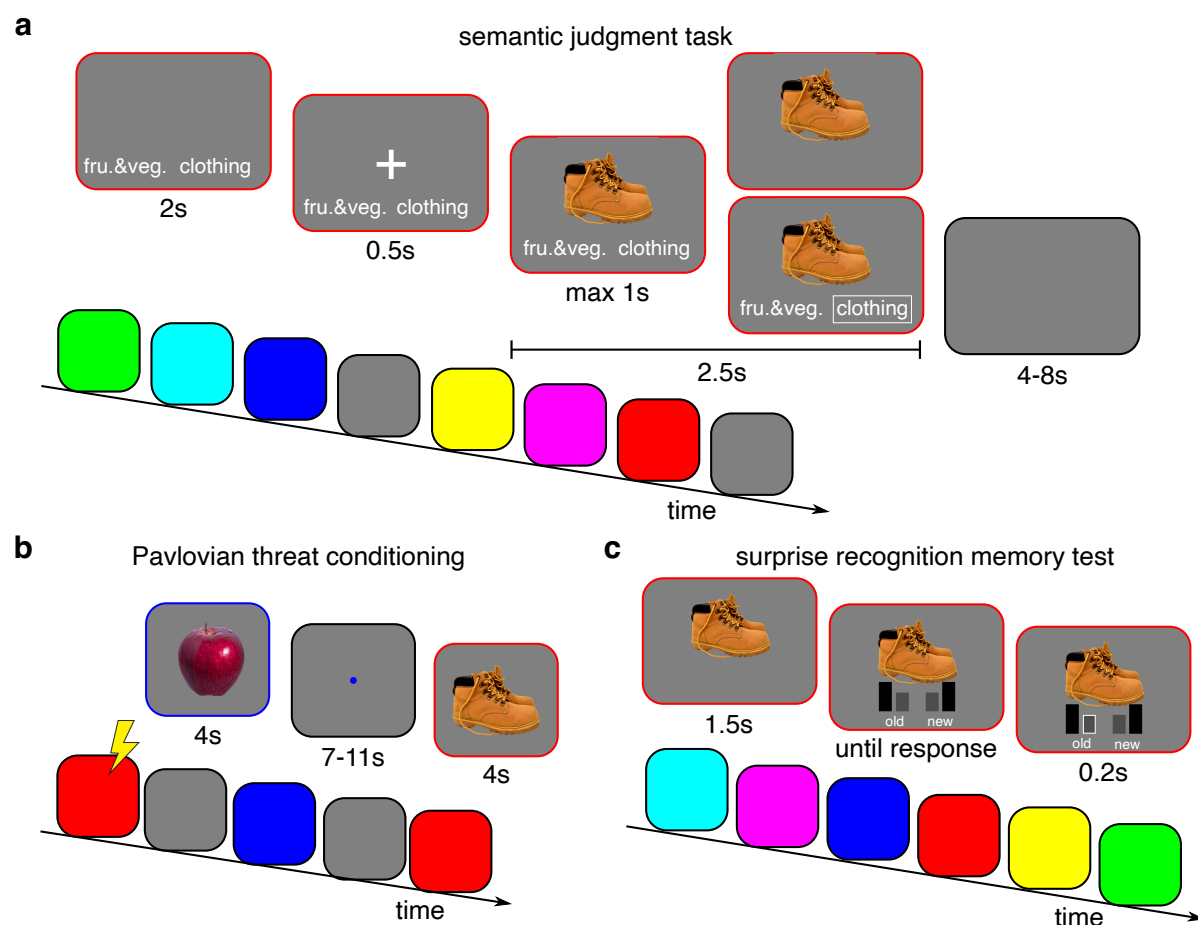depends on prior relational knowledge linking earlier neutral events to later threat-relevant outcomes.

Figure 2. Behavioral tasks. **a)** Semantic judgment task: participants incidentally encoded images of category exemplars while making semantic decisions. Trials were grouped into triplets (three categories per sequence) with variable inter-trial intervals. Each image was shown for 2.5 s, and participants had up to 1 s to respond. **b)** Pavlovian threat conditioning: exemplars from one category predicted a mild electric shock (CS+) under partial reinforcement, while exemplars from another category never predicted shock (CS−). **c)** Surprise recognition test: 24 hours later, participants judged previously seen exemplars and unseen category-matched foils as "old" or "new," providing confidence ratings ("definitely," "maybe"). Colored borders are for illustrative purposes. Images are reproduced under Pixabay CC0 license.

Method

The hypotheses, methods, and analyses were preregistered prior to data collection at [LINK]. Minor deviations from the preregistration protocol are reported in supplemental table S1.

Participants

The study was approved by the local ethics committee (University of Leipzig, Germany) and conducted in accordance with the Declaration of Helsinki. We recruited

healthy, right-handed volunteers aged 18 to 35 from the general population of Leipzig, Germany. Forty-five volunteers gave informed consent. Four participants did not complete the study, and seven did not meet our preregistered inclusion criteria (see the supplemental material). Thus, the final sample consisted of 34 participants (18 female, ages 19-35, mode=24), in line with our preregistered sampling plan. No *a priori* power analysis was conducted; instead, the sample size was determined in line with previous studies demonstrating similar effects at the time of study design (Dunsmoor et al., 2015; Patil, Murty, Dunsmoor, Phelps, & Davachi, 2017; Starita, Kroes, Davachi, Phelps, & Dunsmoor, 2019). *A post-hoc* sensitivity analysis (1,000 simulations with MLE-derived nuisance parameters) indicated our design – a 2 by 2 interaction among preconditioned categories – had >80% power at $\alpha=0.05$ to detect a 6-percentage-point (or 0.5-unit) difference-in-differences in proportion correct (or d-prime) in the recognition memory task.

Stimuli

Memoranda for incidental encoding and the recognition memory test were color photographs of neutral everyday objects (obtained through Google search) on a grey background (RGB: [128/255, 128/255, 128/255]) with equalized luminance. Each image depicted a unique basic-level object with a unique name belonging to one of six semantic categories: "tools", "fruits & vegetables", "furniture", "vehicles", "clothes", and "music instruments". The initial pool contained 62 images per category (372 in total). After a pilot study (see Supplemental Experiment 1), we excluded images that were difficult to perceptually distinguish, atypical for their category, or associated with low or high 15-minute recognition memory scores. The final sample consisted of 50 images per category (300 in total).

Behavioral tasks

Semantic judgment task.   To form knowledge of predictive relations, six semantic categories were split into two temporal sequences of three serial positions, predicting category exemplars in a semantic judgment task (cf., Hsieh et al., 2014). Category allocation to serial positions was pseudo-randomized across participants (for details, see the

supplemental materials). Participants made semantic decisions (e.g., "Is it a furniture or a tool?") following a 2-s presentation of category names, a 0.5-s fixation, and a 2.5-s image. Responses were made during image presentation as long as two response options were on the screen (<1 s). Trials were grouped into triplets with 0.5-s spacing and sequences were separated by 6 ± 2 s with a blue fixation dot. Triplet order was pseudo-randomized, and each triplet followed an order of categories given their serial positions in a sequence. For example, participants saw a sequence of three trials such that a musical instrument was followed by a tool, followed by a vegetable.

The task included a training phase and an incidental encoding phase. First, participants were trained to anticipate the category membership of images, which were repeatedly presented in two sequences, by explicitly focusing on their underlying order. In each block, they saw one new exemplar of each category that was presented four times per sequence (8 triplets in total). To ensure that participants learned the order of categories, they were repeatedly exposed to a "drag-and-drop" test after each of three blocks. In the "drag-and-drop" test, all six category names were simultaneously presented on a screen and participants had to reconstruct the order of categories in which their images were shown before. Participants received feedback on their performance, and in case of an incorrect response, they were asked to reconstruct the sequence again. Having ensured that participants were familiar with the task and learned the order of categories in each sequence, they incidentally encoded 150 new trial-unique (i.e., non-repeating) exemplars (25 per category) in one block (50 triplets in total). These basic-level category exemplars were presented in the absence of any explicit motivation or instruction to remember them as well as without the risk of receiving a mild electric shock.

Category-based Pavlovian threat conditioning.   To form Pavlovian threat memory, participants underwent a category-based threat conditioning task. Two categories ("fruits & vegetables" and "clothes") from the third position of each sequence were used such that exemplars of one category (CS+) were predictive of the unconditioned stimulus (US) – a mild electric shock applied to fingers – while exemplars of the category (CS-) were never predictive of the US (for details, see the supplemental material). Exemplars of the two

categories were already presented during the incidental encoding in the semantic judgment task, and here they were repeated such that there were 25 exemplars of the CS- and 25 exemplars of the CS+, 13 of which co-terminated with the US. The category exemplars were repeated to form overlapping representations across the pre-conditioning and conditioning phases that could presumably facilitate their interaction. The category to condition (CS+/CS-) allocation was counter-balanced across participants. Pictures were presented for 4.0 s with a variable inter-trial interval of 7-11 s (the average ITI was set to 9 s). The order of trials was pseudo-randomized so that no more than 3 trials of the same type could occur in a row. To facilitate aversive learning, the first trial was always the CS+ that was reinforced with the US. Participants were instructed that there was a relationship between the categories and the shocks but were not explicitly instructed the contingency between conditioned category and the US, and had to learn the category-level association through experience. On each trial, participants indicated whether they expected a shock (yes/no) but were told that their responses would not affect the outcome, to prevent them from attributing it to their choice or reaction time.

Surprise recognition memory test. Recognition memory was tested with a surprise old-new recognition test approximately 24 hours after encoding. The test included 300 images – for each category there were 25 old images and 25 new category-matched foils that were never shown before to control for false alarm rate. The test did not include images presented in the training phase of the semantic judgment task on the previous day, because these images were presented multiple times. Images were randomly assigned as old or new across participants. Each image appeared for 1.5 s, followed by response options: "definitely old", "maybe old", "maybe new", and "definitely new". The trial order was pseudo-randomized to prevent long strings of old or new images and avoid repeating sequences from the semantic judgment task. The task was self-paced.

Skin conductance acquisition and response estimation

Skin conductance was recorded using BIOPAC MP35 (Biopac Systems Inc., CA, USA) with Ag/AgCl electrodes at a 500 Hz sampling rate. Continuous raw time series were

pre-processed with a custom script (see the supplemental materials). Skin conductance responses (SCR) were scored if the trough-to-peak deflection occurred within 0.5-4.0 s after CS onset and lasted 0.5-5.0 s (or until the US onset for CS+ trials). Trials not meeting these criteria were scored as zero. Scoring was semi-automatic using Autonomate software (Green, Kragel, Fecteau, & LaBar, 2014), with visual inspection and manual adjustments as needed. Raw SCR values were scaled by the maximum SCR evoked by the shock, and the SCR magnitude was calculated as the mean value across condition-specific trials. The first trial (CS+ with US) was excluded due to orienting responses. SCR values were square-root transformed for group analysis.

Procedure

Two experimental sessions were conducted on consecutive days in the same test room and at similar times. The detailed procedure is described in the supplemental materials. The first session included a pre-conditioning phase (with no US risk) followed by a conditioning phase. Participants began with the semantic judgment task, then had electrodes attached for shock administration, with intensity calibrated to "maximally uncomfortable without being painful." The threat conditioning task was then conducted while measuring skin conductance. Afterward, electrodes were removed, and debriefing questions were asked.

In the second session, participants started with a recognition memory test. They next completed a "drag-and-drop" task to order the six category names into two sequences in the way they remembered experiencing them on the previous day. Finally, participants filled out four self-report questionnaires regarding their personality (trait anxiety and intolerance for uncertainty), childhood traumatic experiences, and social relations (see the supplemental materials). The session ended with debriefing and monetary compensation.

Statistical analyses

Data pre-processing, plotting, and analysis were conducted in Matlab R2016a, GNU Octave 6.4.0, and R 3.3.3 (R Core Team, 2017). To analyze the data, we used (Generalized) Linear Mixed Models (G/LMM) implemented in the R package lme4 (Bates, Mächler,

Bolker, & Walker, 2015). When specifying statistical models, we aimed to include a full random structure; when a model exhibited convergence issues, singular fit, or violated assumptions, we reduced the random structure by successively removing its individual components and/or their correlations. To test each hypothesis, we specified a full model and a null model (i.e., a model reduced by excluding the test predictor), corresponding to the alternative and null hypotheses, respectively (for details, see the supplemental material). The models were compared using likelihood-ratio tests with significance level set at $p = .05$. Model assumptions were checked by inspecting residual normality and homoscedasticity. Sample means and parameter estimates are reported with 95% confidence intervals (CIs) obtained through bootstrapping (n = 5000 and n = 1000, respectively), in the format: mean [lower bound, upper bound]. For gaussian models, parameter estimates are reported on the raw (unstandardized) scale, and for binomial models, estimates are reported on the log-odds scale. We additionally report coefficients of determination (pseudo-R^2) for (G)LMMs as indices of model fit (Nakagawa, Johnson, & Schielzeth, 2017; Nakagawa & Schielzeth, 2012), calculated with the R package MuMIn (Bartoń, 2023). Following convention, we refer to the fixed-effects–only model as the marginal coefficient and the full model (fixed + random effects) as the conditional coefficient. Details of specific analyses are reported in the supplemental material, and the respective random-effect structures are provided in supplemental tables for each model.

Results

Prior knowledge of the predictive sequences facilitated semantic judgments

After participants learned the temporal order of semantic categories during the training phase of the semantic judgment task, we tested whether they used this knowledge while incidentally encoding trial-unique exemplars from all six categories. To do so, we examined the effect of trial position nested within a sequence on response accuracy and its reaction time (RT) using binomial and gaussian linear mixed models, respectively (Figure 3). We observed that the proportion of correct responses to valid trials, i.e., trials in which a response was given within a pre-specified time limit, differed across positions regardless of

the sequence ($\chi^2(4) = 45.23$, $p < .001$, marginal-$R^2 = 0.12$, conditional-$R^2 = 0.24$, $\phi = 0.89$). In both sequences, participants showed a higher probability of correct responses in the second position compared to the first ($\beta_{s1p2-p1} = 1.42$ [0.68, 2.17], $\beta_{s2p2-p1} = 0.72$ [0.22, 1.22]). Accuracy also increased in the third compared to the second position in one sequence ($\beta_{s1p3-p2} = -0.23$ [-1.06, 0.61], $\beta_{s2p3-p2} = 1.23$ [0.42, 2.03]). Indeed, an additional GLMM revealed a significant interaction ($\chi^2(2) = 6.58$, $p = .037$, marginal-$R^2 = 0.12$, conditional-$R^2 = 0.23$, $\phi = 0.88$) between the two sequences in the accuracy between the third and second positions ($\beta_{p2-p1} = -0.74$ [-1.61, 0.13], $\beta_{p3-p2} = 1.46$ [0.30, 2.62]). Yet, when omissions, treated as incorrect responses, were included in the total number of trials, the interaction effect was absent ($\chi^2(2) = 1.68$, $p = .431$, marginal-$R^2 = 0.12$, conditional-$R^2 = 0.20$, $\phi = 0.74$, $\beta_{p2-p1} = -0.16$ [-0.61, 0.30], $\beta_{p3-p2} = 0.39$ [-0.20, 0.97]). Instead, the effect of position nested within a sequence was revealed ($\chi^2(4) = 99.34$, $p < .001$, marginal-$R^2 = 0.00$, conditional-$R^2 = 0.15$, $\phi = 0.74$). Specifically, the accuracy successively increased across positions in the first ($\beta_{s1p2-p1} = 0.89$ [0.56, 1.23], $\beta_{s1p3-p2} = 0.69$ [0.28, 1.10]) and second sequence ($\beta_{s2p2-p1} = 0.72$ [0.31, 1.13], $\beta_{s2p3-p2} = 1.08$, [0.67, 1.49]). The analysis of RT to correct trials revealed successive increase in speed across the positions in both sequences ($\chi^2(4) = 154.09$, $p < .001$, marginal-$R^2 = 0.45$, conditional-$R^2 = 0.94$, $\sigma = 0.03$) without any interaction effect ($\chi^2(2) = 3.72$, $p = .155$, marginal-$R^2 = 0.45$, conditional-$R^2 = 0.96$, $\sigma = 0.02$). Specifically, the RT decreased largely between the first and second position ($\beta_{s1p2-p1} = -0.14$ [-0.16, -0.11], $\beta_{s2p2-p1} = -0.13$ [-0.16, -0.11]) and also between the second and third position ($\beta_{s1p3-p2} = -0.03$ [-0.04, -0.02], $\beta_{s2p3-p2} = -0.02$ [-0.03, -0.01]).

Overall the semantic judgment task was easy for participants (~98% accuracy in valid trials). When participants could use their knowledge of the systematic order among semantic categories, their performance improved in the second and third positions, regardless of the sequence. In contrast, our pilot experiments (see Supplemental Experiments 2 and 3) revealed that when a systematic order among categories was absent, participants exhibited similar performance across sequence positions (or faster responses in the first position). Together, these findings suggest that the acquisition of predictive
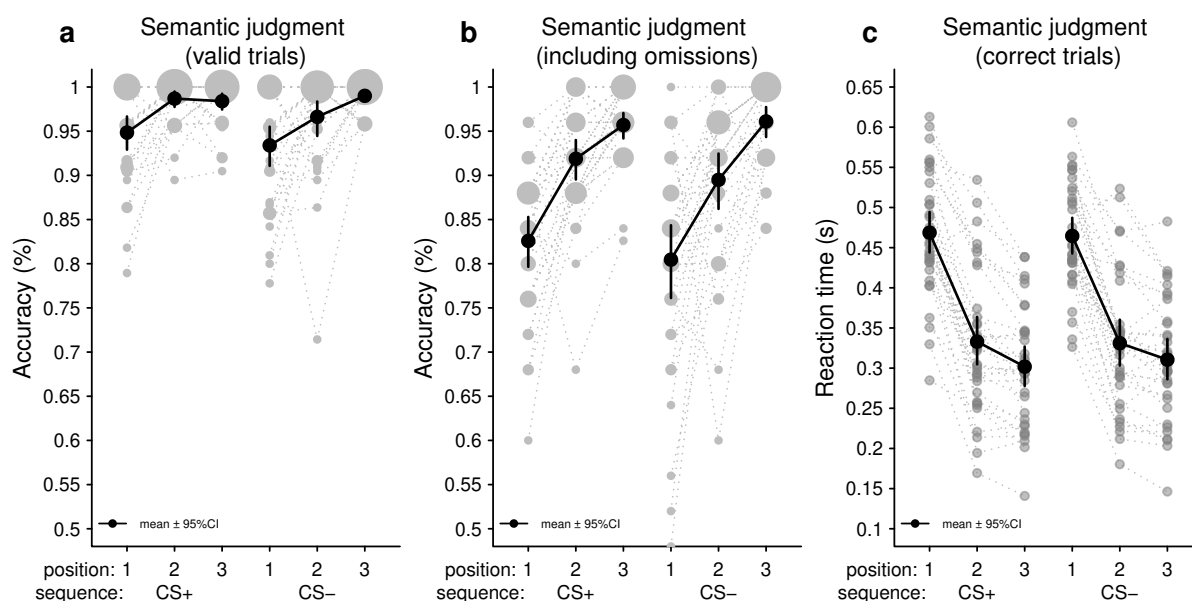
Figure 3. Performance in the semantic judgment task when a systematic order among semantic categories was present in each sequence. **a)** Proportion of correct responses to the total number of valid trials increased from the first to the second position, and the differences between the third and second position depended on a sequence ($\chi^2(2) = 6.58$, $p = .037$, marginal-$R^2 = 0.12$, conditional-$R^2 = 0.23$ $\phi = 0.88$). **b)** However, the interaction effect of sequence disappeared when their proportion was calculated to the total number of trials including omissions. Here, the proportion of correct responses increased across positions in both sequences ($\chi^2(4) = 99.34$, $p < .001$, marginal-$R^2 = 0.00$, conditional-$R^2 = 0.15$, $\phi = 0.74$). **c)** Average reaction time to correct trials decreased along the positions regardless of the sequence ($\chi^2(4) = 154.09$, $p < .001$, marginal-$R^2 = 0.45$, conditional-$R^2 = 0.94$, $\sigma = 0.03$). Size of the circles is proportional to the number of participants with the same data point.

relationships among categories enhanced semantic judgments, and category exemplars were encoded similarly across both sequences.

The final categories became differentially predictive of aversive outcome through conditioning

Having established that participants had learned the predictive relations among the six categories and encoded their trial-unique exemplars, the two categories at the end positions of each sequence were used for category-based Pavlovian threat conditioning. To assess Pavlovian threat memory acquisition, we compared explicit shock expectancy and physiological measures of arousal, i.e., anticipatory skin conductance response (SCR). As expected, participants acquired differential conditioned threat responses (Figure 4). GLMM with a *logit* link function and binomial error structure fitted to a number of responses indicating shock expectancy revealed an expected effect of condition ($\chi^2(1) = 80.04$, $p < .001$, marginal-$R^2 = 0.88$, conditional-$R^2 = 0.98$, $\phi = 0.30$). Shock expectancy was increased in the CS+ condition compared with the CS- condition ($\beta_{gm} = -1.39$ [-1.74, -1.04], $\beta_{slope} = 5.45$ [4.67, 6.24]). These results were corroborated by a corresponding

binomial GLMM fitted to individual binary responses ($\chi^2(1) = 63.20$, $p < .001$, marginal-R^2 = 0.65, conditional-R^2 = 0.77, $\phi = 0.75$). Similarly, LMM fitted to SCR revealed an expected effect of condition ($\chi^2(1) = 25.63$, $p < .001$, marginal-R^2 = 0.16, conditional-R^2 = 0.73, $\sigma = 0.06$). SCR was increased in the CS+ condition compared with the CS– condition ($\beta_{gm} = 0.32$ [0.29, 0.36], $\beta_{slope} = 0.09$ [0.06, 0.12]). Taken together, participants exhibited evidence for the acquisition of Pavlovian threat memory using a category-level association with mild electric shock. This way, the two categories located at the terminating positions of each sequence became differentially predictive of aversive outcome.
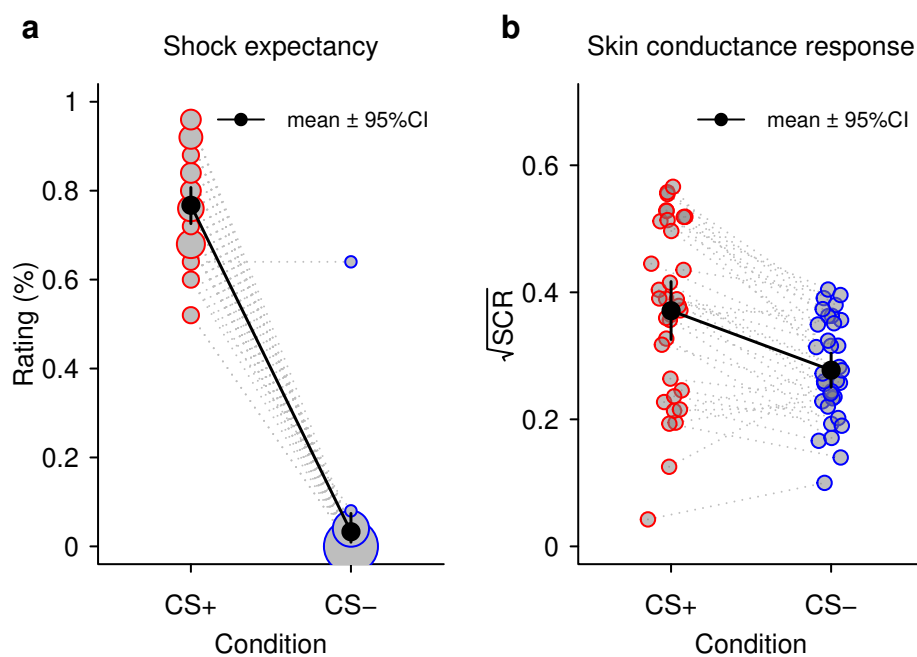

Figure 4. Shock expectancy and skin conductance response (SCR) during category-based Pavlovian threat conditioning. **a)** Proportion of responses indicating shock expectancy revealed an expected effect of condition ($\chi^2(1) = 63.20$, $p < .001$, marginal-R^2 = 0.65, conditional-R^2 = 0.77, $\phi = 0.75$). Size of the circles is proportional to the number of participants with the same data point. Participant with the same proportion of responses in both conditions exhibited increased SCR to the CS+ items compared to the CS- items. **b)** Similarly, magnitude of SCR differed between the conditions ($\chi^2(1) = 25.63$, $p < .001$, marginal-R^2 = 0.16, conditional-R^2 = 0.73, $\sigma = 0.06$).

Threat conditioning selectively altered the recognition memory for preconditioned categories

Finally, approximately 24 hours after encoding, we assessed recognition memory for exemplars from all six categories and their unseen category-matched foils using an old–new surprise test. In line with prior work (de Voogd, Fernández, & Hermans, 2016a; Dunsmoor et al., 2015; Kalbe & Schwabe, 2020), old–new judgments were collapsed across two confidence levels for analysis to ensure an equal number of responses across participants (see

the supplemental material for analyses including only high-confidence responses). We analyzed the old–new judgments using the proportion of correct responses (correctly identifying old items as old and new items as new), which corresponds linearly to the commonly used corrected recognition score (P[hit] − P[false alarm]). To assess the robustness of this analysis, we also computed the d-prime index from signal detection theory (Figure 5). To ease the analysis, we split the data into the pre-conditioned categories (allocated to positions 1 and 2) and conditioned categories (allocated to position 3; for the analysis with full dataset, see the supplemental material).

The comparison between the CS+ and CS- conditions revealed no difference (proportion correct GLMM: $\chi^2(1) = 0.02$, $p = .890$, marginal-$R^2 = 0.29$, conditional-$R^2 = 0.61$, $\phi = 0.81$; d-prime LMM: $\chi^2(1) = 0.10$, $p = .752$, marginal-$R^2 = 0.18$, conditional-$R^2 = 0.38$, $\sigma = 0.44$). Recognition memory test for items of the CS+ category was almost the same as for items of the CS- category (proportion correct: $\beta_{slope} = 0.01$ [-0.17, 0.19]; d-prime: $\beta_{slope} = 0.03$ [-0.18, 0.25]). Although these results may appear contradictory to previous studies reporting the effects of memory prioritization through category-based threat conditioning (de Voogd, Fernández, & Hermans, 2016b; Dunsmoor et al., 2015), they are not directly comparable, as those studies used trial-unique items throughout. In contrast, we used the same exemplars of the CS+ and CS- categories in both the pre-conditioning (semantic judgment task) and conditioning phase, aiming to facilitate the interaction between the two phases through shared elements. Since we anticipated that item repetition might lead to increased (ceiling) recognition memory and potentially overshadow the memory prioritization of CS+ over CS- items, we conducted a pilot experiment before the main experiment and replicated the memory prioritization effect with our current category stimulus set (see Supplemental Experiment 3). In the pilot, the overall demand on episodic memory in the pre-conditioning phase was the same as in the current study but CS+ and CS- items were presented only once, i.e., only in the conditioning phase. The pilot successfully replicated the effect of conditioning on memory recognition (proportion correct GLMM: $\chi^2(1) = 8.99$, $p = .003$, marginal-$R^2 = 0.25$, conditional-$R^2 = 0.46$, $\phi = 0.99$; d-prime LMM: $\chi^2(1) = 5.91$, $p = .015$, marginal-$R^2 = 0.17$,

conditional-$R^2 = 0.20$, $\sigma = 0.43$). Recognition memory was increased for CS+ items compared with the CS- items (proportion correct: $\beta_{slope} = 0.32$ [0.10, 0.53]; d-prime: $\beta_{slope} = 0.39$ [0.10, 0.69]). Considering the two studies side by side revealed that the recognition values in the main study (logit $\beta_{gm} = 1.57$ [1.45, 1.69]) were on average higher than the values in our pilot study (logit $\beta_{gm} = 0.98$ [0.86, 1.12]) which suggests a ceiling performance that likely overshadowed the expected memory prioritization of CS+ items over CS- items. In this context, the lack of differences between the conditioned categories does not impede our hypothesis test as this relies on the recognition memory for the pre-conditioned categories only, trials of which were non-repeating.
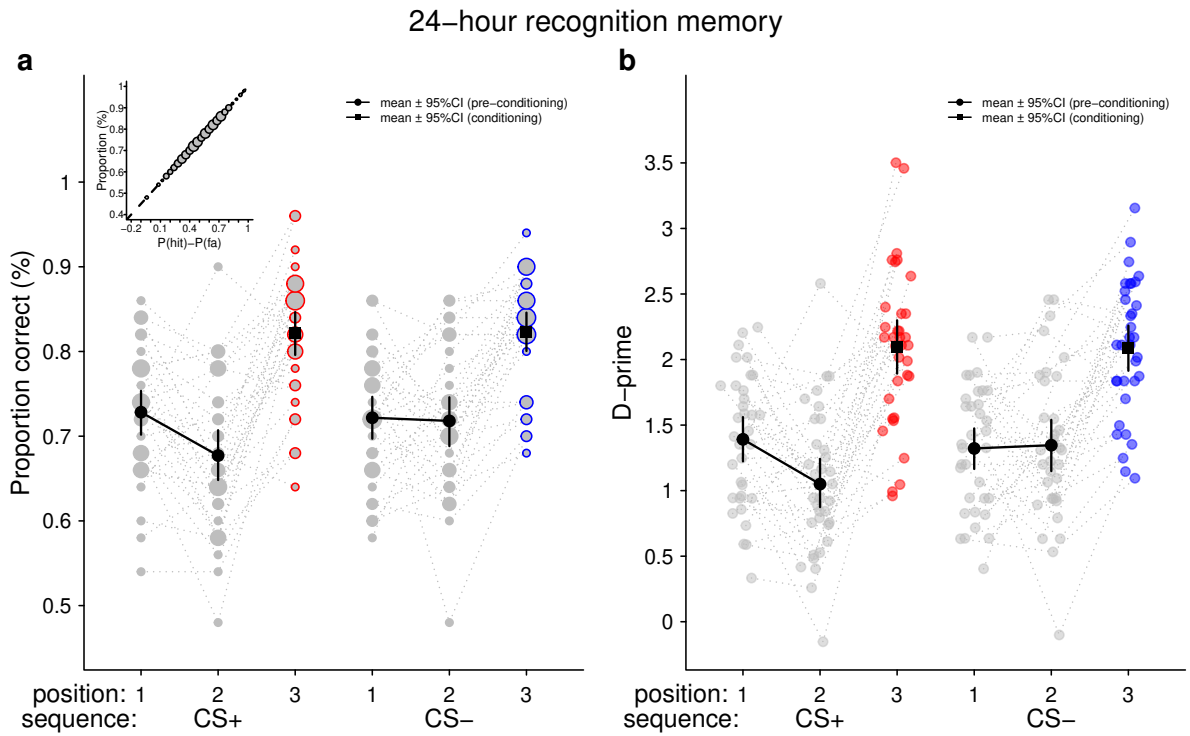

Figure 5. 24-hour recognition memory of the categories incidentally encoded during the semantic judgment task only (pre-conditioning phase) and the categories encoded additionally during the Pavlovian threat conditioning (conditioning phase). **a)** While the probability of correctly recognising items of conditioned categories (CS+ vs. CS-) was similar ($\chi^2(1) = 0.02$, $p = .890$, marginal-$R^2 = 0.29$, conditional-$R^2 = 0.61$, $\phi = 0.81$), it differed in the pre-conditioned categories between the CS+ and CS- sequence as a function of their positions ($\chi^2(1) = 5.20$, $p = .023$, marginal-$R^2 = 0.08$, conditional-$R^2 = 0.53$, $\phi = 0.75$). Size of the circles is proportional to the number of participants with the same data point. Inset plot demonstrates the linear mapping between the scores based on proportion correct vs corrected recognition. **b)** Likewise, results based on d-prime indicated that memory recognition for conditioned categories (CS+ vs. CS-) did not differ ($\chi^2(1) = 0.10$, $p = .752$, marginal-$R^2 = 0.18$, conditional-$R^2 = 0.38$, $\sigma = 0.44$), while we observed interaction effect in the pre-conditioned categories ($\chi^2(1) = 4.91$, $p = .027$, marginal-$R^2 = 0.07$, conditional-$R^2 = 0.20$, $\sigma = 0.48$).

For the pre-conditioned categories, we observed an interaction effect between position and sequence ($\chi^2(1) = 5.20$, $p = .023$, marginal-$R^2 = 0.08$, conditional-$R^2 = 0.53$, $\phi = 0.75$). Inspection of the GLMM estimates revealed that the interaction was mainly driven by

the pronounced decrease in the memory recognition for the items of categories allocated to the second rather than first position and embedded in the CS+ compared with the CS- sequence ($\beta_{gm} = 0.92$ [0.83, 1.00], $\beta_{position} = 0.26$ [0.00, 0.53], $\beta_{sequence} = -0.16$ [-0.42, 0.09], $\beta_{position:sequence} = 0.25$ [0.04, 0.46], which translates into a 5 [1, 10] percentage-point difference-in-differences). The results were corroborated by LMM fitted to d-prime ($\chi^2(1) = 4.91$, $p = .027$, marginal-$R^2 = 0.07$, conditional-$R^2 = 0.20$, $\sigma = 0.48$, $\beta_{gm} = 1.28$ [1.17, 1.38], $\beta_{position} = 0.32$ [-0.01, 0.65], $\beta_{sequence} = -0.23$ [-0.56, 0.09], $\beta_{position:sequence} = 0.37$ [0.04, 0.70]). To mitigate alternative explanations, our pilot experiments showed that the four pre-conditioned categories exhibited similar 24-hour recognition score at baseline (see Supplemental Experiment 1 and 2), and that threat conditioning had no impact on their 24-hour recognition when the systematic order among categories was absent (see Supplemental Experiment 3).

Subsequent exploratory analyses further corroborated the main results (see the supplemental material). These analyses showed that the interaction effect was independent from the performance in the semantic judgment task. We also found no evidence that the interaction effect depended on the physiological arousal (SCR) elicited during conditioning. Further, there was no evidence that the interaction effect related to the individual differences in trait anxiety (STAI-T), intolerance for uncertainty, or childhood traumatic experience.

Taken together, these results suggest that Pavlovian threat conditioning altered the recognition memory for pre-conditioned categories. This effect depended on the position within the temporal sequence if the sequence included the category that later became predictive of shock.

Discussion

Individuals rely on episodic memories of aversive experiences to recognize future threatening encounters (Talmi, 2013; Williams, Ford, & Kensinger, 2022). Such memories often include neutral cues that were encoded in the same context as the aversive event itself due to close spatial or temporal proximity, and can later serve as warning signals (Clewett &

Davachi, 2017; Rouhani, Niv, Frank, & Schwabe, 2023; but see Bisby, Horner, Bush, & Burgess, 2018). For example, neutral items that precede emotionally arousing images (Anderson, Wais, & Gabrieli, 2006; Murayama & Kitagami, 2014; but see Hurlemann et al., 2005; Strange, Hurlemann, & Dolan, 2003) or progressively lead to reward (Braun, Wimmer, & Shohamy, 2018) are often prioritized in long-term episodic memory. Thus, memory for neutral circumstantial events can be enhanced when they share a spatio-temporal context with an aversive experience. However, it is less clear whether such memory enhancements extend to neutral events that occur close in time but outside the immediate emotional context. This question is critical because temporal gaps segment experience into separate episodes, isolating threat from safety to prevent overgeneralization (Dunsmoor et al., 2018; Ezzyat & Davachi, 2021), but may also constrain the capacity of aversive learning to retroactively prioritize neutral events that could support future threat prediction.

The present study addressed this question by adapting the sensory preconditioning paradigm with episodic learning, in which participants first learned predictive relations among three neutral events (semantic categories in our design), followed by aversive learning occurring after a temporal gap. These predictive chains can be viewed as schema-like structures that bridge such boundaries by enabling inferences from prior knowledge (Gilboa & Marlatte, 2017; van Kesteren, Ruiter, Fernández, & Henson, 2012), allowing earlier events to be selectively prioritized when later linked to aversive outcomes. We hypothesized that seemingly inconsequential information at encoding could retrospectively acquire importance for predicting future threats through these inferred predictive relations. Once events are deemed important for survival, they should be selectively prioritized in episodic memory (Nairne et al., 2007; Ritchey, Murty, & Dunsmoor, 2016). We found that conditioning selectively prioritized memory for the first event at the expense of the second event in the deterministic sequence inferred as relevant for aversive outcomes. This asymmetry suggests that experience was first segmented into ordered memory units, which in turn allowed conditioning to act selectively (Laing & Dunsmoor, 2025). Such selectivity nested within a sequence favored early events, as they carry the greatest anticipatory value for the final outcome in highly predictable environments characterized by a chain of successive events

(Klein-Flügge, Wittmann, Shpektor, Jensen, & Rushworth, 2019; Seymour et al., 2004). By contrast, memory for events in the control sequence, irrelevant for aversive outcomes, remained relatively uniform, suggesting that their temporal position carried little consequence for behavior. Taken together, these findings suggests that conditioning can retroactively reshape episodic memory across temporal gaps, and that this modulation is structured by prior knowledge of predictive relations supporting cross-boundary integration (Tompary & Davachi, 2024).

Our findings suggest that events encoded in neutral contexts can later interact with aversive learning despite temporal gaps. This cross-boundary interaction may arise from post-encoding memory reactivation processes shaped by prior knowledge, with predictive schemas providing the scaffolding that may help bridge boundaries and support selective consolidation (Cowan, Chanales, Davachi, & Clewett, 2024; Macias & Persaud, 2024). Reactivation may occur online during conditioning, particularly when neutral experiences and subsequent aversive learning activate shared representations (Shohamy & Wagner, 2008; Wimmer & Shohamy, 2012), enabling their integration into new threat memories in the amygdala, together with hippocampal and prefrontal engagement (Cooper, Hennings, Bibb, Lewis-Peacock, & Dunsmoor, 2024). In line with this, pharmacological interventions in rodents have shown that preconditioning and conditioning memories, though stored in distinct brain regions, can be integrated when prior memory traces are reactivated during conditioning (Wong et al., 2019). Alternatively, reactivation may occur spontaneously offline, during periods of awake rest or sleep (de Voogd et al., 2016a; Hermans et al., 2017), strengthening memory retention (Staresina et al., 2013; Tambini & Davachi, 2013). In particular, offline memory replay offers a compelling route for overcoming temporal gaps. Triggered by salient outcomes, replay can reinstate neutral events initially encoded outside an emotional context (Ambrose, Pfeiffer, & Foster, 2016; Liu, Dolan, Kurth-Nelson, & Behrens, 2019; Singer & Frank, 2009). By linking disjoint experiences, replay builds coherent memory representations structured by prior schemas, which scaffold newly acquired associations (Ghosh & Gilboa, 2014; Tse et al., 2007) and extend knowledge beyond veridical experience but tuned toward anticipated future use (Liu, Mattar, Behrens, Daw, &

Dolan, 2021; Wise, Liu, Chowdhury, & Dolan, 2021). Taken together, these post-encoding mechanisms provide a plausible means by which aversive learning can retroactively prioritize neutral episodic memories.

Limitations

We hope that future research will address some limitations of our study. First, the mechanisms underlying the positional effect in preconditioned categories remain unclear. Although exemplars of pre-conditioned categories were encoded under comparable conditions across sequences, first-position items may have attracted greater attentional resources than second-position items, resulting in stronger encoding. Prior work indicates that memories marked by heightened medial temporal lobe activation at encoding are selectively stabilized during post-encoding consolidation under stress (Ritchey, McCullough, Ranganath, & Yonelinas, 2017). By analogy, first-position items may have been tagged as particularly relevant at encoding and subsequently prioritized during consolidation due to their inferred relevance for predicting aversive outcomes (Sakaki, Fryer, & Mather, 2014). Future research should therefore test whether attentional advantage and its possible interaction with post-encoding processes explain the positional effect in our data.

A second limitation is the uncertainty about whether retrospective inference triggered by conditioning promotes the integration of neutral and aversive events into a single memory representation. While one integrated representation could facilitate more efficient threat prediction, inference-based decision-making can also be supported by separately stored memories, albeit with slower or less accurate performance in associative inference tasks (Schlichting, Mumford, & Preston, 2015; Zeithamova & Preston, 2017). Moreover, pharmacological studies suggest that reliance on two separately stored memories lead to reduced conditioned threat responses in sensory pre-conditioning (Wong et al., 2019). Thus, it remains unclear whether one integrated memory representation built across contexts is advantageous relative to maintaining separate representations. Future work could directly test whether integrated memory representations translate into behavioral advantages, such as more efficient avoidance behavior.

Finally, the reliability and generalizability of retroactive memory enhancements driven by emotional learning are not yet well established. Previous studies have shown that neutral information encoded before conditioning can later be strengthened in long-term episodic memory due to semantic relatedness (Dunsmoor et al., 2015; Patil et al., 2017), but replication attempts have cast doubt on the consistency of these findings (Kalbe & Schwabe, 2021a). Unlike this earlier work, which emphasized semantic knowledge, our study demonstrated selective modulation of memories based on newly acquired schemas of predictive relations. Notably, an independent neuroimaging study observed effects of predictive relations consistent with our results (Cooper et al., 2024). Nevertheless, further research is needed to clarify the reliability and boundary conditions of retroactive memory enhancements driven by emotional learning.

Conclusion

The present findings advance understanding of adaptive memory by showing that Pavlovian threat memory is not stored in isolation, but can retroactively interact with earlier mundane episodes. Neutral events that initially carry little relevance, and are temporally separated from aversive outcomes, can nevertheless acquire significance through inference supported by prior knowledge of event sequences. Those events that occur early in a predictive chain are especially likely to be prioritized in episodic memory, consistent with their adaptive role as early warning signals. These findings highlight that predictive schemas not only organize experiences into structured sequences but also allow disjoint episodes to be integrated across temporal gaps, biasing consolidation toward schema-relevant information. Taken together, episodic memories formed under safe conditions can be retroactively modulated by aversive learning when supported by prior relational knowledge, thereby broadening the defensive repertoire from detecting threats as they occur to predicting them ahead of time.

Research Transparency Statement

General Disclosures. Conflicts of interest: All authors declare no conflicts of interest. Funding: This work was supported by the Max Planck Society. BMB was supported by the Max Planck Society, the International Max Planck Research School "NeuroCom", and the University of Leipzig. AV is supported by the Max Planck Society. Artificial intelligence: ChatGPT (OpenAI, Inc., US) was used for editing language. Ethics: This research received approval from a local ethics board (ID: 482/18-ek).

Main Study. Preregistration: The hypotheses, methods, and analysis plan were preregistered before data collection at [LINK]. Minor deviations from the preregistration are reported in Supplemental Table S1. Materials: All study materials are publicly available at [LINK], with the exception of stimulus images, which cannot be shared due to copyright restrictions. Data: All primary data are publicly available at [LINK]. Analysis scripts: All analysis scripts are publicly available at [LINK].

Pilot Studies. Preregistration: No aspects of the pilot studies were preregistered. Materials: All study materials are publicly available at [LINK], with the exception of stimulus images, which cannot be shared due to copyright restrictions. Data: All primary data are publicly available at [LINK]. Analysis scripts: All analysis scripts are publicly available at [LINK].

References

Ambrose, R. E., Pfeiffer, B. E., & Foster, D. J. (2016). Reverse Replay of Hippocampal Place Cells Is Uniquely Modulated by Changing Reward. *Neuron*, *91*(5), 1124–1136. https://doi.org/10.1016/j.neuron.2016.07.047

Anderson, A. K., Wais, P. E., & Gabrieli, J. D. E. (2006). Emotion enhances remembrance of neutral events past. *Proceedings of the National Academy of Sciences*, *103*(5), 1599–1604. https://doi.org/10.1073/pnas.0506308103

Baczkowski, B. M., Haaker, J., & Schwabe, L. (2023). Inferring danger with minimal aversive experience. *Trends in Cognitive Sciences*, *27*(5), 456–467. https://doi.org/10.1016/j.tics.2023.02.005

Bartoń, K. (2023). *MuMIn: Multi-model inference*. https://CRAN.R-project.org/package=MuMIn.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Usinglme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bisby, J. A., Horner, A. J., Bush, D., & Burgess, N. (2018). Negative emotional content disrupts the coherence of episodic memories. *Journal of Experimental Psychology. General*, *147*(2), 243–256. https://doi.org/10.1037/xge0000356

Braun, E. K., Wimmer, G. E., & Shohamy, D. (2018). Retroactive and graded prioritization of memory by reward. *Nature Communications*, *9*(1), 4886. https://doi.org/10.1038/s41467-018-07280-0

Brogden, W. J. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology*, *25*(4), 323–332. https://doi.org/10.1037/h0058944

Clewett, D., & Davachi, L. (2017). The Ebb and Flow of Experience Determines the Temporal Structure of Memory. *Current Opinion in Behavioral Sciences*, *17*, 186–193. https://doi.org/10.1016/j.cobeha.2017.08.013

Clewett, D., Dunsmoor, J., Bachman, S. L., Phelps, E. A., & Davachi, L. (2022). Survival of the salient: Aversive learning rescues otherwise forgettable memories via neural reactivation and post-encoding hippocampal connectivity.

Neurobiology of Learning and Memory, *187*, 107572.

https://doi.org/10.1016/j.nlm.2021.107572

Cooper, S. E., Hennings, A. C., Bibb, S. A., Lewis-Peacock, J. A., & Dunsmoor, J. E. (2024). Semantic structures facilitate threat memory integration throughout the medial temporal lobe and medial prefrontal cortex. *Current Biology*, *34*(15), 3522–3536.e5. https://doi.org/10.1016/j.cub.2024.06.071

Cowan, E. T., Chanales, A. J., Davachi, L., & Clewett, D. (2024). Goal shifts structure memories and prioritize event-defining information in memory. *Journal of Cognitive Neuroscience*, *36*(11), 2415–2431.

https://doi.org/10.1162/jocn_a_02220

de Voogd, L. D., Fernández, G., & Hermans, E. J. (2016a). Awake reactivation of emotional memory traces through hippocampal-neocortical interactions. *NeuroImage*, *134*, 563–572. https://doi.org/10.1016/j.neuroimage.2016.04.026

de Voogd, L. D., Fernández, G., & Hermans, E. J. (2016b). Disentangling the roles of arousal and amygdala activation in emotional declarative memory. *Social Cognitive and Affective Neuroscience*, *11*(9), 1471–1480.

https://doi.org/10.1093/scan/nsw055

Dudai, Y., Karni, A., & Born, J. (2015). The Consolidation and Transformation of Memory. *Neuron*, *88*(1), 20–32. https://doi.org/10.1016/j.neuron.2015.09.004

Dunsmoor, J. E., Kroes, M. C. W., Moscatelli, C. M., Evans, M. D., Davachi, L., & Phelps, E. A. (2018). Event segmentation protects emotional memories from competing experiences encoded close in time. *Nature Human Behaviour*, *2*(4), 291–299. https://doi.org/10.1038/s41562-018-0317-4

Dunsmoor, J. E., Martin, A., & LaBar, K. S. (2012). Role of conceptual knowledge in learning and retention of conditioned fear. *Biological Psychology*, *89*(2), 300–305. https://doi.org/10.1016/j.biopsycho.2011.11.002

Dunsmoor, J. E., Murty, V. P., Davachi, L., & Phelps, E. A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature*, *520*(7547), 345–348. https://doi.org/10.1038/nature14106

Ezzyat, Y., & Davachi, L. (2021). Neural evidence for representational persistence within events. *The Journal of Neuroscience*, *41*(37), 7909–7920. https://doi.org/10.1523/jneurosci.0073-21.2021

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–114. https://doi.org/10.1016/j.neuropsychologia.2013.11.010

Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013

Green, S. R., Kragel, P. A., Fecteau, M. E., & LaBar, K. S. (2014). Development and validation of an unsupervised scoring system (Autonomate) for skin conductance response analysis. *International Journal of Psychophysiology*, *91*(3), 186–193. https://doi.org/10.1016/j.ijpsycho.2013.10.015

Hermans, E. J., Kanen, J. W., Tambini, A., Fernández, G., Davachi, L., & Phelps, E. A. (2017). Persistence of Amygdala-Hippocampal Connectivity and Multi-Voxel Correlation Structures During Awake Rest After Fear Learning Predicts Long-Term Expression of Fear. *Cerebral Cortex*, *27*(5), 3028–3041. https://doi.org/10.1093/cercor/bhw145

Holmes, N. M., Raipuria, M., Qureshi, O. A., Killcross, S., & Westbrook, F. (2018). Danger Changes the Way the Mammalian Brain Stores Information About Innocuous Events: A Study of Sensory Preconditioning in Rats. *eNeuro*, *5*(1). https://doi.org/10.1523/ENEURO.0381-17.2017

Hsieh, L.-T., Gruber, M. J., Jenkins, L. J., & Ranganath, C. (2014). Hippocampal activity patterns carry information about objects in temporal context. *Neuron*, *81*(5), 1165–1178. https://doi.org/10.1016/j.neuron.2014.01.015

Hurlemann, R., Hawellek, B., Matusch, A., Kolsch, H., Wollersen, H., Madea, B., … Dolan, R. J. (2005). Noradrenergic modulation of emotion-induced forgetting and remembering. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(27), 6343–6349.

https://doi.org/10.1523/JNEUROSCI.0228-05.2005

Kalbe, F., & Schwabe, L. (2020). Beyond arousal: Prediction error related to aversive events promotes episodic memory formation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *46*(2), 234–246. https://doi.org/10.1037/xlm0000728

Kalbe, F., & Schwabe, L. (2021a). On the search for a selective and retroactive strengthening of memory: Is there evidence for category-specific behavioral tagging? *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0001075

Kalbe, F., & Schwabe, L. (2021b). Prediction Errors for Aversive Events Shape Long-Term Memory Formation through a Distinct Neural Mechanism. *Cerebral Cortex*, bhab402. https://doi.org/10.1093/cercor/bhab402

Klein-Flügge, M. C., Wittmann, M. K., Shpektor, A., Jensen, D. E. A., & Rushworth, M. F. S. (2019). Multiple associative structures created by reinforcement and incidental statistical learning mechanisms. *Nature Communications*, *10*(1), 1–15. https://doi.org/10.1038/s41467-019-12557-z

Kurth-Nelson, Z., Barnes, G., Sejdinovic, D., Dolan, R., & Dayan, P. (2015). Temporal structure in associative retrieval. *eLife*, *4*, e04919. https://doi.org/10.7554/eLife.04919

Laing, P. A. F., & Dunsmoor, J. E. (2025). Event segmentation promotes the reorganization of emotional memory. *Journal of Cognitive Neuroscience*, *37*(1), 110–134. https://doi.org/10.1162/jocn_a_02244

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*, 155–184. https://doi.org/10.1146/annurev.neuro.23.1.155

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences*, *15*(8), 343–351. https://doi.org/10.1016/j.tics.2011.06.004

Liu, Y., Dolan, R. J., Kurth-Nelson, Z., & Behrens, T. E. J. (2019). Human Replay Spontaneously Reorganizes Experience. *Cell*, *178*(3), 640–652.e14.

https://doi.org/10.1016/j.cell.2019.06.012

Liu, Y., Mattar, M. G., Behrens, T. E. J., Daw, N. D., & Dolan, R. J. (2021).
Experience replay is associated with efficient nonlocal learning. *Science (New York, N.Y.)*, *372*(6544), eabf1357. https://doi.org/10.1126/science.abf1357

Macias, C., & Persaud, K. (2024). From silos to synergy: Integrating approaches to investigate the role of prior knowledge and expectations on episodic memory. *Psychonomic Bulletin &Amp; Review*, *31*(6), 2390–2409. https://doi.org/10.3758/s13423-024-02505-4

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, *21*(11), 1609–1617. https://doi.org/10.1038/s41593-018-0232-z

McGaugh, J. L. (2018). Emotional arousal regulation of memory consolidation. *Current Opinion in Behavioral Sciences*, *19*, 55–60. https://doi.org/10.1016/j.cobeha.2017.10.003

Mobbs, D., Headley, D. B., Ding, W., & Dayan, P. (2020). Space, Time, and Fear: Survival Computations along Defensive Circuits. *Trends in Cognitive Sciences*, *24*(3), 228–241. https://doi.org/10.1016/j.tics.2019.12.016

Murayama, K., & Kitagami, S. (2014). Consolidation power of extrinsic rewards: Reward cues enhance long-term memory for irrelevant past events. *Journal of Experimental Psychology. General*, *143*(1), 15–20. https://doi.org/10.1037/a0031992

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 263–273. https://doi.org/10.1037/0278-7393.33.2.263

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determinationR2and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Nakagawa, S., & Schielzeth, H. (2012). A general and simple method for obtaining

    R2 from generalized linear mixed-effects models. *Methods in Ecology and*

    Evolution, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Patil, A., Murty, V. P., Dunsmoor, J. E., Phelps, E. A., & Davachi, L. (2017).

    Reward retroactively enhances memory consolidation for related items. *Learning*

    & Memory (Cold Spring Harbor, N.Y.), *24*(1), 65–69.

    https://doi.org/10.1101/lm.042978.116

R Core Team. (2017). *R: A language and environment for statistical computing*.

    Vienna, Austria: R Foundation for Statistical Computing.

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is.

    American Psychologist, *43*(3), 151–160.

    https://doi.org/10.1037/0003-066x.43.3.151

Rescorla, R. A., & Holland, P. C. (1982). Behavioral studies of associative learning in

    animals. *Annual Review of Psychology*, *33*(1), 265–308.

    https://doi.org/10.1146/annurev.ps.33.020182.001405

Ritchey, M., McCullough, A. M., Ranganath, C., & Yonelinas, A. P. (2017). Stress

    as a mnemonic filter: Interactions between medial temporal lobe encoding

    processes and post-encoding stress. *Hippocampus*, *27*(1), 77–88.

    https://doi.org/10.1002/hipo.22674

Ritchey, M., Murty, V. P., & Dunsmoor, J. E. (2016). Adaptive memory systems for

    remembering the salient and the seemingly mundane. *Behavioral and Brain*

    Sciences, *39*. https://doi.org/10.1017/s0140525x15001922

Roozendaal, B., & McGaugh, J. L. (2011). Memory modulation. *Behavioral*

    Neuroscience, *125*(6), 797–824. https://doi.org/10.1037/a0026187

Rouhani, N., Niv, Y., Frank, M. J., & Schwabe, L. (2023). Multiple routes to

    enhanced memory for emotionally relevant events. *Trends in Cognitive Sciences*,

    27(9), 867–882. https://doi.org/10.1016/j.tics.2023.06.006

Sakaki, M., Fryer, K., & Mather, M. (2014). Emotion Strengthens High-Priority

    Memory Traces but Weakens Low-Priority Memory Traces. *Psychological*

Science, *25*(2), 387–395. https://doi.org/10.1177/0956797613504784

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

Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R.
J., … Frackowiak, R. S. (2004). Temporal difference models describe higher-order
learning in humans. *Nature*, *429*(6992), 664–667.
https://doi.org/10.1038/nature02581

Sherman, B. E., & Turk-Browne, N. B. (2020). Statistical prediction of the future
impairs episodic encoding of the present. *Proceedings of the National Academy
of Sciences*, *117*(37), 22760–22770. https://doi.org/10.1073/pnas.2013291117

Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in
Cognitive Sciences*, *14*(10), 464–472. https://doi.org/10.1016/j.tics.2010.08.002

Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain:
Hippocampal-midbrain encoding of overlapping events. *Neuron*, *60*(2), 378–389.
https://doi.org/10.1016/j.neuron.2008.09.023

Singer, A. C., & Frank, L. M. (2009). Rewarded outcomes enhance reactivation of
experience in the hippocampus. *Neuron*, *64*(6), 910–921.
https://doi.org/10.1016/j.neuron.2009.11.016

Staresina, B. P., Alink, A., Kriegeskorte, N., & Henson, R. N. (2013). Awake
reactivation predicts memory in humans. *Proceedings of the National Academy
of Sciences*, *110*(52), 21159–21164. https://doi.org/10.1073/pnas.1311989110

Starita, F., Kroes, M. C. W., Davachi, L., Phelps, E. A., & Dunsmoor, J. E. (2019).
Threat learning promotes generalization of episodic memory. *Journal of
Experimental Psychology: General*, *148*(8), 1426–1434.
https://doi.org/10.1037/xge0000551

Strange, B. A., Hurlemann, R., & Dolan, R. J. (2003). An emotion-induced
retrograde amnesia in humans is amygdala- and beta-adrenergic-dependent.

Proceedings of the National Academy of Sciences of the United States of America, *100*(23), 13626–13631. https://doi.org/10.1073/pnas.1635116100

Talmi, D. (2013). Enhanced Emotional Memory: Cognitive and Neural Mechanisms. *Current Directions in Psychological Science*, *22*(6), 430–436. https://doi.org/10.1177/0963721413498893

Tambini, A., & Davachi, L. (2013). Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory. *Proceedings of the National Academy of Sciences*, *110*(48), 19591–19596. https://doi.org/10.1073/pnas.1308499110

Tompary, A., & Davachi, L. (2024). Integration of overlapping sequences emerges with consolidation through medial prefrontal cortex neural ensembles and hippocampal–cortical connectivity. *eLife*, *13*. https://doi.org/10.7554/elife.84359

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., … Morris, R. G. M. (2007). Schemas and memory consolidation. *Science (New York, N.Y.)*, *316*(5821), 76–82. https://doi.org/10.1126/science.1135935

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

Williams, S. E., Ford, J. H., & Kensinger, E. A. (2022). The power of negative and positive episodic memories. *Cognitive, Affective, &Amp; Behavioral Neuroscience*, *22*(5), 869–903. https://doi.org/10.3758/s13415-022-01013-z

Wimmer, G. E., & Shohamy, D. (2012). Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions. *Science*, *338*(6104), 270–273. https://doi.org/10.1126/science.1223252

Wise, T., Liu, Y., Chowdhury, F., & Dolan, R. J. (2021). Model-based aversive learning in humans is supported by preferential task state reactivation. *Science Advances*, *7*(31), eabf9616. https://doi.org/10.1126/sciadv.abf9616

Wong, F. S., Westbrook, R. F., & Holmes, N. M. (2019). 'Online' integration of sensory and fear memories in the rat medial temporal lobe. *eLife*, *8*, e47085.

https://doi.org/10.7554/eLife.47085

Zeithamova, D., & Preston, A. R. (2017). Temporal Proximity Promotes Integration of Overlapping Events. *Journal of Cognitive Neuroscience, 29*(8), 1311–1323. https://doi.org/10.1162/jocn_a_01116