Distortion of Overlapping Memories Relates to Arousal and Anxiety

Erin Morrow and David Clewett[*]

Department of Psychology, University of California, Los Angeles

*To whom correspondence should be addressed:

Dr. David Clewett
Department of Psychology
5558 Pritzker Hall
University of California, Los Angeles
Los Angeles, CA, 90095

Keywords: memory, arousal, interference, pupil, individual differences, anxiety

Open Practices Statement: Data or materials for the experiments will be available upon acceptance of this manuscript.

Acknowledgements: The authors thank research assistants Lily Zeng and Amna Ali for help with data collection. We also thank members of the ACME and Rissman Labs at UCLA for useful feedback.

ABSTRACT

Everyday experiences often overlap, challenging our ability to maintain distinct episodic memories. One way to resolve such interference is by exaggerating subtle differences between remembered events, a phenomenon known as memory repulsion. Here, we tested if repulsion is influenced by emotional arousal, when resolving memory interference is perhaps most needed. We adapted an existing paradigm in which participants repeatedly studied object-face associations. Participants studied two different colored versions of each object: a to-be-tested 'target' and its not-to-be-tested 'competitor' pairmate. The level of interference between target and competitor pairmates was manipulated by making the object colors either highly similar or less similar, depending on the participant group. To manipulate arousal, the competitor object-face associations were preceded by either a neutral tone or an aversive, arousing white noise burst. Memory distortion for the color of the target objects was tested after each round of learning to examine if memory distortions gradually emerge over time. We found that participants with greater sound-associated pupil dilations, an index of physiological arousal, showed greater memory attraction of target colors towards highly similar competitor colors. Greater memory attraction was also related to greater memory interference at the end of learning. Additionally, individuals who self-reported higher trait anxiety showed greater memory attraction when one of the memories was aversive. Our findings suggest that memories of similar neutral and arousing events may blur together over time, especially in individuals who show higher arousal responses and symptoms of anxiety.

INTRODUCTION

Throughout our lives, we encounter a vast amount of highly similar information. Because we simply cannot process everything, our memory systems are challenged with sorting through this clutter to prioritize and retain what is most important. For instance, most people take a similar route to work each day. Each of these daily events contains common features, such as the identity of the bus driver and the advertisements displayed above the seats. Thus, memories for these experiences significantly overlap with one another, leading to interference — a phenomenon by which shared mnemonic features lead to forgetting (Osgood, 1949; Barnes & Underwood, 1959; Mensink & Raaijmakers, 1988; Anderson & Spellman, 1995; Wixted, 2004; Zhao, Chanales, & Kuhl, 2021). In this case, you may have difficulty remembering the specific details of any particular morning's commute because their respective representations are competing with one another during retrieval.

Forgetting various aspects of interfering memories can either be helpful or harmful depending on one's goals. In certain situations, it is important to preserve highly detailed memories of specific episodes. Here, it is adaptive for memory representations to be kept distinct from one another – that is, for interference to be resolved so that one can retrieve a specific experience. One way to resolve memory interference is through a process known as pattern separation. Pattern separation occurs when episodic memories with similar content are differentiated from each other to help maintain distinct memory representations of both unique events (Müller & Pilzecker, 1900; Underwood, 1957; Underwood & Postman, 1960). Although existing pattern separation paradigms have helped reveal how successfully individuals can discriminate between overlapping memories (Kirwan & Stark, 2007; Yassa & Stark, 2011), they do not fully capture how their underlying memory representations are being transformed to reduce interference. Identifying these mechanisms is essential for understanding how humans accurately store and access important memories to guide specific, context-appropriate behaviors. Pattern separation behaviors are also often explored using one-shot episodic encoding, raising the question of how similar memories become differentiated when individuals have multiple opportunities to learn the differences between those experiences.

Intriguingly, recent studies demonstrate a memory repulsion effect that emerges in time when similar memories interfere with each other. In these instances, subtle differences between similar memory representations become slightly exaggerated or ‘repulsed’ away from their original forms to reduce their overlap (Chanales et al., 2017, 2021; Drascher & Kuhl, 2022; Hulbert & Norman, 2015; Zhao, Chanales, & Kuhl, 2021). For example, if an object stimulus is a red color, it may be remembered as having a deeper red hue than it actually does. Although distorting memory representations might seem counterintuitive for achieving better performance, repulsion is surprisingly effective at resolving interference (Chanales et al., 2021). Ultimately, the price we pay for keeping similar memories distinct from each other is distortion, because this process helps prevent these memories from clashing together during retrieval.

Although work on memory repulsion is relatively sparse, this phenomenon has been observed for a variety of perceptual features learned over time. In one influential study, Chanales and colleagues (2021) demonstrated that – over a gradual learning process – participants remembered similarly colored objects as having an exaggerated hue with respect to their competitor pairmates. Importantly, repulsion was only observed when there was a moderate level of color similarity between the competing object associations, and memory distortions were not observed when color similarity was too low or too high. These findings are consistent with the non-monotonic plasticity model of memory separation, which argues that intermediate levels of similarity present the greatest need for mnemonic discrimination (see Ritvo, Turk-Browne, & Norman, 2019). Extremely similar memories, on the other hand, may be more likely to benefit from memory integration than separation. Likewise, memories that share little overlap do not require strong pattern separation processes, nor do they need to be integrated to acquire knowledge or to generalize. Additionally, Chanales and colleagues (2021) found that systematic memory distortions persisted for 24 hours and were adaptive, as evidenced by greater repulsion being related to better memory for faces associated with each object.

Increasing neuroimaging evidence has also revealed evidence of memory separation and repulsion effects in the brain, particularly in the hippocampus. Both rodent and human research demonstrate that these memory discrimination processes are primarily supported by activation of the dentate gyrus (DG) subregion of the hippocampus. By increasing representational dissimilarity (see Grella & Donaldson, 2024), this subfield promotes encoding of separate representations of highly similar memories (for review, see Yassa & Stark, 2011). Strikingly, several studies have also shown that hippocampal representations of two similar events can become more different from each other than dissimilar events over the course of learning (Xue, 2022; e.g., Chanales et al., 2017; Dimsdale-Zucker et al., 2018; Favila, Chanales, & Kuhl, 2016; Hsieh et al., 2014; Kunz et al., 2019; Kyle et al., 2015). For instance, using a route navigation paradigm, Chanales and colleagues (2017) showed that hippocampal activation patterns become increasingly dissimilar for overlapping paths than for non-overlapping paths over repeated exposures to different city routes. Critical to the idea that repulsion is an adaptive memory process, this kind of hippocampal differentiation – particularly in the DG/CA3 subregions – was correlated with successful learning (Wanjia et al., 2021; Xue, 2022). These findings reveal a neuromechanism that differentiates memories beyond the mere orthogonalization of pattern separation (Xue, 2022). Together, recent behavioral and neuroimaging findings implicate memory repulsion as an important process for resolving interference and improving long-term recall of specific episodic memories.

Yet, an important open question is how memories might become warped to reduce interference for between overlapping neutral and emotionally arousing events. Many of the most significant events in our lives – those for which we want to retain distinctive memories and avoid forgetting – are affective in nature and contain vivid detail (e.g., Kensinger, Garoff-Eaton, & Schacter, 2006; Williams et al., 2022). Up to this point, repulsion effects in memory have only been examined for neutral event features, such as household objects (Chanales et al., 2021), images of spatial routes (Chanales et al., 2017), and scenes (Favila, Chanales, & Kuhl, 2016). In contrast, studies examining learning-dependent distortions in emotional memories are scarce, despite the possibility that memory differentiation processes are likely to be influenced by changes in arousal.

Indeed, much evidence indicates that emotional memories undergo more interference than neutral memories (Barnier, Hung, & Conway, 2004; Hensley, Otani, & Knoll, 2019; Novak & Mather, 2009; Sison & Mather, 2007), perhaps because emotion can serve as a salient category that evokes competition between similar memories (Mather, 2009; see e.g., Schulkind & Woldorf, 2005) or because emotional memories are more resistant to updating (Mather & Knight, 2008; Nashiro et al., 2013). Importantly, emotion-enhanced interference implies a greater need for resolution in emotionally arousing contexts. Given that remembering detailed information about arousing events is often important for maintaining a subjective sense of wellbeing, it would be adaptive to exaggerate the differences between overlapping details to improve discriminations between negative and similar neutral events. Motivated by this idea, the first aim of the current study was to test the hypothesis that arousal would exacerbate memory repulsion between highly similar events.

Support for this prediction comes from a diverse emotional memory literature. In addition to a potentially greater need for interference resolution, emotion has complex – and sometimes opposing – effects on memory. This dichotomy is consistent with the phenomenon of memory repulsion (i.e., a systematic bias in memory that actually improves overall recall). Namely, emotion can induce both subjective memory distortions and objective memory improvements, though these two effects are seldom connected to each other in the same paradigm. Arousal-induced attentional narrowing can increase the likelihood that individuals endorse inaccurate peripheral information, or 'false memories' (Kaplan et al., 2016), even though emotional events are generally remembered more accurately (e.g., Hamann, 2001; Reisberg & Heuer, 1992; Talarico & Rubin, 2003) and with greater perceptual detail (Kensinger, Garoff-Eaton, & Schacter, 2006; Mather, 2007) than neutral events.

Consistent with the prediction that emotion-enhanced interference would necessitate greater memory repulsion, neuroimaging and behavioral evidence show that arousal-driven modulation of the amygdala may enhance the discrimination of similar emotional items in the DG/CA3 subregions of the hippocampus (Leal et al., 2014). This arousal-driven modulation of emotion and memory-related brain regions could be driven, at least in part, by projections from the locus coeruleus (LC), the brain's primary supplier of norepinephrine to the brain (Aston-Jones and Cohen, 2005; Sara, 2009). Of relevance to the current study, noradrenergic projections are especially dense in the DG subregion of the hippocampus (Harley, 2007; Grella & Donaldson, 2024), the subfield that is essential for supporting pattern separation and memory repulsion effects. Recent evidence in rodents demonstrates that LC inputs promote the ‘remapping’ of memory representations in the hippocampus (Grella et al., 2019; Grella & Donaldson, 2024). Moreover, work in humans shows that an indirect salivary marker of noradrenergic system activity, alpha amylase, is associated with enhanced pattern separation performance in the presence of arousal (Segal et al., 2012). Together, these converging findings support the hypothesis that arousal facilitates memory repulsion in the service of resolving interference.

In addition to examining how arousal affects memory repulsion processes, the second goal of the current study was to determine if arousal-related repulsion effects are related to symptoms of affective disorders like anxiety. It is well known that certain clinical populations tend to overgeneralize unpleasant emotions from negative memories to otherwise neutral situations. This fear generalization can produce maladaptive or context-inappropriate behaviors, such as excessive avoidance of contexts similar to the unpleasant memory (e.g., Krypotos et al., 2015). Critically, resolving interference between arousing and non-arousing memories may be especially difficult for those with anxiety disorders or stress due to changes in the structure and function of the hippocampus (see Besnard & Sahay, 2015). These changes could account for poorer behavioral pattern separation among those with anxiety disorders (see Kheirbek et al., 2012). Therefore, it is possible that participants with greater symptom severity – and thus more frequent generalization between similar arousing and neutral events – might show poor discrimination performance between these memories. In this case, memories may become increasingly blended instead of separated, perhaps facilitating the spread of fear or negative emotion.

In support of this idea, individuals with generalized anxiety disorder (GAD) exhibit greater conditioned fear responses to cues similar to arousing images, compared to those without GAD (Lissek et al., 2014). Additionally, a recent meta-analysis of more than 1,000 participants revealed that trait anxiety in non-clinical populations is also associated with the tendency to generalize fear (Sep et al., 2019). In contrast, individuals with major depression do not appear to experience greater fear generalization than healthy individuals (Wurst et al., 2021; see also Park, Lee, & Lee, 2018). Yet, individuals with depression often exhibit overgeneralized autobiographical memories (OGM; Moore & Zoellner, 2007; Williams et al., 2007). Individuals with OGM tend to retrieve general *classes* of events than distinct episodes. Although speculative, it is possible that OGM increases the representational similarity between overlapping negative and neutral memories even further, making them more difficult to differentiate. From this perspective, individuals with depression may be more likely to show smaller subjective repulsion effects between arousing and neutral memories, as well as increased memory interference. This hypothesis is supported by initial evidence indicating that depression severity is associated with changes in brain networks that support emotional pattern separation (Leal et al., 2014), although it is still unclear how depression affects discrimination between memories associated with different levels of arousal.

In summary, the first goal of the present study was to investigate whether arousal influences memory repulsion between overlapping events as a function of perceptual similarity. We hypothesized that arousal would selectively facilitate memory repulsion between highly similar events; that is, inducing a state of arousal during object-face association learning would bias memory for perceptual details of highly overlapping events farther away from their original form. Our second goal was to test if memory repulsion is related to lower interference, demonstrating the adaptive nature of mentally distancing similar memories. Accordingly, we predicted that greater memory repulsion would relate to better associative memory accuracy, irrespective of perceptual similarity. Third, we predicted that one's degree of memory repulsion would be inversely correlated with self-reported levels of trait anxiety and depression.

Importantly, we also used eye-tracking to capture participants' pupil diameter as a read-out of trial-by-trial arousal responses. This physiological measure has been linked to transient activity of the locus coeruleus-norepinephrine (LC-NE) system (Clewett et al., 2018; Joshi et al., 2016; Murphy et al., 2014; Reimer et al., 2016; Varazzani et al., 2015), which responds to stimuli that are relevant, novel, rewarding, or threatening (Aston-Jones & Cohen, 2005). As such, pupil measures may provide mechanistic insights into whether noradrenergic modulation plays a role in shaping adaptive memory distortions.

METHODS

Participants and design. Sixty-nine participants were recruited to participate in this experiment. Eligibility requirements included age (18-35 years), native or fluent English proficiency, normal or corrected-to-normal vision (including color vision), and normal hearing. However, several participants were not included in the final sample; one participant withdrew during the experiment; data from the final memory tests were lost for one participant due to technical issues; and there were technical issues for one other participant. Therefore, the final experiment sample consisted of 66 participants (51 female, 14 male, 1 unspecified; M_{age} = 20.6, SD_{age} = 2.4). All participants provided verbal informed consent approved by the UCLA Institutional Review Board and received course credit for participation.

Sample sizes were estimated using a power analysis based on Chanales et al. (2021). Using the weaker effect size across the two relevant experiments (from one-sample *t*; Cohen's *d* = 0.51), we determined that 33 participants are needed to obtain a power of 0.8. Given that the current study involves two between-subjects conditions, we sought to recruit 66 total participants.

In terms of sample demographics, 3.0% of participants identified as American Indian/Alaskan, 40.9% as Asian, 1.5% as Pacific Islander, 42.4% as White, and 12.0% as more than one race or other. 22.7% of participants identified as being of Hispanic origin and 75.8% as being of non-Hispanic origin, regardless of race (1.5% unspecified).

Stimuli. During object-face encoding, participants viewed 36 unique object images and 72 unique face images. The object images depicted common objects, such as a blender or sofa, displayed on a white background (400 x 400 px total). The majority of these images were sourced from an object dataset intended for color rotation (Brady et al., 2013; Chanales et al., 2021). This stimulus set was supplemented with two additional object images that were created with the assistance of DALL-E 2 and resized to match the size of the other images. All object images were color-rotated within the experiment using open-source MATLAB code from Chanales et al. (2021). The colors to which the objects were rotated were in 4° increments on a 360° color wheel (e.g., 0°, 4°, 8°, etc.).

Face images depicted white, middle-aged to older men with roughly neutral expressions (250 x 250 px total). These images were sourced from a previous study on memory repulsion

(Chanales et al., 2021) and supplemented with faces from the Georgia Tech Face Database (Nefian, Khosravi, & Hayes III, 1997), Caltech 10k Web Faces (Angelova, Abu-Mostafa, & Perona, 2005), and Caltech Face Dataset 1999 (Weber, 2022). Minor brightness adjustments were made in the image editing software GIMP (The GIMP Development Team, 2019) to normalize luminance.

During the Study Phase of the paradigm, each object image was associated with two unique face images. However, these two object-face associations differed with respect to the object's color. For instance, a participant might study both a bright red blender ('target') associated with one face and a maroon blender ('competitor') associated with a different face. Here, 'targets' refer to the specific objects that will be tested during the subsequent memory tests. By contrast, 'competitors' refer to their object pairmates that did not appear during the actual memory tests. Competitors only appeared during learning to induce interference. The degree of color similarity varied systematically between targets and competitors, separated by either 24° (high similarity condition) or 72° (low similarity condition) on the color wheel. Half (*n* = 33) of the participants in the current sample were assigned to the High Similarity group, and the other half of participants were assigned to the Low Similarity group. The color pairmates (e.g., bright red and maroon) assigned to a particular object were randomized for each participant.

To manipulate arousal, one of two different types of sounds was played just prior to studying each object-face association (i.e., within subjects). These sounds consisted of aversive white noise bursts and neutral tones created in Audacity version 3.1.3.0 (Audacity Team, 2021). White noise bursts are often used as aversive auditory stimuli to induce arousal across a variety of contexts (e.g., Hamm et al., 1991; Peri et al., 2000; Wang et al., 2012). We opted to use these simple sounds as opposed to naturalistic sounds (e.g., a scream) to avoid potential confounds associated with semantic content and prevent an additional layer of sound-related learning to the task. In the current study, aversive white noise bursts ranged from 0.8-1 loudness in Audacity (with 1 being maximum volume), whereas pure tones ranged from 230-340 Hz. These values were chosen to evoke a large difference in the subjective aversiveness of the white noise bursts and neutral tones. Laptop system volume was set to approximately two-thirds of the maximum, unless participants explicitly asked for experimenters to lower the volume.

Pupil tracking. We used the EyeLink 1000 Plus system (SR Research Ltd., version 5.15) to measure pupil size continuously throughout encoding. Pupil diameter is known to be a reliable, real-time correlate of physiological arousal (see Huang & Clewett, forthcoming), enabling us to verify that aversive white noise bursts induced significantly greater physiological arousal than neutral tones. To compute sound-evoked pupil dilations, we first measured average pupil diameter in pixels during a window between 0.75-s and 1.25-s following sound onset. This window was chosen to capture the dilatory peak. Pupil diameter during this period was normalized to a 0.5-s pre-sound baseline for each trial (Huang & Clewett, forthcoming). Raw pupil data was preprocessed using the ET-remove-artifacts toolbox to remove blinks and other abnormalities (Mather et al., 2020) and subsequently averaged and analyzed using custom MATLAB code. Pupil data was measured from the left eye unless there was an experimenter note suggesting that data quality appeared higher in the right eye; in that case, the right eye was used (*n* = 2).

Eye-tracking data from *n* = 11 participants was excluded from analyses for the following reasons. Some participants either wore glasses during the experiment, leading to noisy data (*n* = 3), or all of their training rounds had 25% or more missing data (*n* = 8). Additional participants (*n* = 12) had one or more training rounds excluded from analysis because of similarly poor data quality (*n* = 9), technical issues (*n* = 2), or a combination of poor data quality and technical issues (*n* = 1). This left a total of 55 participants with at least partially usable pupil data.

Procedure

Participants completed a 2-hr modified version of an existing associative memory paradigm (Chanales et al., 2021). This version involved five training rounds, each consisting of a Study Phase, an Associative Memory Test, and/or a Color Memory Test. This interleaving pattern of encoding and retrieval is thought to optimize memory differentiation over time (Chanales et al., 2021; Hulbert & Norman, 2015; Storm et al., 2008). After five training rounds, participants completed three Color Memory Tests (averaged) and an Associative Memory Test to assess final memory performance and levels of interference (**Procedure, Figure 1**).

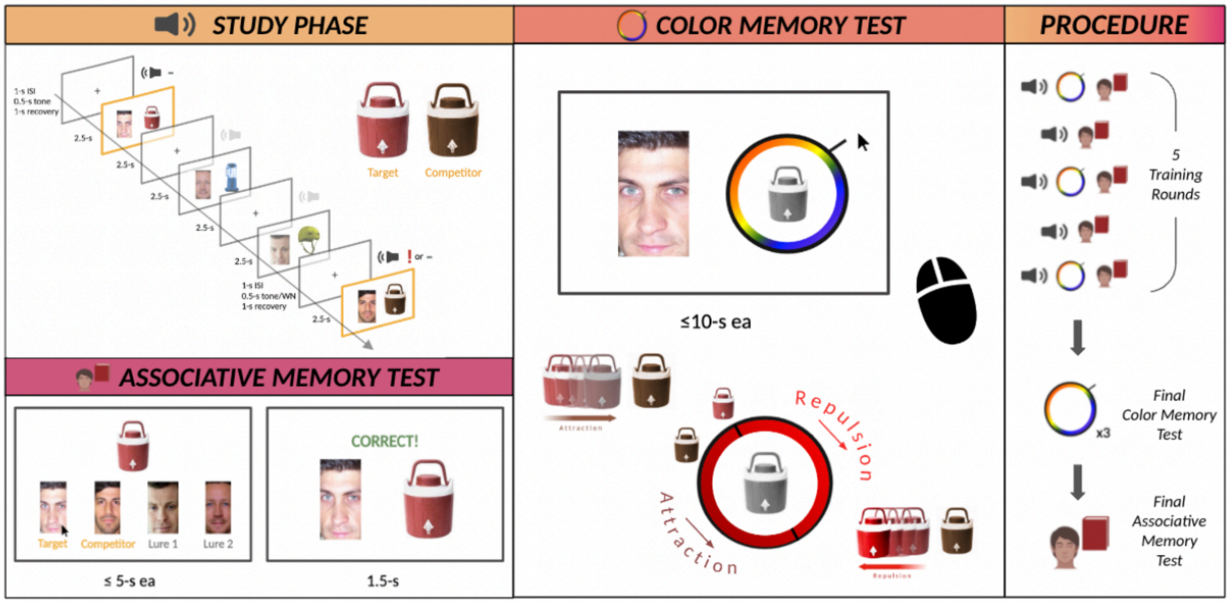

Figure 1. Memory repulsion and associative interference paradigm. The experiment consisted of multiple rounds of a Study Phase, Associative Memory Test, and/or Color Memory Test. In the Study Phase (top left panel), participants sequentially encoded different object-face associations. Critically, there were two versions of each object, referred to as the 'target' and 'competitor,' each associated with a unique face. The High Similarity group encoded targets and competitors that were highly similar in color (24° apart), while the Low Similarity group encoded targets and competitors that were less similar in color (72° apart). Before competitor object-face associations were shown, participants heard either an aversive white noise burst or a neutral tone to manipulate participants' arousal level. The Color Memory Test (middle panel) was used to assess subjective memory biases, or distortions, as participants were presented with target objects in grayscale and asked to select their remembered colors on a 360-degree wheel. In the Associative Memory Test (bottom left panel), participants had to identify the correct object-face associations for target objects only. Interference errors occurred when participants selected the face that had been associated with the target object's competitor. The procedure (right panel) was as follows: first, participants completed five training rounds that contained the above tasks, interleaved. Then, participants completed final Color Memory Tests and a final Associative Memory Test.

Clinical Questionnaires. Participants completed a standard demographic questionnaire, the State-Trait Anxiety Inventory (STAI; Spielberger, 1983), and the Beck Depression Inventory (BDI; Beck et al., 1961), excluding questions regarding suicide. One participant's STAI-S (state anxiety) score was excluded from analyses because they did not fully complete the questionnaire.

Object-Face Association Study Phase. During each Study Phase of the task, participants studied 72 object-face associations one at a time (2.5s each; **Study Phase**, **Figure 1**). Participants were asked to imagine the person whose face was depicted interacting with the object to encourage encoding rich associations. As previously described, each association consisted of a unique face and either a target object (to-be-tested during the memory tests) or competitor object (not tested). Presentation order was pseudorandomized such that targets and their competitors did not appear consecutively. Prior to studying each association, participants viewed a fixation cross for 2.5s. This total fixation period was further divided into an interstimulus interval (ISI) of 1s, sound presentation for 0.5s, and then a pupil recovery period for 1s. The sound presented before target objects was always a neutral tone. By contrast, the sound presented before competitor objects could be either a neutral tone or an aversive burst of white noise. In this way, the to-be-tested target objects were always associated with a neutral sound, reducing the likelihood of reinstating arousal during retrieval. This manipulation also enabled us to query how memory distortions emerge as a function of arousal being associated with a competing, overlapping memory. Participants were instructed that they did not need to memorize the sounds that were presented (**Study Phase, Figure 1**).

Object Color Memory Test. Color Memory Tests were used to assess participants' bias in color memory for the target objects; that is, whether there were color memory repulsion effects. Color Memory Tests only appeared on odd-numbered training rounds (1, 3, and 5) to prevent an excessively long session (i.e., longer than 2 hours). This approach was chosen to match prior working demonstrating color memory repulsion effects (Experiment 2 in Chanales et al., 2021).

During each of the 36 trials of the Color Memory Test, participants were presented with a target object in grayscale alongside its associated face (**Color Memory Test, Figure 1**). Participants were instructed to use their computer mouse to move a dial around a color wheel and select the remembered color of the object (10-s max). As the dial moved around the wheel, the hue of the object changed to provide visual feedback. Participants advanced to the next trial once they made a response. The color wheel was randomly rotated each trial such that the same mouse position did not correspond to the same color across trials, encouraging participants to engage with the task.

For the memory analyses, colors that were selected in the direction of the untested competitor object represented memory attraction, whereas colors selected in the opposite direction of the untested competitor represented memory repulsion. For example, consider that a tested target object's color was at 48° on the wheel and its untested object competitor's color was at 72° on the wheel. Responses between 48° and 228° (48° + 180°), a window which encompasses the location of the competitor object's color, would be considered attraction responses. Responses between 48° and 228° on the opposite side of the wheel (48° - 180°) would be considered repulsion responses. A response exactly equal to the target color (e.g., 48°) or exactly 180° away (e.g., 228°) would have been considered neither attraction nor repulsion (**Color Memory Test, Figure 1**). Color memory data was excluded from a particular training round if participants were unresponsive to >50% of trials.

Object-Face Associative Memory Test. Associative Memory tests were used to assess overall memory performance for each object-face association and to assess interference as a function of arousal during learning and the amount of perceptual overlap, or color similarity, between target-competitor object pairmates (**Figure 1; Associative Memory Test**). During each of the 36 trials of the test, participants were presented with a target object in its original color and were instructed to choose which face was previously associated with that object out of four alternative choices (5-s max). Among the four choices, one was a target (i.e., the correct face association), one was a competitor (i.e., the face associated with the untested competitor object), and the two others were lures (i.e., faces associated with other objects that were neither the target nor competitor object). Choice order was randomized. Trials in which the participant selected the competitor were considered 'interference errors,' as the target and competitor associations were designed to interfere with each other during retrieval. After making their choice, participants immediately advanced to a 1.5-s feedback screen, which displayed either the word "correct" or "incorrect," along with the correct object-face association. Associative memory data was excluded from a particular training round if participants were unresponsive to >50% of trials.

ANALYSIS AND RESULTS

Behavioral and pupil measures

First, we will define several key measures that were used throughout the analyses.

Color memory bias. We used two different metrics to operationalize color memory bias: color distance (in degrees) and the percentage of responses 'away' from the competitor. Color distance is a signed measure that is calculated by subtracting the location of the remembered color on the wheel (in degrees) from the location of the true target color *(true color – remembered color)* on the Color Memory Test. The percentage of responses away from the competitor was computed by assigning a value of '1' to trials with a *positive* color distance value (i.e., repulsion of the target object's color *away from* the competitor object's color) and a value of '0' to trials with a *negative* color distance (i.e., attraction of the target object's color *towards* the competitor object's color). These values were averaged over the entire test, yielding a decimal measure that was rounded to the nearest hundredth. In this way, color distance offered a more precise estimate of memory bias, while percentage of away responses helped diminish the influence of extreme color responses by categorizing each response simply as an instance of either repulsion or attraction.

Memory interference. To operationalize the magnitude of memory interference, we used the total number of interference errors from the Associative Memory Test. As noted previously, an interference error occurred when the participant chose the face that was associated with the target object's competitor on a given trial. These endorsement errors were then summed over all trials of a given round to estimate the amount of memory interference in that round.

Cumulative pupil dilation across encoding. Given that memory distortions develop over the course of learning (see Chanales et al., 2021), we reasoned that pupil responses during all five training rounds would capture the ongoing influence of physiological arousal on encoding and mnemonic discrimination processes. Pupil dilation associated with a given competitor object was summed across the five training rounds, and this score was averaged across all objects within a given participant to form a cumulative sound-evoked pupil dilation measure.

Participants were excluded from analysis if an entire training round of their eye-tracking data had been excluded (*n* = 12). Missing pupil data from one round would bias a measure that sums data over five rounds. Reasons for missing rounds are described in the *Pupil Tracking* section.

Associative and object color learning occurred over training rounds.

We first conducted several tests to determine whether associative learning had occurred over the five training rounds.

Associative memory. First, we focused on performance on the Associative Memory Tests. Here, we excluded *n* = 4 participants that were missing Associative Memory Test data for one training round. Of these participants, *n* = 3 experienced technical issues and *n* = 1 were unresponsive for >50% of trials.

Using a 2 (Color Similarity: high, low) x 5 (Training Round: 1-5) mixed ANOVA on associative memory accuracy, we found both a significant main effect of Color Similarity (*F*(1,60) = 5.26, *p* = .02, $\eta^2{}_p$ = .08) and a significant main effect of Training Round (*F*(4,240) = 239.04, *p* < .001, $\eta^2{}_p$ = .80). An independent two-tailed *t*-test revealed that accuracy was poorer for the High Similarity group (*M* = 0.58, *SD* = 0.20) than the Low Similarity group (*M* = 0.66, *SD* = 0.23; *t*(56.47) = -2.29, *p* = .03, *d* = .58) A paired *t*-test revealed that accuracy increased from Round 1 (*M* = 0.40, *SD* = 0.13) to Round 5 (*M* = 0.81, *SD* = 0.16 ; *t*(61) = 22.25, *p* < .001, *d* = 2.83) (**Figure 2A**). Using one-sample *t*-tests*,* we found that mean accuracy for each of the five rounds was greater than statistical chance, or 25% (all *t*s > 9.38, all *p*s < .001, all *d*s > 1.19).

Additionally, we ran the same 2 x 5 ANOVA on number of interference errors. Similar to memory accuracy, we found both a significant main effect of Color Similarity (*F*(1,60) = 18.05, *p* < .001, $\eta^2{}_p$ = .23) and a significant main effect of Training Round (*F*(4,240) = 38.80, *p* < .001, $\eta^2{}_p$ = .39). An independent two-tailed *t*-test revealed the High Similarity group made more interference errors (*M* = 7.72, *SD* = 3.54) than the Low Similarity Group (*M* = 5.56, *SD* = 3.43; *t*(59.71) = 4.25, *p* < .001, *d* = 1.08). A paired *t*-test revealed that interference errors decreased from Round 1 (*M* = 8.23, *SD* = 2.72) to Round 5 (*M* = 3.84, *SD* = 3.06; *t*(61) = 8.55, *p* < .001, *d* = 1.09). Overall, these findings suggest that participants successfully learned the object-face associations and experienced less interference over time. However, in relative terms, participants that studied objects of highly similar colors had poorer associative memory accuracy and experienced more interference.

Absolute color error. Next, we examined performance on the Color Memory Tests using *absolute color error*, which represents the unsigned distance between a target's remembered color and its true color (i.e., the absolute value of color distance on the color wheel). Here, we excluded *n* = 3 participants that were missing Color Memory Test data for one training round. Of these participants, *n* = 1 experienced technical issues and *n* = 2 misunderstood instructions. Color Memory Tests were administered on training rounds 1, 3, and 5.

Using a 2 (Color Similarity: high, low) x 3 (Training Round: 1, 3, 5) mixed ANOVA on absolute color error, we found only a significant main effect of Training Round (*F*(2,122) = 327.30, *p* < .001, ${\eta^2}_p$ = .84). An independent two-tailed t-test revealed that there was no significant difference in absolute color error between the High Similarity group (*M* = 42.62°, *SD* = 20.88°) and the Low Similarity group (*M* = 47.08°, *SD* = 21.40°; *t*(60.27) = -1.49, *p* = .14, *d* = .38). A two-tailed paired *t*-test revealed that absolute color error decreased between Round 1 (*M* = 66.84°, *SD* = 15.11°) and Round 5 (*M* = 29.44°, *SD* = 13.53°; *t*(62) = -19.68, *p* < .001, *d* = 2.48) (**Figure 2B**). These findings suggest that participants successfully learned the colors of the objects.

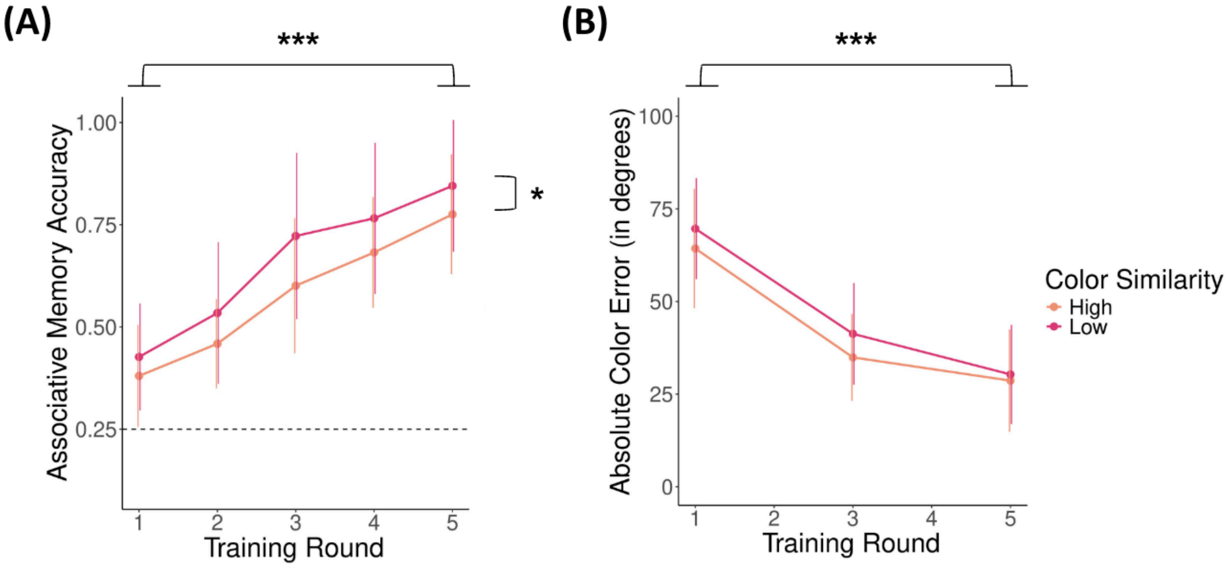

Figure 2. Participants successfully learned object-face associates and object colors over time. **(A)** Mean object-face associative memory accuracy over five training rounds, for both the High and Low Similarity groups. Mean $\pm$ SD. Dashed line represents memory performance at chance, 0.25. **(B)** Mean absolute color error (in degrees; 179° maximum) over the three training rounds that included a Color Memory Test, for both the High and Low Similarity groups. Mean $\pm$ SD. **p* < .05; ****p* < .001.

Aversive white noise bursts elicited pupil-linked arousal across all training rounds.

As an arousal manipulation check, we examined whether white noise bursts elicited significantly larger pupil dilations than neutral pure tones. Here, we excluded *n* = 23 participants total (*n* = 11 missing all eye-tracking data, *n* = 12 missing one or more rounds). Reasons for missing data are described in **Methods**.

Using a 2 (Sound: aversive, neutral) x 5 (Training Round: 1-5) within-subjects ANOVA, we investigated whether pupil dilation differed by sound type, and whether this relationship changed over time. We found a significant Sound-by-Training Round interaction effect (*F*(4,168) = 11.58, *p* < .001, $\eta^2{}_p$ = .22), such that the *degree* to which aversive white noise bursts led to larger pupil dilations varied by training round. Paired *t*-tests revealed that aversive white noise bursts led to larger pupil dilations in every training round (all *t*s > 6.51, all *p*s < .001), with the greatest evidence of arousal enhancement in Round 1 (*t*(42) = 10.9, *p* < .001) and the smallest arousal enhancement in Round 4 (*t*(42) = 6.51, *p* < .001) (**Figure 3A**).

For illustrative purposes, we replotted the main effect of Sound on pupil dilation in **Figure 3B.** Overall, pupil dilation was greater in response to aversive white noise bursts (*M* = 186.79 pixels, *SD* = 113.26) than to neutral tones (*M* = 81.07, *SD* = 84.68; t(42) = 10.39, *p* < .001, *d* = 1.29)

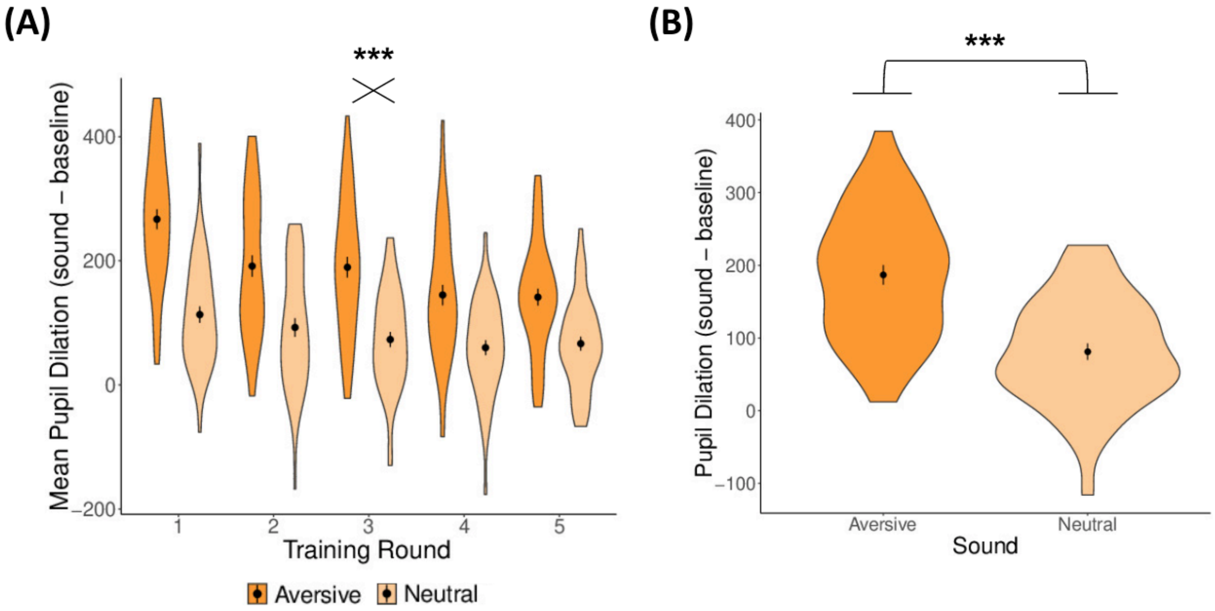

Figure 3. White noise bursts elicited significant increases in physiological arousal. (A) Mean pupil dilation to aversive white noise bursts and neutral tones over the five training rounds of the task, calculated by subtracting pupil size during baseline from pupil size during sound window. In each round, aversive noise led to greater pupil dilation than neutral sound (*p*s < .001). **(B)** Overall comparison between mean pupil dilation to aversive white noise bursts and to neutral tones, Mean ± SE. ****p* < .001.

Higher color similarity, but not aversive noise, triggered attraction effects in color memory.

Next, we conducted 2 (Sound: aversive, neutral) x 2 (Color Similarity: high, low) mixed ANOVAs to test our main hypothesis that arousal elicits greater repulsion effects for highly similar associations.

Percentage of away responses for color memory judgments. We first examined the effect of Sound and Color Similarity on the percentage of away responses. The 2 x 2 ANOVA showed a significant main effect of Color Similarity (*F*(1,64) = 2.12, *p* < .001, $\eta^2{}_p$ = .25). There was no significant main effect of Sound (*F*(1,64) = .002, *p* = .97, $\eta^2{}_p$ = 0). Contrary to our expectations, one-sample *t*-tests indicated that the High Similarity group demonstrated significant memory attraction, for both pairmates that included an aversive association (*M* =

37% away responses, *SD* = 15%; *t*(32) = -5.44, *p* < .001, *d* = .95) and those that were neutral (*M* = 37%, *SD* = 13%; *t*(32) = -6.15, *p* < .001, *d* = 1.07). That is, participants remembered target objects as being closer to their pairmates' colors than they actually were. In contrast, the Low Similarity group did not show significant memory bias, for neither pairmates that included an aversive association (*M* = 48%, *SD* = 14%; *t*(32) = -1.22, *p* = .23, *d* = .21) nor those that were neutral (*M* = 48%, *SD* = 13%; *t*(32) = -1.23, *p* = .23, *d* = .21) (**Figure 4A**).

Color distance. We also performed the same type of mixed ANOVA analysis on the second outcome measure, color distance, which generally yielded a similar pattern of results. There was a significant main effect of Color Similarity (*F*(1,64) = 4.46, *p* = .04, ${\eta^2}_p$ = .07), and no main effect of Sound (*F*(1,64) = 0.45, *p* = .51, ${\eta^2}_p$ = .01) on color distance memory. One sample *t*-tests indicated that the High Similarity group demonstrated significant memory attraction, for both pairmates that included an aversive association (*M* = -6.64°, *SD* = 10.62°; *t*(32) = -4.19, *p* < .001, *d* = .73) and those that were neutral (*M* = -7.30°, *SD* = 11.67°; *t*(32) = -4.74, *p* < .001, *d* = .83). The Low Similarity group also demonstrated significant memory attraction for pairmates that included an aversive association (*M* = -4.50°, *SD* = 11.04°; *t* = -2.62, *p* = 0.013, *d* = .46). However, this group did not show significant memory bias for pairmates that were neutral (*M* = -1.87°, *SD* = 11.52°; *t*(32) = -1.09, *p* = 0.28, *d* = .19) (**Figure 4B**).

Overall, we found that the High Similarity group showed significant memory attraction rather than the expected memory repulsion effect. This memory bias was not influenced by the type of sound that was heard (i.e., an aversive noise or neutral tone). There was weaker evidence for memory bias in the Low Similarity group: when examining percentage of away responses, there were no significant biases; when examining color distance, there was only significant bias towards attraction for pairmates that included an aversive association.

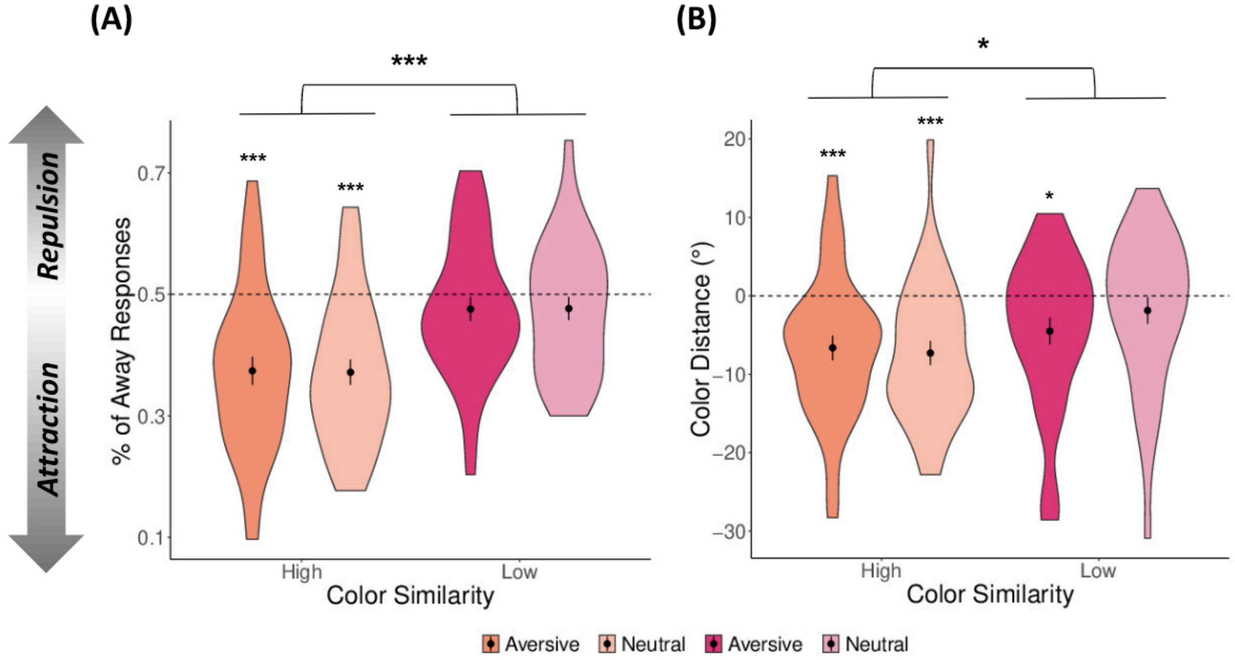

Figure 4. Participants remembered highly similar objects as being closer to their pairmate's color than they actually were. (A) Mean percentage of away responses for target objects, depending on color similarity and sound during encoding. Mean ± SE. Dashed line represents no memory bias (50% of away responses), and asterisks above each violin represent significant differences from this value in each condition. **(B)** Mean color distance (in degrees) for target objects, depending on color similarity and sound during encoding. Mean ± SE. Dashed line represents no memory bias (0 degrees), and asterisks above each violin represent significant differences from this value in each condition. ****p* < .001; **p* < .05.

Aversive noise led to greater memory interference for highly overlapping associations.

To test whether color similarity and sound type influenced memory interference, we conducted a 2 (Sound: aversive, neutral) x 2 (Color Similarity: high, low) mixed ANOVA on the final number of interference errors. We found a significant Sound-by-Color Similarity interaction effect ($F(1,64) = 6.77$, $p = .011$, $\eta^2_p = .10$) on memory interference, such that aversive noise selectively led to significantly higher interference for highly similar memories. Follow-up paired *t*-tests clarified this effect, with the High Similarity group making significantly more interference errors for pairmates that included an aversive association (*M* = 1.45 errors; *SD* = 1.15) than pairmates that were neutral (*M* = 0.85; *SD* = 0.83; *t*(32) = 3.12, *p* = .004, *d* = .59). In contrast, there was no significant difference between aversive-related (*M* = 0.91, *SD* = 1.28) versus neutral-related interference in the Low Similarity group (*M* = 1.09, *SD* = 1.91; *t*(32) = -0.78, *p* = .44, *d* = -.10) (**Figure 5**).

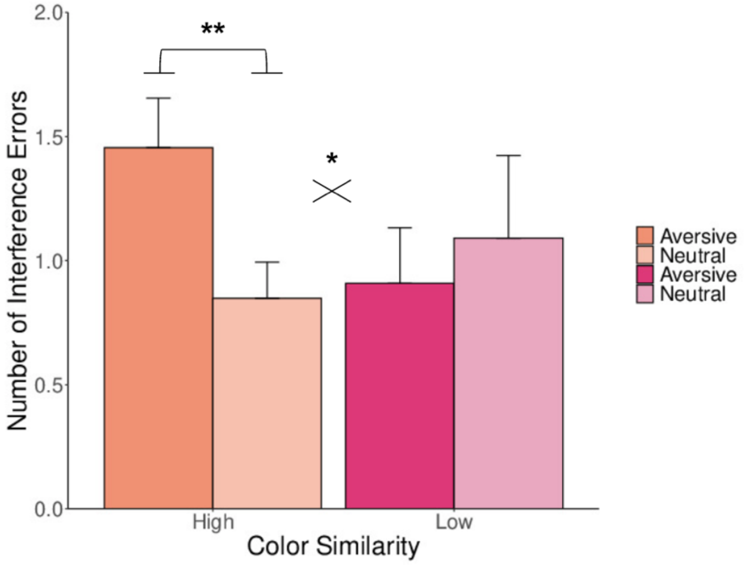

Figure 5. Arousal intensified memory interference when object-face pairmates were highly overlapping. Plot shows mean number of interference errors made on the final Associative Memory Test by color similarity and sound type. Error bars represent SEM. X = significant interaction effect; ***p* < .01; **p* < .05.

Higher pupil-linked arousal across the course of learning was associated with greater attraction effects between highly similar memories.

Beyond manipulating the nature of the arousal-eliciting stimulus (i.e., by presenting an aversive white noise burst or neutral tone), we also investigated whether the actual amount of physiological arousal experienced by the participant was related to the degree of color memory bias. That is, arousal may relate to mnemonic discrimination irrespective of whether it was elicited by something relatively neutral or aversive. To test this idea, we correlated the cumulative pupil measure with the degree of color memory bias across participants using a Spearman rank coefficient correlation. Color distance for each trial on the final Color Memory Tests was averaged to compute a mean color distance score for each participant. Color distance was selected as the preferred outcome measure over the percentage of away responses due to its ability to capture greater nuance in memory biases. We excluded *n* = 23 participants total (*n*

= 11 missing all eye-tracking data, *n* = 12 missing one or more rounds). Reasons for missing data are described in **Methods**.

We found a significant negative correlation between cumulative pupil dilation and color distance in the High Similarity group, such that greater pupillary arousal was associated with greater memory attraction (*ρ* = -.51; *p* = .017) (**Figure 6A**). In contrast, there was no significant pupil-color memory association in the Low Similarity group (*ρ* = .21; *p* = .34) (**Figure 6B**). The interaction between the two groups' pupil-memory slopes was also not statistically significant (*t* ratio = -1.21, *p* = .23).

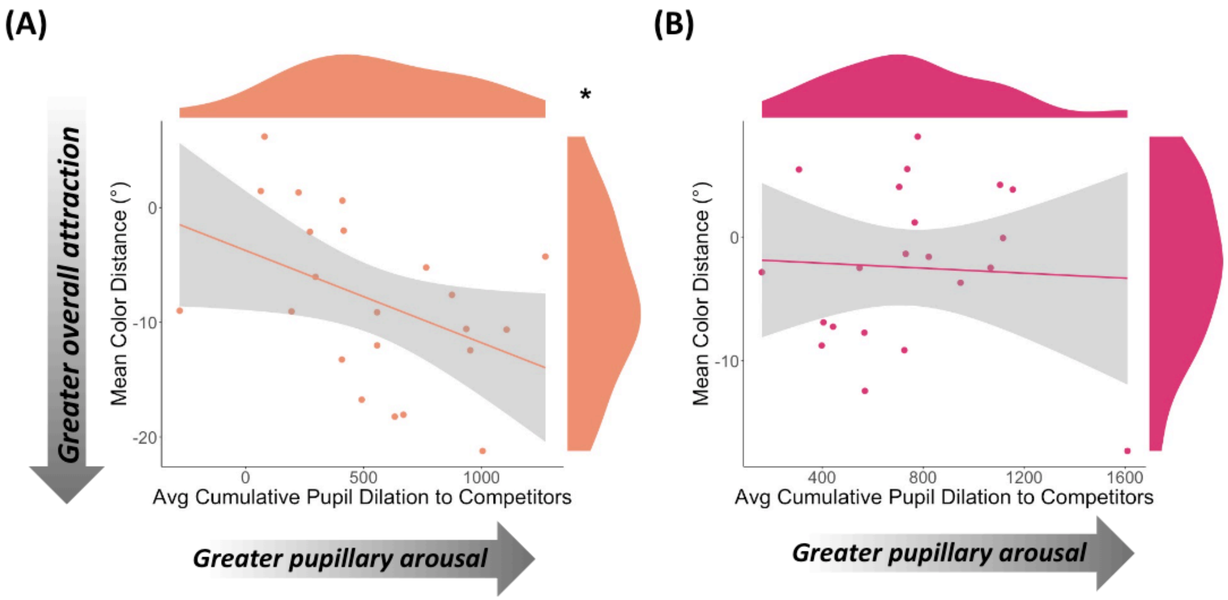

Figure 6. Individual differences in cumulative pupil dilation across learning were related to the amount of memory attraction between highly similar memories. Plots show average cumulative pupil dilation to competitor object versus mean color distance in the final Color Memory Test, for **(A)** the High Similarity group and **(B)** Low Similarity group. Trendlines and 95% confidence intervals shown. *$p < .05$.

Greater memory attraction was related to more memory interference at the end of training.

To test whether biases in color memory were adaptive – that is, whether they were associated with lower memory interference – we conducted Spearman correlations between color memory bias and the number of interference errors in the Associative Memory Test.

First, we focused on performance during the final Associative Memory Test. We tested linear correlations between the number of interference errors on this associative memory test with both measures of color memory bias. For percentage of away responses, we found that the relationship between color memory bias and associative memory interference was in the expected negative direction, but not statistically significant (ρ = -.17; *p* = .18). For color distance, we found a similar, non-significant relationship (ρ = -.21; *p* = .10).

Next, we ran exploratory correlation analyses that focused on the last training round (Round 5). For percentage of away responses, we found a significant negative relationship between color memory bias and associative memory interference (ρ = -.36; *p* = .003), such that greater memory attraction was related to greater interference (**Figure 7A**). For color distance, we found a similar, significant relationship (ρ = -.37; *p* = .002) (**Figure 7B**). This distortion-interference association appeared to be mostly driven by effects in the Low Similarity group (percentage of away responses: ρ = -.35; *p* = .048; color distance: ρ = -.39; *p* = .025), as the correlation was not statistically significant in the High Similarity group (percentage of away responses: ρ = -.13; *p* = .46; color distance: ρ = -.12; *p* = .50).

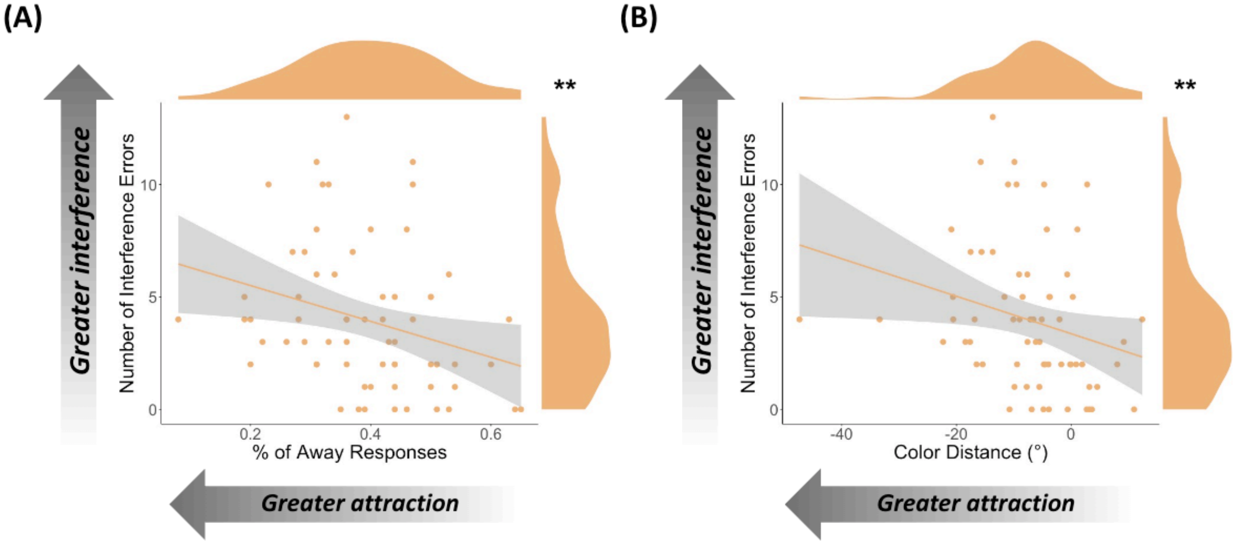

Figure 7. Greater memory attraction was related to greater memory interference across similarity levels at the end of training. (A) Mean percentage of away responses versus number of interference errors in Round 5. **(B)** Mean color distance versus number of interference errors in Round 5. Trendlines and 95% confidence intervals also shown. ***p* < .01.

Higher self-reported trait anxiety was related to greater memory attraction towards aversive memories.

In a final set of individual differences analyses, we used Spearman correlations to test whether scores on our three clinical measures of interest – state anxiety, trait anxiety, and depression – were related to color memory bias for pairmates that included an aversive association. We did not use multiple regression for these analyses because there was moderate multicollinearity between the *z*-scored clinical measures (state anxiety: VIF = 1.90; trait anxiety: VIF = 3.22; depression: VIF = 2.86). This multicollinearity makes it difficult to disentangle which clinical measures best predict color memory bias. We excluded the state anxiety score (STAI-S) for *n* = 1 participant, who did not fully answer the questionnaire.

Trait anxiety scores (*ρ* = -.26; *p* = .038) showed a significant negative correlation with aversive-related color memory bias, such that higher trait anxiety scores were associated with greater memory attraction for pairmates that included an aversive association (**Figure 8A**). Neither state anxiety scores (*ρ* = -.15, *p* = .22) nor depression scores (*ρ* = -.22; *p* = .076) showed a significant relationship with aversive-related color memory bias.

Additionally, we tested whether our three clinical scores were related to the difference in color memory bias for pairmates that included an aversive association versus those that did not. This clinically relevant measure – which we will refer to as *color memory bias difference score* – captured the degree to which participants exhibited selective memory attraction towards events encoded under aversive conditions compared to neutral conditions.

Trait anxiety scores showed a significant negative correlation with color memory bias difference scores, such that higher trait anxiety scores were associated with more selective memory attraction for pairmates that included an aversive association (*ρ* = -.31; *p* = .012) (**Figure 8B**). Neither state anxiety scores (*ρ* = -.11, *p* = .40) nor depression scores (*ρ* = -.18, *p* = .15) showed a significant relationship with color memory bias difference scores.

Regarding pupil-linked responses, no clinical measures were significantly correlated with cumulative pupil dilation across the learning rounds (*ps* > .05). Additionally, no clinical measures were significantly correlated with the difference in cumulative pupil dilation between aversive white noise bursts and neutral tones (*ps* > .05). In summary, higher trait anxiety was selectively related to greater attraction between neutral memories and overlapping aversive memories.

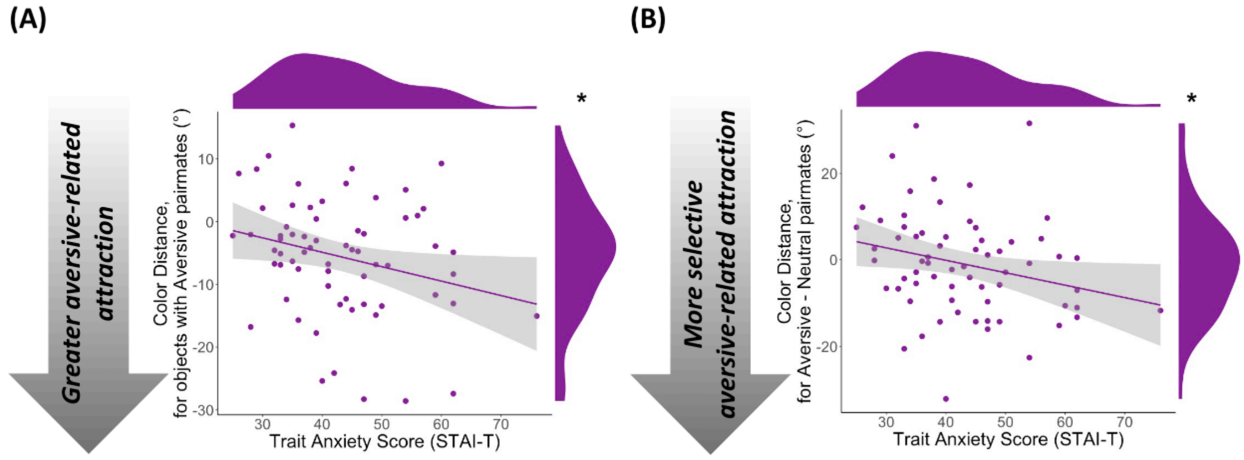

Figure 8. Higher trait anxiety was associated with greater memory attraction effects between neutral target objects and their aversive pairmates. (A) Trait anxiety (STAI-T) scores were negatively associated with color memory bias for neutral targets with aversive pairmates. Trendline and 95% confidence interval shown. **(B)** Trait anxiety (STAI-T) scores were also negatively associated with color memory bias for neutral targets with aversive pairmates versus neutral pairmates. Trendline and 95% confidence interval shown. **p* < .05.

DISCUSSION

Memory repulsion is thought to be an adaptive way of resolving interference between highly similar memories. By exaggerating slight perceptual differences between overlapping representations in memory, repulsion helps reduce competition and allow us to retain memories of similar yet distinct events. Most work to date has demonstrated how repulsion effects emerge for competing neutral associations. By comparison, less is known about how emotional arousal might influence these memory distortions and their relation to associative memory performance, despite a large body of research indicating that emotion often enhances interference. To test this idea, we used aversive white noise bursts to induce arousal while participants repeatedly studied competing object-face associations. Our findings showed that higher pupil-linked arousal responses were related to greater attraction – not repulsion – effects between overlapping memories, such that the colors of highly similar memories were remembered as being more similar to each other than they actually were. After repeated rounds of associative learning, greater memory attraction was also associated with higher interference, and interference was particularly pronounced for pairmates that included an aversive association. Thus, our results demonstrate that very similar experiences are remembered as being even more alike by participants that experience high arousal.

Importantly, we found that the nature of the arousal-eliciting stimulus (i.e., an aversive burst of white noise or a neutral tone) did not significantly influence distortions in color memory. In general, the tested 'target' object-face associations and their highly similar pairmates were remembered as being more alike than they actually were. Yet, exposure to an aversive burst of white noise did not lead to additional repulsion or attraction effects in memory. Instead, it appeared that the actual physiological arousal induced by the sounds – rather than their emotional or aversive properties – were related to changes in how similar events were remembered.

Indeed, participants who showed greater sound-associated pupil dilation across time showed stronger attraction between highly similar memories. We chose to sum participants' pupil responses across all training rounds, given our interest in assessing the cumulative impact of physiological arousal over the course of learning. We believe this approach is preferable to alternative measures, because it is sensible that arousal should alter encoding processes whenever it is elicited – in this case, during every training round. Yet, it is possible that arousal effects are time and learning-dependent, whereby arousal has differing levels of impact on distinct stages of learning and differentiation. For example, because pupil responses were strongest earlier on, its effects on memory may only be evident during the first rounds of learning. However, given that memory repulsion is a gradual effect that emerges only over repeated rounds of learning, we reasoned that arousal-related modulation of distortions occur across the entire associative learning period, particularly since our task involves fewer training rounds than other repulsion studies (Chanales et al., 2021; Drascher & Kuhl, 2022). Nevertheless, other approaches to studying interactions between pupil-linked arousal and repeated rounds of learning could uncover discrete timepoints when arousal is more important for influencing potential biases in memory.

Given our unexpected finding that greater pupil-linked arousal was associated with greater memory attraction effects, it is important to consider whether this relationship might actually be adaptive for learning and behavior. One possibility is that remembering similar arousing and neutral events as being more alike than they actually are might promote useful generalizations of fear or salience. Although this 'blending' of memories might be unhelpful for distinguishing minute features of each environment, it could protect an individual from situations that closely resemble a threatening past experience. This idea is consistent with work in human fear conditioning demonstrating that aversive stimuli can reduce discrimination between low-level perceptual features (Resnik, Sobel, & Paz, 2011). Further, it has been shown that greater physiological responses to arousing stimuli are linearly related to poorer discrimination (Resnik, Sobel, & Paz, 2011). This converging evidence suggests that higher perceptual discrimination thresholds could promote the adaptive generalization of fear to similar stimuli (Dunsmoor & Paz, 2015; Resnik, Sobel, & Paz, 2011).

However, it is less straightforward to reconcile this hypothesis with existing evidence from the pattern separation and mnemonic discrimination literatures. In particular, one behavioral study showed that higher arousal levels, as indexed by salivary alpha-amylase, was associated with greater discrimination of similar memories (Segal et al., 2012; see also Szőllősi & Racsmány, 2020). Similarly, previous neuroimaging work has shown that the pattern separation signal in DG/CA3, which is important for differentiating highly similar inputs and decreasing memory interference (Leal et al., 2014; McClelland et al., 1995; Yassa & Stark, 2011), is amplified when correctly discriminating negative items from other similar, negative items (Leal et al., 2014).

At first glance, these findings appear to conflict with our finding that pupil-linked arousal was related to greater memory attraction rather than repulsion. However, previous paradigms differ from the current study in several important ways. First, in prior neuroimaging work, participants had to discriminate between overlapping emotional memories (Leal et al., 2014). Importantly, reducing interference between similarly emotional items might require different neural processes than reducing interference between emotional items and similar neutral items. In particular, discriminating between two threatening events is likely less useful than discriminating between a threatening event and a safe event, which uniquely prevents the overgeneralization of fear. While pattern separation-linked DG/CA3 activation is enhanced when accurately discriminating two emotional items (Leal et al., 2014), encountering a neutral pairmate at the time of retrieval may affect the manner in which salience-relevant networks – such as the LC-NE system – bias memory discrimination processes in hippocampus. Future neuroimaging work could examine whether these different types of valence and arousal-related discriminations are supported by separate neural pathways.

Another possibility is that high physiological arousal responses may have disproportionally enhanced memory for the competitors. Because competitor object-face associations were specifically modulated by aversive noise – which also induced significantly more pupil-linked arousal than neutral tones – it is possible that participants formed much stronger and perceptually-detailed memories for those competitors. Because the competitors would stand out in memory, they would also be more likely to interfere with retrieval of their neutral pairmates. Consistent with this possibility, we found that interference was higher for aversive competitor memories that were highly overlapping. This result aligns with the idea that arousing associations tend to elicit more memory interference than neutral associations, perhaps disrupting mnemonic differentiation (see Mather, 2007; Mather & Knight, 2008). Due to methodological limitations, we were unable to explicitly test memory for the competitor object-face associations. However, future research could test this interference hypothesis by testing memory for all studied associations and contingencies between their memory distortions and memory accuracy.

One potential way to increase the likelihood of observing memory repulsion, rather than memory attraction, is to ensure there is sufficient behavioral demand for this effect. Memory repulsion, or the exaggeration of subtle differences between similar memories, is most beneficial when one needs to discriminate and recall specific details about similar events (Chanales et al., 2021). As mentioned previously, memory tests in the current study were only administered for one of the object-face associations. Therefore, since participants were only asked to remember the color and face of one of the overlapping associations, it remains unknown whether more direct competition may be required. To encourage memory repulsion, future studies could pit the overlapping associations against each other more explicitly by displaying them simultaneously during the color memory test (e.g., with two color wheels on the screen at the same time).

Alternatively, a future paradigm could reward participants for accurately discriminating between the two interfering objects or threaten punishment when participants make inaccurate judgments. One caveat to this approach is that arousal would also likely be induced during the retrieval phase, which would confound interpretations about when arousal is affecting memory separation.

Third, researchers could manipulate the semantic nature of the arousal-eliciting stimulus that is presented alongside the to-be-learned object-face fairs. For example, participants could be shown a neutral or negative image before the competitor object-face association (e.g., an image of an ordinary car or an image of a car accident). It is possible that participants could semantically integrate or unitize this image with the competitor object-face association, enhancing its distinctiveness. Whether arousal would enhance or impair inter-item binding through unitization strategies is slightly unclear, as arousal can sometimes benefit and other times impede relational processing (Murray and Kensinger, 2013). Nevertheless, this approach could provide interesting insights into how arousal processes influence the binding of discrete memories and the ability to maintain separate memory representations of distinct events.

Finally, we were also interested in testing for individual differences in the relationship between symptoms of affective disorders and memory biases for pairmates that included an aversive association. As predicted, we found that individuals who self-reported higher symptoms of trait anxiety showed greater memory attraction effects for neutral events towards their aversive pairmates, consistent with findings that individuals with anxiety disorders (Lissek et al., 2005; 2010; 2014) and trait anxiety (Sep et al., 2019) show greater fear generalization compared to less anxious individuals. Importantly, this association was also selective to conditions that included an aversive association. In contrast, we did not find a significant correlation between depression symptoms and memory bias, consistent with the observation that depression is typically not associated with fear generalization (Park, Lee, & Lee, 2018; Wurst et al., 2021). Our findings provide specific evidence that people with high trait anxiety may generalize memories of aversive events to encompass similar neutral events, blurring the line between fear and safety.

These findings could help inform future neuroimaging studies examining arousal and memory interference in individuals with anxiety disorders. Of particular interest are prefrontal networks, which have been linked to memory differentiation (Nash et al., 2021) and show impaired function in anxiety disorders (for review, see e.g., Kenwood, Kalin, & Barbas, 2021). In particular, one human fear conditioning study found that increased threat generalization was related to lower performance on a memory differentiation task. Poorer memory differentiation, in turn, was associated with lower activation in the subcallosal cortex, a prefrontal region that may play a role in threat appraisal and related behavior (Etkin et al., 2015; Fullana et al., 2016; Greenberg et al., 2013; Lange et al., 2017). Future work could investigate whether prefrontal cortical networks influence memory interference under arousing conditions, and how these dynamics might change in anxiety disorders. In particular, the subcallosal cortex is well-positioned to help shape the interaction of memory and arousal, especially considering its rich anatomical connectivity with the hippocampus (see Joyce & Barbas, 2017).

In summary, the current study demonstrated that individual differences in pupil-linked arousal and trait anxiety were associated with greater attraction between overlapping memories. These findings align with suggestions from the fear conditioning literature that threat generalization across similar stimuli may promote adaptive behavior, allowing individuals to respond efficiently to situations that likely share comparable risks (Resnik, Sobel, & Paz, 2011). However, this mnemonic blending effect may not be as beneficial for wellbeing in anxious individuals, where neutral events resembling arousing events might become excessively integrated in memory and lead to the overgeneralization of fear. Indeed, such a mechanism that is generally adaptive can contribute to pathological behavior in excess (Asok, Kandel, & Rayman, 2019; Dunsmoor & Paz, 2015; in general, see Burnell, Rasmussen & Gary, 2020). Additional research should aim to identify the boundary conditions for the relationship between memory biases and arousal, such as through altering task demands and the emotional nature of the arousal-eliciting stimuli. Along with the current findings, identifying these parameters will provide valuable insights into potential strategies for reducing the disruptive effects of aversive events on the lives and wellbeing of anxious individuals.

DECLARATIONS

Funding: This project was funded by UCLA faculty retention start-up funds to D.C.

Conflicts of interest: The authors declare no conflicts of interest.

Ethics approval: This study was approved by the University of California, Los Angeles Institutional Review Board (IRB#23-000066).

Consent to participate: All participants provided verbal informed consent approved by the University of California, Los Angeles Institutional Review Board.

Availability of data and materials: The data and materials for the experiment will be made publicly available on the OSF account of E.M. (osf.io/yt6hm) upon acceptance of this manuscript.

Code availability: Experiment code was modified from a previous study by Chaneles et al. (2021) (https://osf.io/d4pmg/) and will be made publicly available on the OSF account of E.M. (osf.io/yt6hm) upon acceptance of this manuscript.

REFERENCES

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: memory retrieval as a model case. *Psychological Review*, *102*(1), 68.

Angelova, A., Abu-Mostafam, Y., & Perona, P. (2005, June). Pruning training sets for learning of object categories. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, pp. 494-501). IEEE.

Asok, A., Kandel, E. R., & Rayman, J. B. (2019). The neurobiology of fear generalization. *Frontiers in Behavioral Neuroscience*, *12*, 329.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403-450.

Audacity Team. (2021). *Audacity*. Retrieved from https://www.audacityteam.org/

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology*, *58*(2), 97.

Barnier, A., Hung, L., & Conway, M. (2004). Retrieval-induced forgetting of emotional and unemotional autobiographical memories. *Cognition and Emotion*, *18*(4), 457-477.

Beck, A. T., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. J. A. G. P. (1961). Beck depression inventory (BDI). *Archives of General Psychiatry*, *4*(6), 561-571.

Besnard, A., & Sahay, A. (2016). Adult hippocampal neurogenesis, fear generalization, and stress. *Neuropsychopharmacology*, *41*(1), 24-44.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as bound units: independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General*, *142*(3), 791.

Burnell, R., Rasmussen, A. S., & Garry, M. (2020). Negative memories serve functions in both adaptive and maladaptive ways. *Memory*, *28*(4), 494-505.

Chanales, A. J., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among spatial memories triggers repulsion of hippocampal representations. *Current Biology*, *27*(15), 2307-2317.

Chanales, A. J., Tremblay-McGaw, A. G., Drascher, M. L., & Kuhl, B. A. (2021). Adaptive repulsion of long-term memory representations is triggered by event similarity. *Psychological Science*, *32*(5), 705-720.

Clewett, D. V., Huang, R., Velasco, R., Lee, T. H., & Mather, M. (2018). Locus coeruleus activity strengthens prioritized memories under arousal. *Journal of Neuroscience*, *38*(6), 1558-1574.

Dimsdale-Zucker, H. R., Ritchey, M., Ekstrom, A. D., Yonelinas, A. P., & Ranganath, C. (2018). CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields. *Nature Communications*, *9*(1), 294.

Drascher, M. L., & Kuhl, B. A. (2022). Long-term memory interference is resolved via repulsion and precision along diagnostic memory dimensions. *Psychonomic Bulletin & Review*, *29*(5), 1898-1912.

Dunsmoor, J. E., & Paz, R. (2015). Fear generalization and anxiety: Behavioral and neural mechanisms. *Biological Psychiatry*, *78*(5), 336-343.

Etkin, A., Büchel, C., & Gross, J. J. (2015). The neural bases of emotion regulation. *Nature Reviews Neuroscience,* 16(11), 693-700.

Favila, S. E., Chanales, A. J., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, *7*(1), 11066.

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2016). Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry*, *21*(4), 500-508.

Greenberg, T., Carlson, J. M., Cha, J., Hajcak, G., & Mujica-Parodi, L. R. (2013). Neural reactivity tracks fear generalization gradients. *Biological Psychology*, *92*(1), 2-8.

Grella, S. L., & Donaldson, T. N. (2024). Contextual memory engrams, and the neuromodulatory influence of the locus coeruleus. *Frontiers in Molecular Neuroscience*, *17*, 1342622.

Grella, S. L., Neil, J. M., Edison, H. T., Strong, V. D., Odintsova, I. V., Walling, S. G., ... & Harley, C. W. (2019). Locus coeruleus phasic, but not tonic, activation initiates global remapping in a familiar environment. *Journal of Neuroscience*, *39*(3), 445-455.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, *5*(9), 394-400.

Hamm, A. O., Greenwald, M. K., Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1991). The fear potentiated startle effect: Blink reflex modulation as a result of classical aversive conditioning. *Integrative Physiological and Behavioral Science*, *26*(2), 119-126.

Harley, C. W. (2007). Norepinephrine and the dentate gyrus. *Progress in Brain Research*, *163*, 299-318.

Hensley, C. J., Otani, H., & Knoll, A. R. (2019). Reducing negative emotional memories by retroactive interference. *Cognition and Emotion*, *33*(4), 801-815.

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.

Huang, R., & Clewett, D. The locus coeruleus: where cognitive and emotional processing meet the eye. In forthcoming book, *Modern Pupillometry.*

Hulbert, J. C., & Norman, K. A. (2015). Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cerebral Cortex*, *25*(10), 3994-4008.

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*(1), 221-234.

Joyce, M. K. P., & Barbas, H. (2018). Cortical connections position primate area 25 as a keystone for interoception, emotion, and memory. *Journal of Neuroscience*, *38*(7), 1677-1698.

Kaplan, R. L., Van Damme, I., Levine, L. J., & Loftus, E. F. (2016). Emotion and false memory. *Emotion Review*, *8*(1), 8-13.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *Journal of Memory and Language*, *54*(1), 99-112.

Kenwood, M. M., Kalin, N. H., & Barbas, H. (2022). The prefrontal cortex, pathological anxiety, and anxiety disorders. *Neuropsychopharmacology*, *47*(1), 260-275.

Kheirbek, M. A., Klemenhagen, K. C., Sahay, A., & Hen, R. (2012). Neurogenesis and generalization: a new approach to stratify and treat anxiety disorders. *Nature Neuroscience*, *15*(12), 1613-1620.

Kirwan, C. B., & Stark, C. E. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory*, *14*(9), 625-633.

Krypotos, A. M., Effting, M., Kindt, M., & Beckers, T. (2015). Avoidance learning: a review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience*, *9*, 189.

Kunz, L., Wang, L., Lachner-Piza, D., Zhang, H., Brandt, A., Dümpelmann, M., ... & Axmacher, N. (2019). Hippocampal theta phases organize the reactivation of large-scale electrophysiological representations during goal-directed navigation. *Science Advances*, *5*(7), eaav8192.

Kyle, C. T., Smuda, D. N., Hassan, A. S., & Ekstrom, A. D. (2015). Roles of human hippocampal subfields in retrieval of spatial and temporal context. *Behavioural Brain Research*, *278*, 549-558.

Lange, I., Goossens, L., Michielse, S., Bakker, J., Lissek, S., Papalini, S., ... & Schruers, K. (2017).

Behavioral pattern separation and its link to the neural mechanisms of fear generalization. *Social Cognitive and Affective Neuroscience*, *12*(11), 1720-1729.

Leal, S. L., Tighe, S. K., Jones, C. K., & Yassa, M. A. (2014). Pattern separation of emotional information in hippocampal dentate and CA3. *Hippocampus*, *24*(9), 1146-1155.

Lissek, S., Kaczkurkin, A. N., Rabin, S., Geraci, M., Pine, D. S., & Grillon, C. (2014). Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. *Biological Psychiatry*, *75*(11), 909-915.

Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., & Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behaviour Research and Therapy*, *43*(11), 1391-1424.

Lissek, S., Rabin, S., Heller, R. E., Lukenbaugh, D., Geraci, M., Pine, D. S., & Grillon, C. (2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. *American Journal of Psychiatry*, *167*(1), 47-55.

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, *2*(1), 33-52.

Mather, M. (2009). When emotion intensifies memory interference. *Psychology of Learning and Motivation*, *51*, 101-120.

Mather, M., Huang, R., Clewett, D., Nielsen, S. E., Velasco, R., Tu, K., ... & Kennedy, B. L. (2020). Isometric exercise facilitates attention to salient events in women via the noradrenergic system. *Neuroimage*, *210*, 116560.

Mather, M., & Knight, M. (2008). The emotional harbinger effect: poor context memory for cues that previously predicted something arousing. *Emotion*, *8*(6), 850.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419.

Mensink, G. J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, *95*(4), 434.

Moore, S. A., & Zoellner, L. A. (2007). Overgeneral autobiographical memory and traumatic events: an evaluative review. *Psychological Bulletin*, *133*(3), 419.

Müller, G. E., & Pilzecker, A. (1900). *Experimentelle beiträge zur lehre vom gedächtniss* (Vol. 1). JA Barth.

Murphy, P. R., O'connell, R. G., O'sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, *35*(8), 4140-4154.

Murray, B. D., & Kensinger, E. A. (2013). A review of the neural and behavioral consequences for unitizing emotional and neutral information. *Frontiers in Behavioral Neuroscience*, *7*, 42.

Nash, M. I., Hodges, C. B., Muncy, N. M., & Kirwan, C. B. (2021). Pattern separation beyond the hippocampus: A high-resolution whole-brain investigation of mnemonic discrimination in healthy adults. *Hippocampus*, *31*(4), 408-421.

Nashiro, K., Sakaki, M., Huffman, D., & Mather, M. (2013). Both younger and older adults have difficulty updating emotional memories. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *68*(2), 224-227.

Nefian, A. V., Khosravi, M., & Hayes III, M. (1997). Real-time human face detection from uncontrolled environments. *SPIE Visual Communications on Image Processing*.

Novak, D. L., & Mather, M. (2009). The tenacious nature of memory binding for arousing negative items. *Memory & Cognition*, *37*, 945-952.

Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review*, *56*(3), 132.

Park, D., Lee, H. J., & Lee, S. H. (2018). Generalization of conscious fear is positively correlated with anxiety, but not with depression. *Experimental Neurobiology*, *27*(1), 34.

Peri, T., Ben-Shakhar, G., Orr, S. P., & Shalev, A. Y. (2000). Psychophysiologic assessment of aversive conditioning in posttraumatic stress disorder. *Biological Psychiatry*, *47*(6), 512-519.

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, *7*(1), 13289.

Reisberg, D., & Heuer, F. (1992). Remembering the details of emotional events.

Resnik, J., Sobel, N., & Paz, R. (2011). Auditory aversive learning increases discrimination thresholds. *Nature Neuroscience*, *14*(6), 791-796.

Ritvo, V. J., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic plasticity: how memory retrieval drives learning. *Trends in Cognitive Sciences*, *23*(9), 726-742.

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, *10*(3), 211-223.

Schulkind, M. D., & Woldorf, G. M. (2005). Emotional organization of autobiographical memory. *Memory & Cognition*, *33*, 1025-1035.

Segal, S. K., Stark, S. M., Kattan, D., Stark, C. E., & Yassa, M. A. (2012). Norepinephrine-mediated emotional arousal facilitates subsequent pattern separation. *Neurobiology of Learning and Memory*, *97*(4), 465-469.

Sep, M. S., Steenmeijer, A., & Kennis, M. (2019). The relation between anxious personality traits and fear generalization in healthy subjects: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *107*, 320-328.

Sison, J. A. G., & Mather, M. (2007). Does remembering emotional items impair recall of same-emotion items?. *Psychonomic Bulletin & Review*, *14*, 282-287.

Spielberger, C. D. (1983). State-trait anxiety inventory for adults. In *Manual for the State-Trait Anxiety Inventory STAI.*

SR Research Ltd. (2017). *SR Research*. Retrieved from https://www.sr-research.com/.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2008). Accelerated relearning after retrieval-induced forgetting: the benefit of being forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 230.

Szőllősi, Á., & Racsmány, M. (2020). Enhanced mnemonic discrimination for emotional memories: the role of arousal in interference resolution. *Memory & Cognition*, *48*, 1032-1045.

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, *14*(5), 455-461.

The GIMP Development Team. (2019). *GIMP*. Retrieved from https://www.gimp.org

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*(1), 49.

Underwood, B. J., & Postman, L. (1960). Extraexperimental sources of interference in forgetting. *Psychological Review*, *67*(2), 73.

Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in the reward/effort trade-off: a direct electrophysiological comparison in behaving monkeys. *Journal of Neuroscience*, *35*(20), 7866-7877.

Wang, P., Baker, L. A., Gao, Y., Raine, A., & Lozano, D. I. (2012). Psychopathic traits and physiological responses to aversive stimuli in children aged 9–11 years. *Journal of Abnormal Child Psychology*, *40*, 759-769.

Wanjia, G., Favila, S. E., Kim, G., Molitor, R. J., & Kuhl, B. A. (2021). Abrupt hippocampal remapping signals resolution of memory interference. *Nature Communications*, *12*(1), 4816.

Weber, M. (2022). Caltech Face Dataset 1999 (1.0) [Data set]. CaltechDATA.

Williams, J. M. G., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin*, *133*(1), 122.

Williams, S. E., Ford, J. H., & Kensinger, E. A. (2022). The power of negative and positive episodic memories. *Cognitive, Affective, & Behavioral Neuroscience*, *22*(5), 869-903.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, *55*, 235-269.

Wurst, C., Schiele, M. A., Stonawski, S., Weiß, C., Nitschke, F., Hommers, L., ... & Menke, A. (2021). Impaired fear learning and extinction, but not generalization, in anxious and non-anxious depression. *Journal of Psychiatric Research*, *135*, 294-301.

Xue, G. (2022). From remembering to reconstruction: The transformative neural representation of episodic memory. *Progress in Neurobiology*, 102351.

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515-525.

Zhao, Y., Chanales, A. J., & Kuhl, B. A. (2021). Adaptive memory distortions are predicted by feature representations in parietal cortex. *Journal of Neuroscience*, *41*(13), 3014-3024.