Retrieval practice prevents stress-induced inference impairment by restoring rapid memory reactivation

Abbreviated title: Retrieval Rescues Stressed Inference

Jinpeng Guo (郭金朋)[1,2], Ruixin Chen (陈瑞欣)[1,2], Qi Zhao (赵琦)[1,2], Xiaojun Sun (孙晓军)[1,2*], Wei Liu (刘威)[1,2*]

1. Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan, China
2. Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology, Central China Normal University, Wuhan, China

Correspondence:
Prof. Xiaojun Sun (sunxiaojun@ccnu.edu.cn), School of Psychology, Central China Normal University
Dr. Wei Liu (weiliu1991@ccnu.edu.cn), School of Psychology, Central China Normal University
Address: Nanhu Complex Building, No 152 Luoyu Road, Wuhan. Postal Code: 430079

- **Number of pages:** 59
- **Number of figures, tables, multimedia, and 3D models (separately):** figures-Main Text (7); figures-Supplymental Materials (6); tables (0), multimedia (0), and 3D models (0)
- **Number of words for abstract, introduction, and discussion (separately)**: abstract (234 words); significance statement (118 words); introduction (610 words); discussion (1497 words)

Competing Interests Statement: The authors declare no competing financial interests.

Acknowledgments: This study was supported by the National Natural Science Foundation of China (grant No. 32300879 and W2421004 to W.L.), Humanities and Social Sciences Fund, Ministry of Education (grant No. 22YJCZH109 to W.L.).

Abstract

A hallmark of human memory is its ability to form novel inferences by linking discrete but related events. We examined whether acute stress impairs memory inference and if retrieval practice can buffer this effect. Participants were trained on image pairs, AB and BC, to establish interconnected triads (ABC) with a shared bridge element B. Twenty-four hours later, we induced acute stress in half of the participants and then tested their capacity to infer the indirect AC associations. Behavioral results indicated that acute stress reduced memory inference accuracy and speed, yet targeted retrieval practice of the AB and BC pairs after encoding prevented the stress-induced impairment in both A-C inference accuracy and speed. Using multivariate decoding analysis of human electroencephalogram (EEG) recordings, we found neural evidence that bridge element B is rapidly reactivated during the inferential process, a neural signature predictive of subsequent successful inference. Importantly, stress disrupts this rapid neural reactivation of the bridge element, but retrieval practice buffers the stress effect and enhances the strength of reactivation signals beyond the non-stress condition. Time-frequency analyses of theta oscillations mirrored these effects of stress and retrieval practice on inference performance and neural reactivation. Collectively, our findings pinpoint rapid reactivation of the bridge element as an essential neural mechanism for memory-based inference. Although susceptible to stress, this mechanism can be enhanced through retrieval practice, suggesting that building robust memory traces renders subsequent inferences resilient to stress.

Keywords: retrieval practice; acute stress; neural reactivation; memory inference; retrieval

Significance Statement

Stress often disrupts our ability to form new connections from past experiences, thereby limiting cognitive flexibility. Our study demonstrates that stress impairs memory inference-the ability to link related events to generate novel links-by disrupting the rapid reactivation of critical memory elements. However, we found that a simple yet effective technique, retrieval practice, which involves recalling related information, can prevent this negative effect. By reactivating prior memory traces through repeated recall, we can restore the ability to make inferences and even enhance the strength of memory reactivation. These findings highlight the potential of using strategic learning methods to protect and enhance cognitive flexibility, particularly under stress, with broad implications for improving memory resilience in high-pressure environments such as exam.

Introduction

The ability to make novel inferences from distinct yet related memories is essential for adaptive behavior. Successful memory inference is thought to depend on the effective retrieval of constituent memories. However, acute stress is known to impair memory retrieval, potentially disrupting subsequent inference (Vogel and Schwabe, 2016). In the present study, we examined how acute stress impairs the cognitive and neural mechanisms underlying memory inference. Furthermore, we investigated whether retrieval practice, a potent learning strategy, could buffer against this stress-related impairment and restore the associated neural processes.

To investigate the neural underpinnings of memory inference, we recorded electroencephalogram (EEG) signals while participants performed an associative inference task (Preston and Eichenbaum, 2013; Schlichting and Preston, 2015). This paradigm requires recombining information across overlapping pairs (e.g., AB, BC) to make a novel inference (e.g., A-C). Previous fMRI studies have highlighted the roles of the prefrontal cortex (PFC) and medial temporal lobe (MTL) in integrating these overlapping associations (Zeithamova and Preston, 2010; Zeithamova et al., 2012). More recently, multivariate decoding analyses have begun to elucidate how the representation of the linking "bridge element" (i.e., B) is reactivated to support subsequent inference (Morton et al., 2020; Schlichting et al., 2021). However, because these studies often lacked neural data during the active A-C inference test, the direct role of the bridge element in supporting online inference remains unclear. We posited that the rapid neural reactivation of the bridge element B is crucial for successful

inference. By leveraging the high temporal resolution of EEG and employing time-resolved multivariate decoding analysis (Cichy et al., 2014; Linde-Domingo et al., 2019), we aimed to capture this reactivation during the inferential process itself, focusing on distinct stimulus categories (i.e., faces and buildings) as bridge elements.

Acute stress is known to disrupt episodic memory retrieval, often by affecting the prefrontal-hippocampal circuit (de Quervain et al., 1998; Buchanan et al., 2006; Wolf, 2017; Gagnon et al., 2019). Importantly, under stress, individuals tend to shift from flexible, MTL-dependent cognitive strategies to more rigid, striatal-dependent habitual behaviors (Schwabe et al., 2022), such as relying on familiar paths rather than discovering novel shortcuts during navigation tasks (Brown et al., 2020). Given these neurocognitive alterations, we anticipated that acute stress would hinder the flexible recombination of stored memories (i.e., AB and BC) required for inference (i.e., from A to C).

Considering the detrimental effect of stress on memory, we sought to identify a strategy to protect inferential abilities. While various methods have been explored to mitigate stress-induced memory deficits (J-F de Quervain et al., 2007; Schwabe et al., 2009; Smith and Thomas, 2018), we focused on retrieval practice—the act of actively testing memory. Retrieval practice is a highly effective learning technique that not only enhances overall memory (Karpicke and Blunt, 2011; Roediger and Butler, 2011) but has also been shown to prevent stress-induced impairments in free recall (Smith et al., 2016). This protective effect may arise because retrieval practice promotes rapid consolidation (Antony et al., 2017; Liu et

al., 2020; Zhuang et al., 2021), creating stress-resilient cortical memory representations that are less susceptible to stress-induced glucocorticoid release (De Quervain et al., 2003).

The present study integrates these theoretical frameworks. We used EEG decoding, along with behavioral measures of inference accuracy and speed in humans, to test a series of hypotheses (Figure 1). First, we hypothesized that the rapid neural reactivation of the bridge element B supports successful memory inference. Second, we predicted that acute stress would disrupt this neural reactivation process and impair inferential performance. Finally, we hypothesized that retrieval practice would counteract these deficits by strengthening memory traces, thereby restoring the critical neural reactivation of bridge information and preserving memory inference under stress.

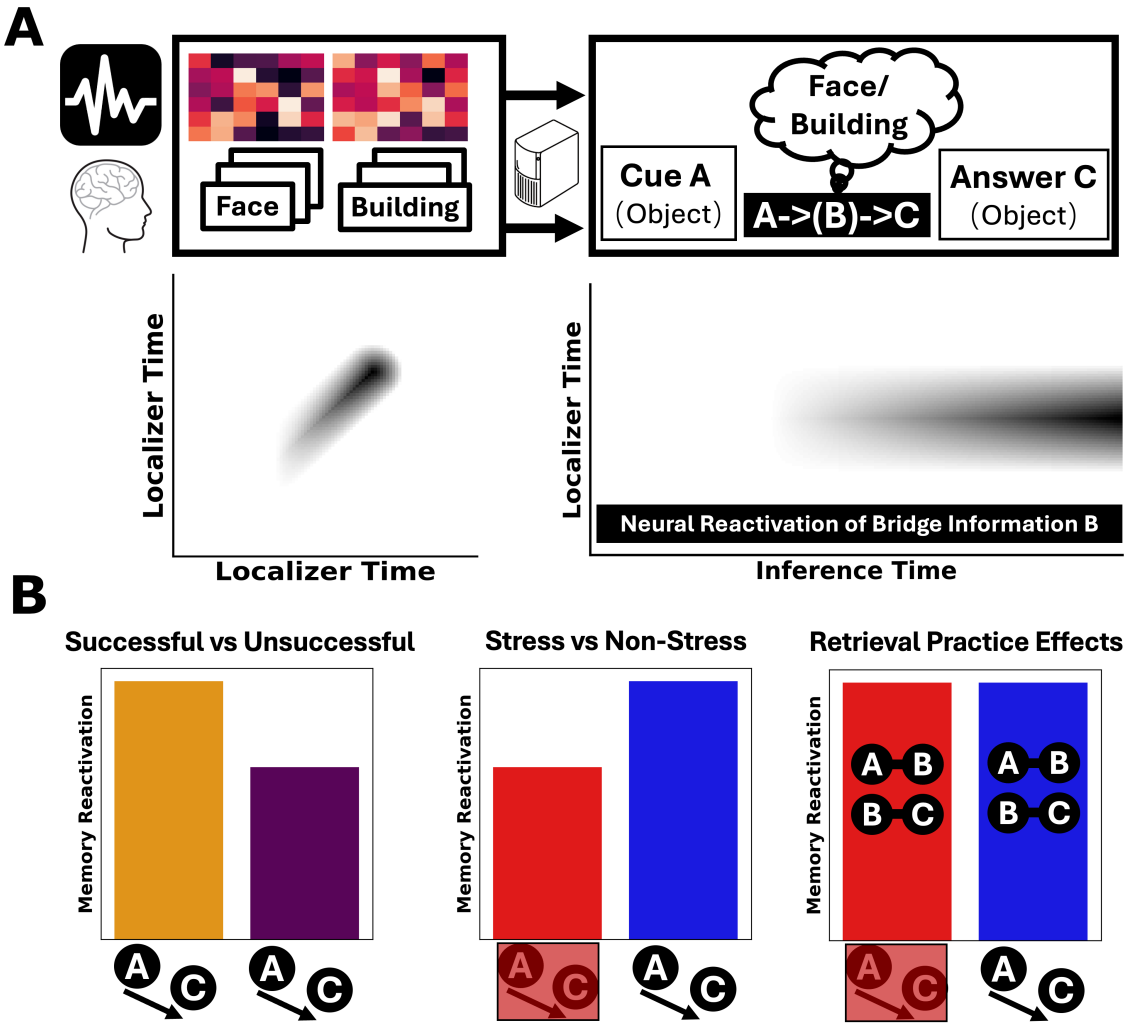

Figure 1 Design for EEG experiment and neural decoding for memory reactivation. (A) In the EEG experiment, participants performed an inference task in which they were asked to infer the correct answer-C based on the presented cue-A, using pretrained AB and BC associations. Element B, the "bridge information," could be either a face or a building image. Both the cue-A and the answer-C were pictures of daily-life objects. EEG data were recorded during both the inference and an independent localizer phase. During the localizer phase, participants actively viewed a different set of face and building images. A classifier was trained on category-specific neural response patterns from the localizer phase and then applied to the inference phase. This approach enabled time-series decoding analyses to detect, within individual trials, whether category information of the bridge element-B was reactivated at any moment. **(B)** Using neural decoding to capture the memory reactivation of bridge information-B, we examined whether this neural signature was stronger during successful inference trials (left panel) and weaker under acute stress (middle panel). Additionally, we assessed whether the neural signature could be protected through retrieval practice (right panel).

Method

2.1 Participants

We tested a total of 128 healthy college students across both experiments. Experiment 1 involved 68 participants (50 females; M_{age}=20.43 ± 1.80 years), who were randomly assigned to either a stress condition (n=34; M_{age} = 20.35 years, SD_{age} = 1.87, 26 females) or a control condition (n=34; M_{age} = 20.5 years, SD_{age} = 1.75, 24 females). There were no significant differences in age or gender between the two conditions (age: t(66) = -0.335, p = 0.739, Cohen's d = -0.081, 95% CI = [-1.02,0.73]; gender: $\chi^2(1,68)=0.302$, p=0.582). The remaining 60 participants (50 females; M_{age} 20.12 ± 1.46 years) were tested in the Experiment 2, which involved random assignment to inference under stress (n=30; M_{age} = 19.93 years, SD_{age} = 1.41, 28 females) or inference without stress (n=30; M_{age} = 20.3 years, SD_{age} = 1.51, 23 females). Again, there were no significant differences in age or gender between these groups (age: t(58) = -0.971, p = 0.336, Cohen's d = -0.251, 95% CI = [-1.12,0.39]; gender: χ^2（1,60）=2.092, p=0.148). All participants were right-handed and had normal or corrected-to-normal vision. The sample size was determined based on precedents in the relevant literature examining memory and stress using similar factorial designs and neuroimaging methods (Backus et al., 2016; Smith et al., 2016). A formal a priori power analysis was not conducted, as the effect size for the interaction between acute stress and retrieval practice on neural markers of inference was not known. Our sample size is comparable to or larger than those used in prior related studies. A post-hoc sensitivity analysis confirmed that our sample was sufficiently large to detect the critical stress-induced impairment on inference

performance (observed effect size Cohen's d = 0.56). The study received approval from the Ethics Review Committee for Life Sciences at Central China Normal University, and all participants provided informed consent prior to participation.

2.2 Experimental procedure

2.2.1 Task overview

We conducted two two-day experiments to investigate whether retrieval practice can mitigate stress-induced memory inference impairments and explore the associated neural reactivation mechanisms. The only difference between the two experiments was that Experiment 2 included a retrieval practice phase following the initial encoding phase. Detailed procedures are illustrated in **Figure S1**.

On Day 2, they returned for baseline stress measurements, which included heart rate recording and the Positive and Negative Affect Scale (PANAS, PA1, and NA1). Participants were then randomly assigned to either the stress or control group. After the stress induction or control procedure, and once cortisol levels peaked, they completed a three-alternative forced-choice test assessing inference (AC) while EEG recorded brain activity. Memory for directly learned associations (AB, BC) was also tested for subsequent trial-by-trial behavioral analysis. The second PANAS (PA3 and NA3, PA4 and NA4) was administered 10 minutes post-stress, with continuous heart rate monitoring. Finally, participants completed a localizer task to train an EEG-based machine learning model to classify two distinct types of visual stimuli—famous faces and famous buildings—based on neural response patterns. Data from the localizer task served as the training dataset for the model, which was used to decode

neural responses during the memory inference phase, specifically targeting the reactivation of bridge information B. We primarily used behavioral data and neural responses during the AC inference phase to address our main research questions.

2.2.2 Stimuli

We employed two distinct sets of visual stimuli for this study. One set was used in the main memory task, comprising the encoding, inference, and retrieval phases, while the other was used for the localizer phase. The two sets were independent of each other, and all images are publicly available (see our statement on Data and code availability below) for future experiments or replication purposes.

Stimuli for the main memory task: The main memory task used 72 images divided into three categories: faces, buildings, and objects. Specifically, this set included 12 images of famous faces (e.g., Andy Lau), 12 images of iconic buildings (e.g., the Great Wall), and 48 images of everyday objects (e.g., an iPad). These images were selected to reflect items commonly encountered in the daily lives of Chinese college students. The stimuli were organized into 24 "ABC" memory triads, where A and C were always two distinct object images, and B was either a face or a building image. Stimuli for the localizer phase: The localizer phase included 108 images, evenly distributed across the same three categories (36 famous faces, 36 iconic buildings, and 36 everyday objects). The images of everyday objects were sourced from the THINGS database(Hebart et al., 2019), while images of famous faces and buildings were obtained from publicly available online sources. All images were resized to a resolution of 474 × 658 pixels and edited to remove distracting elements.

2.2.3 Encoding phase

During the encoding phase, participants learned 48 memory pairs (AB and BC) derived from ABC triads, with each triad divided into two separate components (AB and BC) for independent encoding. Participants were explicitly instructed to learn the relationship between the two images displayed on the screen; however, no hints were provided regarding the potential to form a complete triad. Each encoding trial began with a fixation cross presented for 100 milliseconds, followed by the display of an AB or BC image pair for 2500 milliseconds. The next trial began immediately after the fixation cross was presented again. Each AB or BC pair was presented twice, with the positions of the images reversed between presentations. For example, if image A appeared on the left and image B on the right during the first presentation, their positions were swapped during the second presentation, with image B on the left and image A on the right. Notably, images A and C were never presented together during encoding, enabling a memory inference test on Day 2. To control the temporal interval between the presentation of AB and BC pairs, a mini-block design was employed, ensuring a trial interval of 1 to 4 trials between corresponding AB and BC pairs. The encoding phase consisted of two runs, with participants allowed to take breaks between runs to maintain focus throughout the task (**Figure S1**).

2.2.4 Retrieval practice phase

In Experiment 2, a retrieval practice phase was conducted following the encoding phase. Participants were given a sheet of paper and were instructed to perform a free recall test of the previously studied pairs (both AB and BC). They were given two minutes to write down

as many pairs as they could remember. No actual pictures were presented; participants need to write down the content of picture pairs as retrieval. At the end of the two-minute period, the experimenter collected the response sheets. Importantly, no corrective feedback was provided on their performance. This phase was not included in Experiment 1.

2.2.5 Stress procedure

To ensure equivalent encoding contexts for both control and stress groups and to confine the stress effects to Day 2, all participants underwent the same encoding procedure on Day 1, without knowledge of their Day 2 group assignment (i.e., stress or control).

Stress induction

To induce acute stress, we employed a modified version of the Trier Social Stress Test (TSST), a well-established protocol for activating the hypothalamic-pituitary-adrenal (HPA) axis and the sympathetic nervous system(SNS)(Kirschbaum et al., 1993). The procedure was administered by a panel of three evaluators who maintained neutral expressions throughout. To amplify the socio-evaluative threat, participants were informed that their performance was being video-recorded for later analysis; no recordings were retained to ensure participant privacy. The stress protocol began with a 3-minute anticipation period, during which participants received instructions for an impromptu speech. This was followed by a 4-minute public speaking task, where participants were required to present themselves as candidates for a teaching certification, explaining how they would teach a collection of prose poems. ly, participants engaged in an 8-minute mental arithmetic task. To increase the cognitive load and associated stress, we modified the standard TSST by replacing the serial subtraction task

with challenging sequential two-digit multiplication problems (e.g., 56×47). Participants submitted answers via a keyboard within a 5-second window, and an audible alarm was triggered for incorrect or delayed responses. Pilot testing had confirmed that completing these calculations within the allotted time was exceedingly difficult.

Control condition

The control procedure was carefully designed to match the duration and environmental context of the TSST while eliminating its stressful components. This approach, which removes socio-evaluative threat and time pressure while retaining non-stressful cognitive tasks, is consistent with validated control procedures used in previous TSST studies (von Dawans et al., 2012; Buchanan et al., 2014; Smith et al., 2016). The procedure consisted of two parts. First, participants quietly read the same collection of prose poems for 7 minutes. Second, they completed a simple, self-paced arithmetic task (e.g., 2+3) for 8 minutes without time constraints or performance feedback. Participants were not disturbed, even for incorrect or delayed responses.

2.2.6 Memory inference phase

Before the formal test, we used slides and practice trials featuring completely unrelated images to explain the structure of the ABC task and to clarify that picture A and C were indirectly related through B. After confirming participants' understanding of the inference task, they completed a three-alternative forced-choice memory inference phase. Each inference trial began with the presentation of a fixation point, displayed for a randomly selected duration (500 ms, 1000 ms, 1500 ms, or 2000 ms), followed by the memory cue (i.e.,

picture A) for 4000 ms. Participants were instructed to vividly retrieve and imagine the answer (i.e., picture C) during this period. This phase was the primary window of interest for EEG analyses, as successful inference required the recall of picture B (which we referred to as the "bridge information") and picture C during this phase. Following the presentation of the memory cue, three candidate pictures were shown, and participants were required to select the correct answer within 3000 ms using the keyboard. One of the pictures was the correct answer (i.e., picture C), while the other two were distractor object images. These distractors had been presented during the Day 1 encoding phase but were not paired with the current memory cue A(**Figure S4**). This setup ensured that participants had to perform the full A-B, B-C inference to select the correct answer, rather than choosing based on familiarity alone (e.g., recognizing the picture from earlier encoding phase). If participants selected the correct option, the trial was labeled as correct. Additionally, we took further steps, analyzing trial-by-trial data from the retrieval phase to rule out trials where the correct answer was selected due to guessing (See *Statistical Analyses* below).

2.2.7 Direct memory retrieval phase

In addition to the A-C memory inference task, participants were also tested on their originally learned AB/BC pairs. We refer to this phase as direct memory retrieval to emphasize its distinction from the A-C inference, which involves both memory retrieval and inference. At the start of each direct memory retrieval trial, a fixation point was presented for a randomly selected duration (500 ms, 1000 ms, 1500 ms, or 2000 ms). Next, a memory cue (i.e., picture A or C) and two response options (one being the correct picture B) were

presented simultaneously, with a 3000 ms window for participants to make a selection. Similar to the inference phase, the distractor image was a previously presented image from another AB or BC pair and belonged to the same category as the correct option. This task required direct recalling information from the encoding phase without engaging in memory inference. The direct retrieval phase consisted of 48 trials, testing all 24 AB and 24 BC pairs learned (**Figure S1**).

For the crucial A-C inference test, we used a three-alternative forced-choice (3-AFC) design where participants saw item A and had to choose the correct item C from three options. This design served two methodological purposes. First, it controlled for non-inferential response strategies. The two foil items were selected such that they were paired with other bridge elements (B items), with one foil's bridge sharing the same stimulus category (face or building) as the correct triad's bridge and the other foil's bridge belonging to the different category. This forced participants to rely on the specific A-B-C association rather than a general categorical link. Second, the 3-AFC design reduces the probability of a correct guess to 33.3%, thereby increasing confidence that trials scored as 'correct' truly reflect successful inference. This is particularly important for the statistical power and validity of our subsequent EEG analyses contrasting successful and unsuccessful trials.

In contrast, the direct memory tests for A-B and B-C pairs employed a standard two-alternative forced-choice (2-AFC) design. Given that memory for these directly learned pairs was expected to be high, this format was sufficient to confirm successful encoding without adding unnecessary complexity.

2.2.8 Localizer phase

At the end of the experimental procedure, participants completed a visual category localizer task employing a 1-back paradigm with images from three distinct categories: famous buildings, famous faces (male and female), and common objects. EEG was used to measure brain activity. The stimuli for this task were unique and did not overlap with those used in the main memory tasks (encoding, inference, and retrieval phases). Each run of the task comprised nine mini-blocks, with three mini-blocks assigned to each image category, and each mini-block containing 12 trials. During each trial, an image was displayed for 1500 ms, followed by a 500ms inter-trial interval. Participants were instructed to respond quickly and accurately to indicate whether the current image was identical to the previous one (**Figure S5)**. This task requirement served as a cover story to ensure participants remained engaged with the stimuli; however, their behavioral responses were not included in subsequent analyses. To prevent category-specific fatigue or habituation, the order of mini-blocks was pseudo-randomized, ensuring that two mini-blocks from the same category did not appear consecutively. The primary objective of the localizer task was to collect neural response patterns specific to each image category, which were later used to train a machine learning model for image classification based on brain activity (see *Decoding of EEG Signals* below).

2.2.9 Physiological and Psychological measures of acute stress response

In Experiment 1, heart rate, as a physiological measure of acute stress, was continuously monitored using the BIOPAC MP160 system. In Experiment 2, heart rate was tracked with

portable equipment, specifically the POLAR V2 system, which includes a Polar WearLink and a heart rate monitor. Heart rate was initially measured for 1 minute to establish a baseline and was continuously monitored throughout the subsequent tasks. Average heart rates were calculated for each experimental stage: a 3-minute preparation, a 4-minute public speaking, and an 8-minute arithmetic stage.

The Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) was employed to assess participants' current subjective affective states as a psychological measure of acute stress. This 20-item scale comprises 10 items for positive affects (e.g., "interested," "excited") and 10 items for negative affects (e.g., "nervous," "scared"). Participants rated each item on a 5-point scale reflecting their current state, from 1 (very slightly or not at all) to 5 (extremely). Average scores for both positive and negative affect were then computed.

We acknowledge that cortisol, a key glucocorticoid hormone, is central to the physiological cascade initiated by acute stress and its subsequent effects on memory. However, this study was conducted during the COVID-19 pandemic under strict institutional safety protocols designed to minimize infection risk. As collecting saliva samples for cortisol assay would have posed an unnecessary risk of viral transmission to participants and research staff, we made the a priori decision to forgo this measure. The successful induction of stress was therefore confirmed using a combination of physiological (heart rate) and psychological (PANAS) metrics. We recommend that future research under normal conditions incorporates cortisol analysis to further elucidate the precise hormonal mechanisms.

2.3 EEG acquisition and preprocessing

EEG data were collected using Brain Products GmbH actiCHamp Plus amplifiers at a 1000 Hz sampling rate. Sixty-four electrodes recorded continuous EEG, referenced to the vertex (FCz), with signals amplified through a 0.1-100 Hz band-pass filter and digitized at 250 Hz for storage. Offline processing involved band-pass filtering between 0.1 Hz and 30 Hz, down-sampling to 500 Hz, and re-referencing to the average of the left and right mastoids. To reduce line noise at 50 Hz, a band-stop filter was applied between 49 and 51 Hz. Subsequently, we visually inspected the ERP data and discarded any trials with residual ocular or other artifacts, such as those caused by movement. If an artifact was detected at any scalp site, data from all sites for that trial were excluded from further analysis. Interpolation was used to replace bad channels. Independent component analysis was performed to remove eye-blink and horizontal eye movement artifacts. The continuous EEG signal was segmented into 2500 ms periods for the localizer task and 4500 ms for the inference task, each starting 500 ms before stimulus presentation. Segments were baseline-corrected against the mean voltage of the 500 ms pre-stimulus period. Behavioral data were used to label each trial with a true or false category, indicating whether the trial involved a famous building or a male or female famous face. The EEG data were preprocessed using the MNE-Python software package (Version 1.9.0), applying a band-pass filter from 0.1 to 30 Hz. Following filtering, the entire 0.1–30 Hz bandwidth was analyzed without further frequency decomposition.

2.4 Multivariate time-resolved decoding of EEG signals

2.4.1 General overview

To investigate the neural reactivation of memory content, we performed a time-resolved multivariate decoding analysis on the EEG data using a Linear Discriminant Analysis (LDA) classifier, implemented with custom Python code adapted from the NeuroRA toolbox (Lu and Ku, 2020). For all decoding analyses, features were extracted from the preprocessed EEG signals using a sliding window approach. A 20 ms window (5 consecutive time points) was shifted in 20 ms increments across the epoch. The raw amplitude values from all 63 channels within each window were concatenated to form a feature vector (63 channels × 5 time points = 315 features) for each time step, preserving fine-grained temporal information.

To enhance the signal-to-noise ratio (SNR) of the EEG data before classifier training, we created trial aggregates (Isik et al., 2014; Grootswagers et al., 2017). For each participant and stimulus category (faces, buildings), we randomly sampled and averaged sets of 6 trials without replacement. This process was repeated until all trials for a given category were used, creating a balanced set of averaged trials (⌊N_trials / 6⌋ per category). To ensure the stability of our results, the entire decoding procedure, including the random sampling for trial averaging, was repeated 10 times, and the resulting accuracy maps were averaged across these iterations. We performed three distinct decoding analyses: (1) within-task classification to validate classifier performance, (2) cross-task classification to measure bridge element reactivation during inference, and (3) single-trial classification to link reactivation to behavioral success.

2.4.2 Within-task classfication

To validate the classifier's ability to distinguish between face- and building-related neural patterns, we first conducted a cross-temporal decoding analysis exclusively on the localizer task data. Using the averaged trials, we employed a stratified 5-fold cross-validation procedure, repeated twice for stability. In each fold, the classifier was trained on 80% of the data and tested on the held-out 20%. The model trained at each time point was tested on all other time points, yielding a temporal generalization matrix for each participant. These matrices were then smoothed using a 2D moving average filter (5 × 5 time points, i.e., 20 ms × 20 ms) to highlight consistent decoding patterns.

2.4.3 Cross-task classfication

To test for the rapid neural reactivation of the bridge element B, we performed a cross-task classification. The LDA classifier was trained on the entirety of the averaged trial data from the localizer task (faces vs. buildings). This fixed decoder was then applied to EEG data from the memory inference task to classify neural patterns at each time point, despite no face or building images being present. This analysis yielded a temporal generalization matrix for each participant, where the training time corresponds to the localizer task and the testing time corresponds to the inference task. As a secondary analysis, the same decoder was applied to EEG data from the A-C choice phase to confirm the persistence of reactivation.

2.4.4 Single-trial analysis of inference success

To directly link neural reactivation to memory inference performance, we applied the localizer-trained decoder to single trials from the inference phase, without trial averaging. We sorted trials based on whether the participant's subsequent response was correct (successful

inference) or incorrect (unsuccessful inference). This procedure generated separate temporal generalization matrices for successful and unsuccessful trials for each participant, allowing us to directly compare the strength of bridge element B reactivation as a function of behavioral outcome.

2.5 Time-Frequency Analysis of EEG Signals

Theta-band (2-8 Hz) neural oscillations were analyzed. EEG epochs (0 to 4 s relative to cue onset) were baseline-corrected using the -0.5 to 0 s pre-cue interval. Morlet wavelets were applied to compute time-frequency representations (TFRs). Power was extracted and decimated by a factor of 10 from the original 500 Hz sampling rate, yielding a 20ms temporal resolution. Group comparisons of this averaged theta power (within the 0 to 4 s window) utilized a non-parametric cluster-based permutation test (10,000 permutations) to correct for multiple comparisons across time points. Independent samples t-tests were performed at each time point; contiguous points exceeding a t-value threshold of ±1.96 formed clusters. Cluster significance ($p < 0.05$) was assessed against a permuted null distribution of maximum cluster-level statistics (sum of t-values within a cluster). For visualizing group differences across the full time-frequency plane (2-8 Hz, 0 to 4 s), t-statistics were computed for each time-frequency bin from channel-averaged power. These t-value matrices were smoothed using 3rd-order bivariate spline interpolation for plotting.

2.6. Statistical analysis

2.6.1 Validation of Acute Stress Induction:

To assess the induction of acute stress, two-way mixed ANOVAs were performed on heart rate and subjective affect (separately for positive and negative affect), with group (stress vs. control) as a between-subjects factor and acquisition period as a within-subjects factor. We also defined three indices to represent individual differences in acute stress responses. For heart rate and negative affect ratings, we subtracted the baseline value from the post-induction value, with higher values indicating stronger stress responses. In contrast, for positive affect ratings, we subtracted the post-induction value from the baseline value, with a larger reduction indicating a stronger stress response.

2.6.2 Analysis of Memory Inference Performance

To examine whether retrieval practice mitigated stress-induced impairments in memory inference, two-way ANOVAs were conducted on memory inference performance, with group (stress vs. control) and strategy (Study 1: no retrieval practice vs. retrieval practice) as between-subjects factors. Additionally, an independent t-test was performed in Study 1 to isolate the effect of acute stress on memory inference performance (i.e., whether acute stress impaired memory inference). Another independent t-test was conducted in Study 2 to compare the stress and control groups, aiming to determine whether retrieval practice could prevent stress-induced inference impairments. Furthermore, individual differences in memory inference were correlated with acute stress response indices and memory reactivation indices to establish a comprehensive model incorporating physiological, neural, and cognitive measures.

2.6.3 Analyses of Decoding Accuracy Maps

We employed two approaches to analyze the generated decoding accuracy maps. The first analysis focused on determining whether mean classification accuracy exceeded the chance level of 50%. We applied permutation-based cluster-level correction to identify time windows with significant decoding accuracy, thereby suggesting the optimal training features and the timing of memory trace reactivation. Specifically, a cluster-based permutation test was applied to the 2D spatial data, which is well-suited for analyzing spatially correlated data, such as neuroimaging and EEG results. Initially, a one-sample t-test was performed for each spatial point in the 2D grid, comparing the results to chance level. A one-sided test with a p-value threshold of 0.05 identified significant points, which were assigned a value of 1, while non-significant points were marked as 0. Significant points were then grouped into clusters of adjacent significant points based on spatial connectivity. To assess cluster significance, a permutation procedure was applied, where the data were shuffled, and t-statistics were recalculated for 5000 iterations. The maximum t-statistic from each iteration was recorded, and clusters with observed t-statistics exceeding the 95th percentile of the permutation distribution were considered significant. The final output was a 2D matrix of binary values indicating the significance of each spatial point, ensuring robust control over Type I errors and identifying spatially contiguous regions of significant clusters.

2.6.4 Optimal time window identification and validation

To compare neural reactivation strength between experimental conditions (e.g., Stress vs. Non-Stress) without incurring the problem of multiple comparisons across all time points, we focused our statistical analyses on an "*optimal time window*" that captured the peak of the

reactivation signal. This window was defined a priori at the group level and validated using the following three-step procedure: (1) Identification of Subject-Level Peak Latency: We first identified the most robust time window for decoding in the classifier training data (the localizer task) for each participant individually. We performed a cluster-based permutation test on each participant's time-series of decoding accuracy. We identified temporal clusters of significant decoding (defined as contiguous time points exceeding 50% accuracy for at least 10 ms). The "peak latency" for each participant was then defined as the center-of-mass of the cluster with the highest summed accuracy score. (2) Definition of the Group-Level Optimal Window: We then aggregated these peak latencies from all 128 participants. The group-level peak was the mean of this distribution, which was 207.17 ms (SD = 228.17 ms). Based on this, we defined the optimal training time window as 207±50 ms for our primary analyses. (3) Cross-Task Validation: To ensure this time window was also representative of reactivation during the inference task, we performed a time-generalization MVPA. This analysis produced a 2D matrix showing classifier performance for every combination of training time (from the localizer task) and testing time (from the inference task). We applied the same center-of-mass procedure described in Step 1 to this cross-task decoding matrix to independently identify the peak reactivation time during inference. This validation yielded a peak latency of 230.48 ms (SD = 139.81**).** This latency did not differ significantly from the one identified during model training ($t(127)=1.39$, $p=0.16$), indicating no significant temporal shift in the reactivation signature between tasks. This result validates our use of the fixed 207±50 ms window for all subsequent statistical comparisons of reactivation strength.

2.6.5 Linking Neural Reactivation to Subsequent Inference Effect, Stress, and Retrieval Practice

Each participant generated three decoding accuracy matrices: (1) all trials, (2) trials with correct inferences (successful inferences), and (3) trials with incorrect inferences (unsuccessful inferences). Since each matrix is a high-dimensional structure containing over 200 × 450 accuracy outputs, we applied dimensionality reduction by averaging the values within the identified optimal training time window prior to further analysis. To quantify individual differences in memory reactivation within this window, we defined two indices representing the temporal and strength aspects of memory reactivation: *decodability* and *mean decoding accuracy*. Decodability was calculated by determining the percentage of time points with decoding accuracy greater than chance within the specified time window. This index ranges from 0 to 1, with higher values indicating greater memory reactivation across time points. From a cognitive processing perspective, higher decodability suggests sustained memory reactivation during the inference phase, reflecting the temporal aspect of reactivation. In contrast, mean decoding accuracy represents the average decoding accuracy and ranges from 0 to 50, as raw accuracy was corrected for chance and scaled by 100. Higher mean decoding accuracy indicates more accurate decoding of bridging information from brain activity, which may suggest stronger underlying memory representations.

To compare the neural indices of memory reactivation across experimental conditions, we defined the subsequent inference effect as the comparison of mean decoding accuracy and decodability between successful and unsuccessful inference trials, analogous to the

subsequent memory effect, which compares neural activity between correctly and incorrectly recalled trials. Because successful and unsuccessful trials were compared within participants, a paired t-test was conducted on the two memory reactivation indices with a threshold of $p = 0.05$ (two-tailed). Additionally, we compared the two neural indices between Experiment 1 and Experiment 2 to examine the general effects of retrieval practice on memory reactivation, between stress and control participants in Experiment 1 to assess the effect of acute stress on memory reactivation, and between stress and control participants in Experiment 2 to investigate the protective effect of retrieval practice on memory reactivation under stress. To test the directional hypotheses derived from our behavioral findings, we used one-tailed independent t-tests. The alpha level for these analyses was set at 0.05. Each specific hypothesis is outlined alongside its corresponding statistical test in the Results section.

2.7 Code availability

The custom code used in this study is available in https://osf.io/v54g8/?view_only=dd87e13d50b04c30b9ed839c58b80d95

2.8 Data availability

The data and that support the findings of this study are in https://osf.io/v54g8/?view_only=dd87e13d50b04c30b9ed839c58b80d95

3. Results

We conducted two independent EEG experiments with healthy young adults (total N = 128; Experiment 1: N = 68; Experiment 2: N = 60) to investigate how acute stress impacts memory inference and whether retrieval practice can mitigate such effects. In both experiments, participants encoded overlapping associative pairs (AB and BC) on Day 1, forming integrative ABC triads. On Day 2, they were tested on their ability to infer the indirect A-C relationship **(Figure 2)**. Experiment 1 investigated how acute stress, induced at the start of Day 2, affects inference performance and the underlying neural reactivation of the bridge element (B) compared to a non-stress control group. Experiment 2 assessed whether retrieval practice on the AB and BC pairs, implemented after encoding on Day 1, could buffer against these stress-induced impairments by modulating memory reactivation. The results are presented in five sections. First, we report manipulation checks confirming the efficacy of the stress induction at both psychological and physiological levels. Second, we present the effects of acute stress on memory inference accuracy and speed, and the mitigating role of retrieval practice. Third, we detail the results from our multivariate EEG decoding analyses of neural reactivation during inference. Fourth, we provide group comparisons of the time-resolved decoding outcomes to specify the effects of stress and retrieval practice. Finally, we report two control analyses: one to determine whether our decoding results were specific to the inference period versus the subsequent choice phase, and a time-frequency analysis to assess whether theta-band oscillatory power was consistent with the observed behavioral and reactivation patterns.

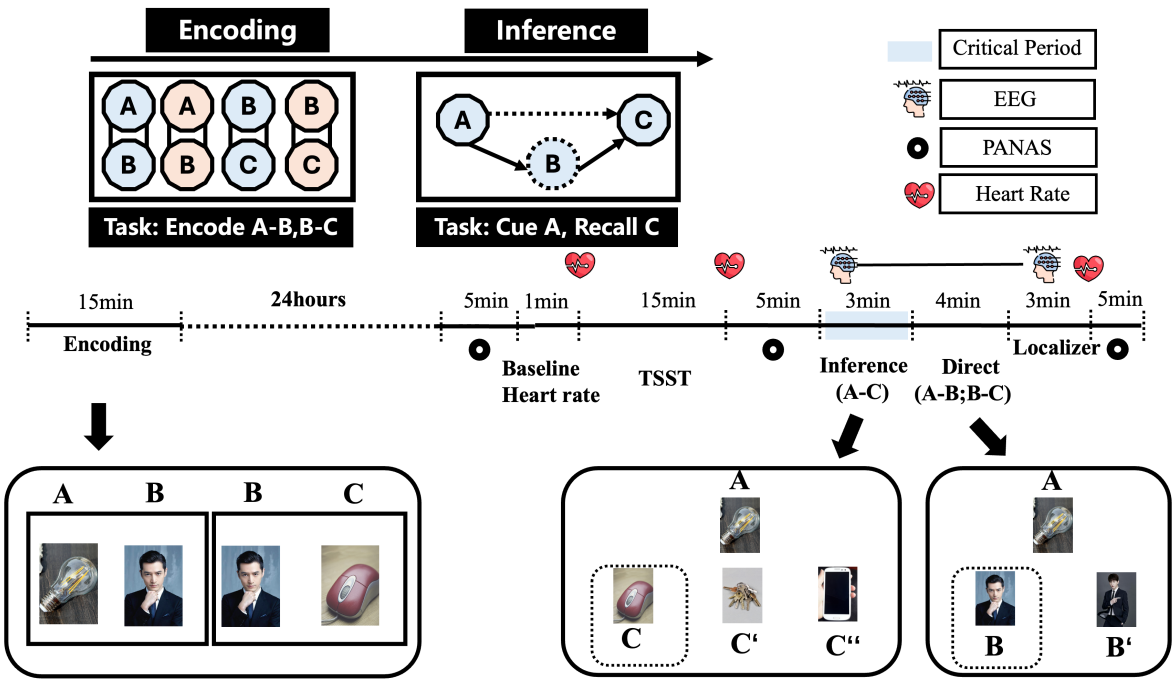

Figure 2: Task Overview. During the encoding phase, participants learned two types of associations: AB and BC, where each association consisted of pairs of images, forming interconnected triads (ABC). During the inference phase, picture A was presented as a cue to recall picture C. Importantly, participants were not informed about the inference requirement during the AB and BC encoding phases. The task was divided into two sessions, separated by a 24-hour interval. On Day 1, the encoding phase lasted approximately 15 minutes, during which two pictures were presented side-by-side on the screen. On Day 2, participants underwent either a stress (i.e., Trier Social Stress Test [TSST]) or a control manipulation, and then finish the memory tasks. Acute stress responses were assessed using heart rate measurements and subjective emotion ratings (i.e., Positive and Negative Affect Schedule [PANAS]). EEG data were recorded during both the inference and localizer phases, with the inference phase being the critical period of our analysis. During the inference phase, Cue-A was presented alongside the correct answer-C and two lure images from other learned ABC triads. The original AB and BC associations were also tested to enable subsequent trial-by-trial analyses of memory inference. Experiment1 and Experiment2 used the same task procedure, with experiment2 included the additional retrieval practice session in Day1, immedialy after the encoding. Trial structure of each task phase can found in **Figure S1**.

3.1 Manipulation checks of stress induction

In both experiment1 and 2, we evaluated the effectiveness of acute stress induction using both physiological (heart rate in beats per minute, BPM) measures and psychological (self-reported Positive and Negative Affect Schedule, PANAS).

For physiological responses, we monitored heart rate throughout the stress induction procedure both experiments and during the memory inference phase in Experiment 1 and Experiment 2. Heart rate responses were quantified as the difference between the average during the Trier Social Stress Test (TSST) phase and the baseline phase (i.e., before stress induction). Independent t-tests showed significant higher heart rate elevation in the stress condition compared to the control condition in both experiments (exp1: $t(66) = 9.377$, $p < .001$, Cohen's d = 2.274, 95% CI = [8.04,12.40], **Figure 3A**; exp2: $t(58) = 5.643$, $p < .001$, Cohen's d = 1.457, 95% CI = [4.19,8.81], **FigureS2A**).

For psychological stress responses, we analyzed changes in PANAS scores. Differences between the second measurement (after stress induction) and the first measurement (before stress induction) served as indicators of stress responses for both negative and positive affect. Independent samples t-tests were conducted with group (stress vs. control) as the between-subject factor. In both experiments, the stress induction procedure led to significantly increased negative affect (exp1: $t(66) = 2.335$, $p = 0.023$, Cohen's d = 0.566, 95% CI = [0.38,4.86], **Figure 3B**; exp2: $t(58) = 3.023$, $p = 0.004$, Cohen's d = 0.781, 95% CI = [1.05,5.15], **Figure S2B**), and decreased positive affect in the stress group compared to the control group (exp1: $t(66) = -2.155$, $p = 0.034$, Cohen's d = -0.523, 95% CI = [-6.35,-0.24], **Figure 3C**; exp2: t(58)

= -2.149, p = 0.036, Cohen's d = -0.555, 95% CI = [-3.73,-0.13], **Figure S2C**).

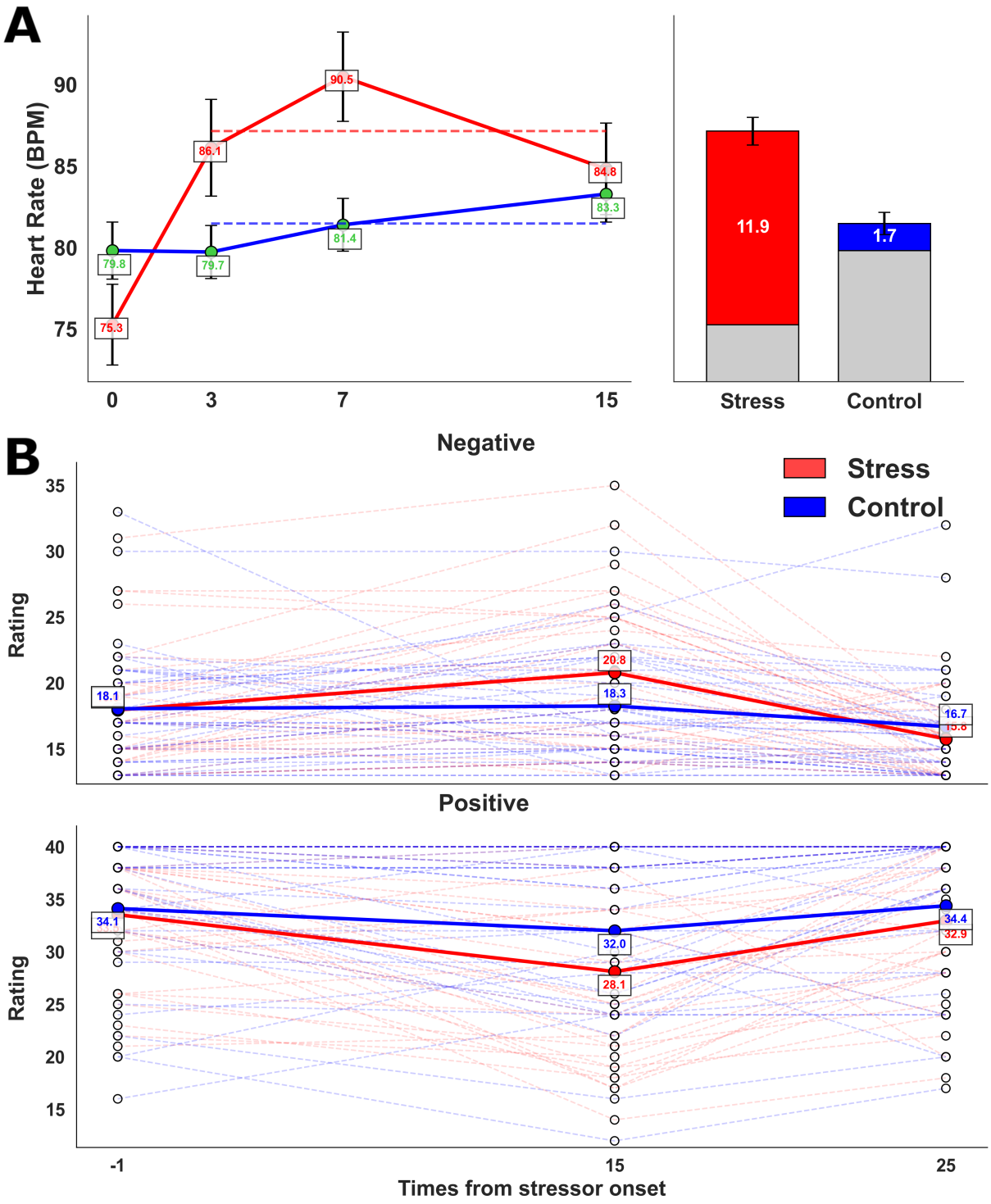

Figure 3. Manipulation Checks. (A) Acute stress induction significantly increased heart rate (HR), measured in beats per minute (BPM), during the Trier Social Stress Test (TSST). The

heart rate response was defined as the mean BPM difference between the TSST and baseline phases, with the stress group showing a mean increase of 11.9 BPM (SD = 4.95) and the control group showing a mean increase of 1.7 BPM (SD = 3.98). **(B)** Stress significantly increased subjective ratings of negative valence. **(C)** Stress significantly decreased subjective ratings of positive valence. Note: Results from Experiment 1 are presented here. Similar stress responses were observed in Experiment 2, which are shown in **Figure S2**.

3.2 Memory performance

3.2.1 Inference accuracy

In Experiment 1, we examined whether acute stress impaired memory inference (i.e., the ability to recall item C based on cue A in previously learned ABC triads). Memory inference performance was calculated as the percentage of correct AC inferences, contingent on successful recall of both AB and BC pairs to control for individual differences in memory encoding. Results revealed that acute stress significantly impaired memory inference compared to the control condition ($t(66) = -2.160$, $p = 0.034$, Cohen's d = -0.523, 95% CI = [-0.10，-0.00], **Figure 4A left panel**).

We further conducted exploratory correlational analyses to investigate relationships between individual differences in stress responses and memory inference. Stronger heart rate responses to stress were associated with poorer memory inference ($r(126) = -0.175$, $p = 0.048$), while higher reductions in positive affect correlated with worse performance ($r(126) = -0.162$, $p = 0.068$). To avoid interpretational issues with pooled data, we additionally conducted correlation analyses separately for each group and experiment. No significant associations were found between heart rate and A-C inference accuracy. In Experiment 1, this

was true for both the stress group (r=-0.171, *p*=0.333; r=-0.171, *p*=0.333) and the control group (r=-0.041, *p*=0.816; r=−0.041, *p*=0.816). We observed the same null result in Experiment 2 for the stress group (r=-0.198, *p*=0.294; r=-0.198, *p*=0.294) and the control group (r=0.143, *p*=0.452; r=0.143, *p*=0.452).

Next, Experiment 2 tested whether retrieval practice could mitigate the stress-induced inference impairment reported in the Experiment1. Compared to Experiment 1, Experiment 2 included a retrieval practice phase immediately following the initial encoding of AB and BC pairs. On Day 2 participants underwent the same stress procedure and memory inference task as in Experiment 1. Unlike Experiment 1, no significant difference in memory inference performance was observed between the stress and control group (t(58) = -0.488, p = 0.627, Cohen's d = -0.125, 95% CI = [-0.06，0.04], **Figure 4A right panel**). Notably, the retrieval practice phase included only AB and BC pairs, with no AC inferences. Participants were informed that they would be tested on these practiced pairs and were unaware that they would instead be assessed on AC associations on Day 2, thereby ruling out the possibility that the observed protective effects of retrieval practice resulted from explicit AC practice on Day 1.

To further examine the interaction between acute stress and retrieval practice, we combined data from both experiments and revealed significant main effect of retrieval practice (F(1, 124) = 4.589, p = 0.034, η2 = 0.036), suggesting that the retrieval practice enhanced inference performance regardless of acute stress induction. However, the main effect of acute stress was not significant (F(1, 124) = 3.609, p = 0.060, η^2 = 0.028), largely because the retrieval practice reduced the stress effect in experiment2. Finally, a

complementary analysis integrating data from both experiments revealed no significant interaction between acute stress and retrieval practice (F (1,124) = 1.232, p = 0.269, η2 = 0.009). We note for transparency that this analysis was planned as a complementary step, given that Experiment 2 was designed after the results of Experiment 1 were known. In summary, we found that acute stress significantly impairs memory inference (F(1, 124) =4.599, p = 0.034, η^2 = 0.036, 95% CI = [0.004,0.095]), while with the protection of retrieval practice, acute stress did not affect memory inference anymore (F(1, 124) =0.242, p =0.624, η^2 = 0.002, 95% CI = [-0.037,0.061])).

An alternative explanation is that retrieval practice in Experiment 2 reduced stress levels during the memory reference phase, resulting in undamaged performance under stress. However, this is unlikely for two reasons: (1) between-group design ensured participants were unaware of retrieval practice manipulation in the other group, preventing confidence biases; and (2) retrieval practice and memory inference are distinct tasks, such that inference task-related stress would not be alleviated by prior retrieval practice on AB and BC. Additionally, we conducted data analyses to rule out this alternative explanation. We analyzed the mean PANAS scores before and after the memory inference phase. Independent samples t-tests were performed with retrieval practice as the between-subjects factor and subjective positive and negative scores as the dependent variables. For both subjective positive and negative ratings, no significant differences between groups were observed in the memory inference phase between Experiment 1 and Experiment 2 (subjective positive

ratings: t(62) = 1.895, *p* = 0.063, Cohen's d = 0.475, 95% CI = [-0.15,6.21]; subjective negative ratings: t(62) = -0.985, *p* = 0.329, Cohen's d = -0.247, 95% CI = [-3.85， 1.37]).

3.2.2 Inference speed

To investigate the effects of acute stress and retrieval practice on the speed of memory inference, we analyzed response times (RTs) during the A-C inference test. A clear speed-accuracy relationship was observed, where successful inferences were associated with significantly faster RTs than incorrect ones (t(124) = -12.823, p < .001, Cohen's d = -1.241, 95% CI [-429.830, -286.080] ms; Figure 4C). We next examined how inference speed was modulated by stress, retrieval practice, and successful vs unscucesful inference (Figure 4D). In Experiment 1, which did not included retrieval practice, participants under acute stress exhibited significantly longer RTs than the control group (t(133) = 2.219, p = .028, Cohen's d = 0.382, 95% CI [12.738, 222.151] ms), indicating that stress impairs the efficiency of memory inference. In contrast, this stress-induced slowing was eliminated in Experiment 2 following retrieval practice, with no significant RT difference observed between the stress and control groups (t(116) = -0.785, p = .434, Cohen's d = -0.144, 95% CI [-183.812, 79.505] ms). Lastly, to probe the mechanism of this buffering effect, we assessed whether retrieval practice differentially impacted RTs on successful versus unsuccessful inference trials regardless of stress induction. The analysis revealed that retrieval practice significantly accelerated RTs for successful inferences (t(126) = -2.800, p = .006, Cohen's d = -0.307, Cohen's d = -0.499, 95% CI [-249.031, -42.694]) but had no significant effect on RTs for unsuccessful trials (t(123) = -1.707, p = 0.090, Cohen's d = -0.307, 95% CI [-182.657,

13.539]). These findings suggest that retrieval practice selectively enhances the processing efficiency of successful memory inferences, thereby mitigating stress-induced deficits.

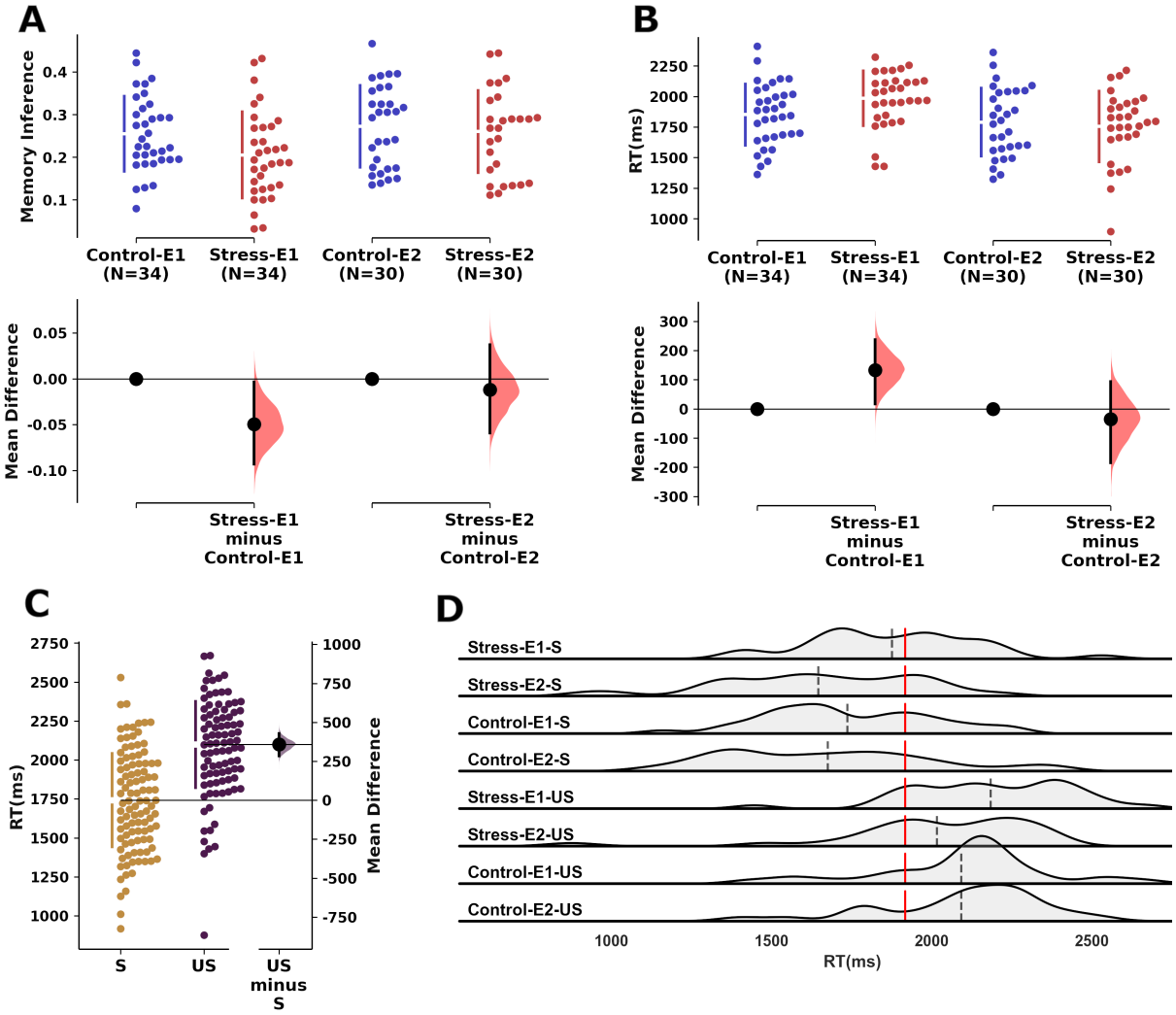

Figure4 Behavioral effects of stress and retrieval practice on memory inference. (A) In Experiment 1, acute stress significantly reduced memory inference accuracy (Stress-E1 vs. Control-E1). In Experiment 2, retrieval practice following encoding prevented this stress-induced deficit; no significant difference in accuracy was found between the stress and control groups (Stress-E2 vs. Control-E2). Moreover, retrieval practice enhanced overall inference accuracy relative to Experiment 1, irrespective of the stress condition. **(B)** Analysis of reaction times (RTs) revealed a similar pattern: stress impaired inference speed in Experiment 1, and this impairment was buffered by retrieval practice in Experiment 2. **(C)**

Reaction times were significantly faster for successful compared to unsuccessful inference trials across all conditions. **(D)** Distribution of RTs for successful and unsuccessful trials, shown for each experiment and stress condition. E1, Experiment 1; E2, Experiment 2; S, Successful inference; US, Unsuccessful inference.

3.2.3 Control Analyses: retrieval practice protects direct memory retrieval from stress without affecting premise pairs counts

We next examined whether stress had a broader impact on memory retrieval (i.e., for the individual premise pairs). Using a mixed-effects model to analyze direct memory retrieval performance for A-B and B-C pairs, we found that in Experiment 1, the stress group showed significantly impaired retrieval performance compared to the control group ($\beta = -2.21$, $z = -2.42$, $p = .016$, 95% CI [-4.00, -0.42]; Figure S3A left panel). This indicates that acute stress did indeed produce a general deficit in retrieving the directed encoded memories. Critically, this deficit was eliminated by retrieval practice in Experiment 2: we found no significant difference in direct memory performance between the stress and control groups ($\beta = -0.30$, $z = -0.47$, $p = .638$, 95% CI [-1.55, 0.95]; Figure S3B left panel). This result demonstrates that retrieval practice not only protects the complex process of memory inference but also bolsters the underlying memories, making them resilient to the general retrieval impairment caused by stress.

A critical prerequisite for interpreting our A-C inference results is to ensure that any observed group differences were not driven by disparities in statistical power. There is a possibility that if the stress group recalled fewer premise pairs (A-B and B-C), fewer trials

would be included in the conditional A-C inference analysis, potentially confounding the results. To directly address this, we calculated the number of "foundation pairs" for each participant, defined as triads for which both the A-B and B-C premise pairs were correctly recalled. A comparison of these counts revealed no significant difference between the stress and control groups in either experiment. In Experiment 1, the groups had a similar number of foundation pairs available for inference (t(66) = -1.60, p = .114, 95% CI [-3.50, 0.38], Cohen's d = -0.39; Figure S3B left panel). Likewise, in Experiment 2, there was no significant difference between the groups (t(58) = -0.62, p = .541, 95% CI [-2.41, 1.28], Cohen's d = -0.16; Figure S3B left panel). This key control analysis confirms that the groups were well-matched and that our main findings on A-C inference performance are not attributable to a difference in the number of analyzable trials.

3.2.4 Retrieval practice enhanced direct memory retrieval but not baseline inference

To further probe the mechanism by which retrieval practice protected inference against stress, we performed two additional analyses comparing performance across Experiment 1 (no retrieval practice) and Experiment 2 (with retrieval practice). First, we examined whether retrieval practice enhanced the memory for the directly encoded pairs. A t-test comparing the combined A-B and B-C accuracy across all participants from both experiments confirmed a significant benefit of retrieval practice, with higher accuracy in Experiment 2 than in Experiment 1 (t(126)=2.404, p=0.018, 95% CI: [0.369,3.804], Cohen's d=0.423; Figure S4). This finding indicates that, as intended, the retrieval practice manipulation successfully strengthened the foundational memory traces.

Second, we investigated whether this strengthening of direct memories led to a general improvement in inference, even in the absence of stress. To this end, we compared A-C inference accuracy between the non-stressed control group of Experiment 2 and the control group of Experiment 1. This analysis revealed no significant difference in performance between the two groups (t(62)=0.754, *p*=0.454, 95% CI: [−0.029,0.064], Cohen's d=0.189; FigureS5). Together, these results suggest that while retrieval practice bolstered the direct associations, its primary functional benefit was not a general enhancement of inference but rather a specific buffering effect that rendered the inferential process resilient to the disruptive impact of acute stress.

3.3 Rapid neural reactivation of bridge information supports subsequent memory inference

To capture the neural dynamics underlying memory inference, we employed multivariate pattern analysis (MVPA) on time-resolved EEG data. Our rationale was that if bridge information (i.e., picture B) is mentally reactivated during inference, then a classifier trained to discriminate between face and building images—using neural activity collected during an independent localizer phase—would successfully classify the category of the reactivated bridge information during inference, even though only object images (i.e., cue A and answer C) were presented (**Figure 5A**). We used this neural decoding strategy as a means to capture underlying mental processes that are difficult to assess via self-report and examined how this neural measure of memory reactivation supports memory inference. Following the principle of subsequent memory effect, we defined the subsequent inference

effect contrast (i.e., successful vs. unsuccessful inference trials) to investigate whether neural reactivation of bridge information predicted subsequent correct behavioral responses (**Figure 5B**). We present the EEG results in the following order: (1) within-task classification, in which we trained and tested the model to classify the category of pictures on localizer phase data; (2) cross-task classification, where the classifier was trained on localizer data and tested on inference phase data; and (3) analyses of the subsequent inference effect intergrating outcomes of neural reactivation analyses and trial-level behavioral responses.

Before analyzing the neural reactivation of memory items during the inference task, we first validated the performance of our multivariate classifier. We trained and tested the classifier on an independent localizer dataset using a cross-validation procedure (see Methods). The classifier was highly effective at distinguishing the stimulus categories (e.g., faces vs. buildings) from the EEG data (Figure 5C). The mean peak decoding accuracy across participants was 77.42% (SD=8.97%, 95% CI [75.85%,78.99%]), with individual peak accuracies ranging from 52.60% to 95.80%. On average, decoding accuracy was significantly above chance for a duration of 488.73 ms (SD=195.35 ms). The mean latency of the peak decoding accuracy was 479.54 ms (SD=170.28 ms) post-stimulus onset.

Next, to investigate whether and when the visual category (i.e., face or building) of bridge information (B) can be decoded from brain activity during an A–C inference trial, we applied cross-task classification to the time-resolved EEG data collected during the inference phase. The classifier was trained on neural activity elicited when participants viewed pictures from specific categories and then applied to the time window during which memory cue A

was presented, prior to the selection of answer C. We hypothesized that the category information of bridge information (B) would be reactivated, despite the absence of face or building stimuli during this phase. To examine the temporal dynamics of rapid neural reactivation at the single-trial level, we conducted a series of time-resolved decoding analyses. Linear discriminant analysis (LDA; see Methods) was employed to classify visual category information based on the EEG topography at each time point following the presentation of memory cue A. Two primary cross-task classification clusters were identified in the decoding accuracy maps. The first cluster extended from 180 to 340 ms during the localizer phase and from 710 to 3930 ms during the inference phase. The second cluster spanned 1100 to 1500 ms during the localizer phase and comprised two distinct intervals during the inference phase: 180 to 1520 ms and 2760 to 3440 ms (**Figure S6**).

Lastly, to validate the behavioral relevance of the identified neural reactivation of bridge information (B), we examined whether this reactivation predicted subsequent successful inference (i.e., Subsequent Inference Effect). Each trial was labeled as either successful or unsuccessful, and we anticipated that rapid neural reactivation would be more pronounced in successful trials compared to unsuccessful ones. For each participant, two decoding matrices were generated: one for successful inference and one for unsuccessful inference. Two participants were excluded from the Subsequent Inference Effect analyses because two types of trials could not be defined for them, resulting in a final sample of 126 participants. When focusing on neural activity preceding successful inference, we observed significant cross-task classification, ranging from 80 ms to 260 ms in the localizer phase and from 420 ms to 4000

ms in the inference phase (**Figure 5D, upper organge panel**). In contrast, when applying the classifier to neural activity preceding unsuccessful inference, we observed significant cross-task classification, ranging from 570 ms to 1500 ms in the localizer phase and from 70 ms to 610 ms in the inference phase (**Figure 5D, lower purple panel)**. Moreover, we identified the optimal time window to capture memory reactivation (See Methods for details), and then average each participant's 2D maps across time points within that time window and adjusting by subtracting the 50% chance level, thereby setting 0% as the new baseline. An initial paired t-test revealed that the classifier decoded reactivated bridge information (B) more accurately prior to successful inference than unsuccessful inference (t(124) = 2.063, p = 0.04, Cohen's d = 0.260, 95% CI = [0.126, 5.450]; **Figure 5E**). Follow-up tests indicated that the category of bridge information was significantly reactivated above chance levels prior to a successful inference response (t(125) = 1.941, $\boldsymbol{p}$ **= 0.027, one-tailed,** Cohen's d = 0.173, 95% CI = [–0.041, 4.221]). Conversely, for unsuccessful inference trials, decoding accuracy did not significantly differ from chance (t(125) = -0.855, $\boldsymbol{p_{one\text{-}side}}$ **= 0.19, one-tailed,** Cohen's d = –0.076, 95% CI = [–2.313, 0.917]). These findings underscore the behavioral relevance of the detected reactivation of bridge information during the inference phase.

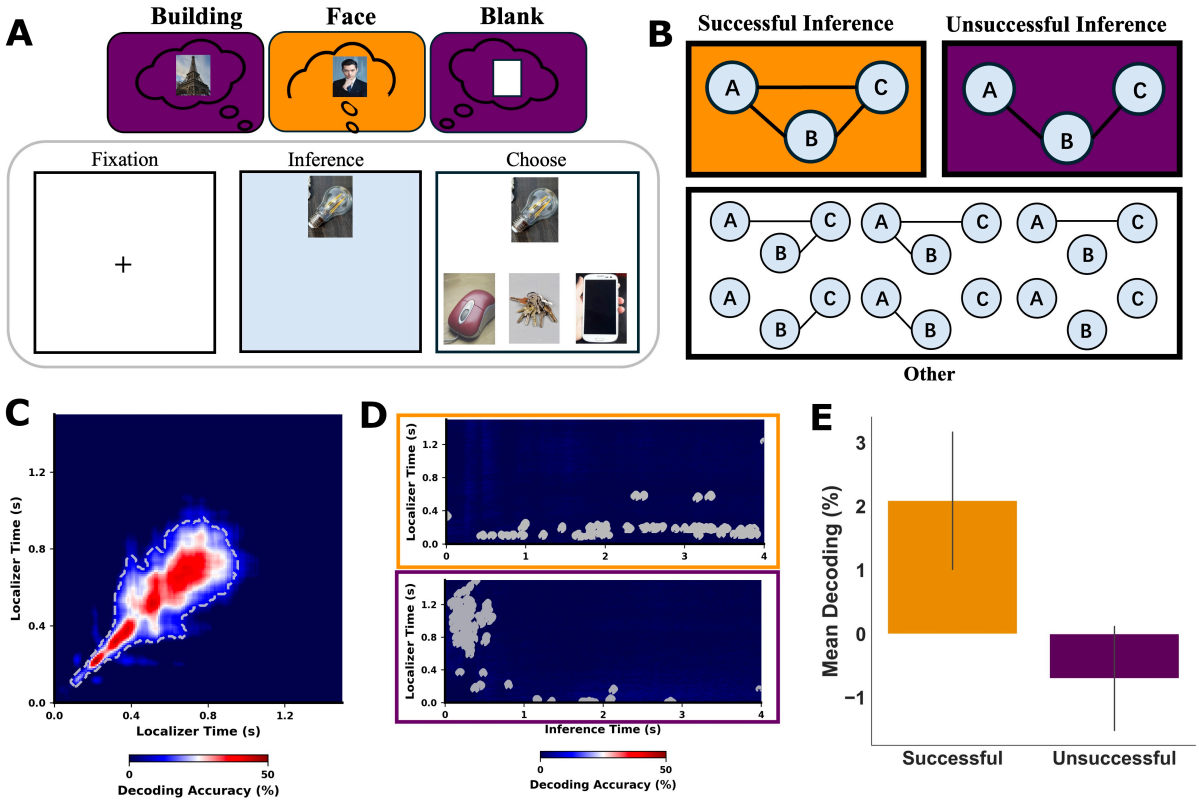

Figure5 Bridge information were reactivated to support subsequent successful inference. (A) Hypothesized reactivation of bridge information during the inference phase. To select the correct image-C based on cue-A, participants were hypothesized to mentally retrieve the bridge information-B (e.g., a face in the example trial, in orange). In some cases, the incorrect image (e.g., a building, in purple) might also be reactivated, or no image might be recalled during this phase. **(B)** Definition of successful and unsuccessful inference trials. In successful trials, both associations (A–B and B–C) were remembered, allowing for the correct inference of the A–C relationship (Orange Box). In contrast, in unsuccessful trials, despite remembering A–B and B–C, participants failed to correctly infer the A–C relationship (Purple Box). Trials in which either A–B or B–C associations were not encoded prior to inference were excluded from the analysis (White Box). **(C)** Results of time-generalization analysis from within-task classification. EEG data from the localizer phase were used to train a machine learning model to classify faces and buildings based on localizer-phase EEG activity. **(D)** Results of time-generalization analysis from cross-task classification. EEG data from the localizer phase were used to train a machine learning model to classify faces and buildings using EEG data from the inference phase. Decoding accuracy matrices were

computed separately for successful and unsuccessful inference trials (depicted in orange and purple, respectively). **(E)** Mean decoding accuracy within the optimal decoding time window, relative to chance, for successful and unsuccessful inference trials. Significantly above-chance decoding of bridge information predicted subsequent successful inference, demonstrating the Subsequent Inference Effect.

3.4 Acute stress disrupts reactivated neural representation of bridge information, but retrieval practice restores this process

Behavioral data indicate that acute stress impairs memory inferences in humans. To investigate how acute stress influences the neural mechanisms underlying the flexible use of encoded memories to infer A–C associations, we employed rapid neural reactivation of bridge information as a neural lens. We used two neural indices to capture the dynamics of memory reactivation: *Decodability* and *Mean Decoding Accuracy* (see Methods for details). *Decodability* reflects the temporal consistency of memory reactivation, with higher scores suggesting sustained activation during the inference phase. In contrast, *Mean Decoding Accuracy* quantifies the strength of memory reactivation by averaging the accuracy of decoding bridging information from brain activity, adjusted for chance levels; higher values indicate stronger retrieval of memory representations.

Initially, we provided neural evidence that stress impairs inference by disrupting the reactivation of bridge information. In Experiment 1, we compared *Decodability* and *Mean Decoding Accuracy* between stress (Stress-E1) and control participants (Control-E1). Acute stress tended to disrupt bridge information reactivation compared to the control condition.

This was indicated by a marginally significant reduction in Decodability measures (t(66)=−1.535, *p*=.065, one-tailed; Cohen's d=−0.372, 95% CI for [-0.854, 0.108]) and a similar, non-significant trend for Mean Decoding Accuracy (t(66)=−1.271, *p*=.104, one-tailed; Cohen's d=−0.308, 95% CI for [-0.789, 0.171]).To further support the notion that acute stress disrupts neural reactivation of bridge information, we correlated individual differences in acute stress responses (i.e., changes in HR and PA) with *Decodability* and *Mean Decoding Accuracy* measures. Stronger HR responses under stress predicted weaker subsequent neural reactivation of bridge information, as shown by both *Decodability* (r(62) = -0.338, *p* = 0.006) and *Mean Decoding Accuracy* (r(62) = -0.259, *p* = 0.039). Critically, HR—a rapid biological response to acute stress—had already returned to baseline during the inference phase, indicating that HR changes were measured prior to the reactivation of bridge information. This temporal ordering further strengthens the conclusion that acute stress causally impairs subsequent neural reactivation, as measured by EEG decoding. We also examined the associations between changes in PA and neural reactivation measures, which revealed a similar negative relationship that did not reach significance (Decodability: r(62) = -0.22, *p* = 0.08; Mean Decoding Accuracy: r(62) = -0.20, *p* = 0.10). To avoid interpretational issues with pooled data, we also conducted correlation analyses separately for each group and experiment.In Experiment 1, we found no correlation for either decodability (Stress: r=0.182, *p*=0.302; Control: r=0.117, *p*=0.509) or mean decoding amplitude (Stress: r=0.125, *p*=0.480; Control: r=0.037, *p*=0.835). This pattern was replicated in Experiment 2 for both decodability

(Stress: r=0.233, *p*=0.216; Control:r=0.045, *p*=0.814) and mean decoding amplitude (Stress: r=0.161, *p*=0.395; Control: r=0.061, *p*=0.751).

In Experiment 1, participants under stress were less likely to infer correct A–C associations. However, behavioral data from Experiment 2 indicate that retrieval practice of AB and BC associations can prevent this stress-induced impairment and enhance overall inference performance. If neural reactivation of bridge information contributes to successful inference and reflects the impact of stress, then it should be modulated by retrieval practice. Accordingly, we compared the decoded neural dynamics of bridge information during inference between participants who did and did not use retrieval practice (i.e., Experiment 1 vs. Experiment 2), irrespective of stress manipulation, to reveal the general effect of retrieval practice. We found that participants in Experiment 2 exhibited significantly stronger reactivation of bridge information than those in Experiment 1, as indexed by higher *Decodability* (t(126) = 1.995, *p* = 0.048, Cohen's d = 0.355, 95% CI = [0.076, 20.034]) and higher *Mean Decoding Accuracy* (t(126) = 2.099, *p* = 0.038, Cohen's d = 0.374, 95% CI = [0.162, 5.609]).

Subsequently, we examined participants in Experiment 2 who all underwent retrieval practice but received either stress (Stress-E2) or control (Control-E2) manipulations prior to inference, to elucidate why the memory inference process is resilient to stress following retrieval practice. We observed that participants in Experiment 2 who received stress not only avoided the stress-induced impairment in reactivation but also exhibited significantly stronger reactivation of bridge information compared to those receiving control manipulations, as

reflected in both higher Decodability (t(126) = 3.132, p = 0.003, Cohen's d = 0.809, 95% CI = [8.068, 36.972]) and higher Mean Decoding Accuracy (t(126) = 2.549, p = 0.014, Cohen's d = 0.658, 95% CI = [1.136, 9.513]).

Finally, as a complementary analysis, we integrated data from both experiments to formally test for an interaction, noting that Experiment 2 was designed after the results of Experiment 1 were known. For the decodability of the bridge element, the analysis revealed a significant main effect of retrieval practice (F(1,124)=4.374, p=0.039, η_p^2=0.034), while the main effect of stress was not significant (F(1,124)=1.226, p=0.270, η_p^2=0.010). Crucially, a significant interaction between acute stress and retrieval practice emerged (F(1,124)=11.344, p=0.001, η_p^2=0.084; Figure 6A). Post-hoc tests confirmed that stress significantly reduced decodability in the No-RP condition (p<.05) but not in the RP condition (p>.05), indicating that retrieval practice prevented the stress-induced disruption of reactivation. A similar pattern was observed for the mean decoding amplitude. The 2 × 2 ANOVA revealed a significant main effect of retrieval practice (F(1,124)=4.772, p=0.031, η_p^2=0.037) and no significant main effect of stress (F(1,124)=1.085, p=0.300, η_p^2=0.009). Again, the acute stress × retrieval practice interaction was significant (F(1,124)=7.925, p=0.006, η_p^2=0.060; Figure 6B), providing converging evidence that retrieval practice buffered the neural correlates of inference from stress.

3.5 Neural reactivation of bridge element during the choice phase predicts successful inference but is unaffected by stress or retrieval practice

We next investigated whether neural reactivation of the bridge element (B) persisted into the subsequent choice phase of the trial and, if so, whether it was modulated by stress or retrieval practice. Applying the same multivariate decoder to the choice phase—during which participants viewed the cue (A) alongside the potential target (C)—revealed significant reactivation of the bridge element across all conditions. Cluster-based permutation tests on the decoding time course identified two significant temporal clusters: an early cluster from 350–410 ms ($p<.05$) and a more sustained later cluster from 670–1460 ms ($p<.05$). Both clusters emerged well before the mean response time (M=1917 ms, SD=340 ms). To quantify this effect, we extracted the mean decoding accuracy within these time windows, confirming significant above-chance reactivation ($t(126)=2.18$, $p=0.01$ one-tailed, Cohen's d = 0.39). Consistent with its functional relevance, this choice-phase reactivation of the bridge element was significantly greater on successfully inferred trials compared to unsuccessful trials ($t(124) = 2.258$, $p = 0.01$, Cohen's d = 0.20, 95% CI =[0.02, 0.38]).

Crucially, in a key divergence from the initial inference period, this choice-phase reactivation was not modulated by acute stress or retrieval practice. We formally tested this using two complementary metrics. First, a 2 (Stress: Stress vs. No-Stress) × 2 (Session: Retrieval Practice vs. No-Retrieval Practice) ANOVA on overall decodability yielded no significant main effect of retrieval practice ($F(1,124)=1.49$, $p=.225$, $\eta_p^2=0.012$), no main effect of acute stress ($F(1,124)=0.31$, $p=.581$, $\eta_p^2=0.002$), and no interaction ($F(1,124)=0.01$, $p=.923$, $\eta_p^2<0.001$; Figure 6C). A second parallel one ANOVA on mean decoding accuracy yielded no significant main effect of retrieval practice ($F(1,124)=1.73$, $p=.191$, $\eta_p^2=0.014$), no

main effect of acute stress ($F(1,124)=1.26$, $p=.264$, $\eta_p^2=0.010$), and no interaction ($F(1,124)=0.02$, $p=.886$, $\eta_p^2<0.001$; Figure 6D). Exploratory simple effects analyses confirmed that stress did not significantly influence decoding accuracy in either the no-retrieval-practice condition ($F(1,66)=1.10$, $p=.299$, $\eta_p^2=0.016$) or the retrieval-practice condition ($F(1,58)=0.48$, $p=.493$, $\eta_p^2=0.008$). In sum, while bridge element reactivation occurs during the choice phase and predicts performance, this neural signal—unlike the rapid reactivation during the initial inference stage—is not susceptible to impairment by stress or enhancement by retrieval practice (Figure 6E).

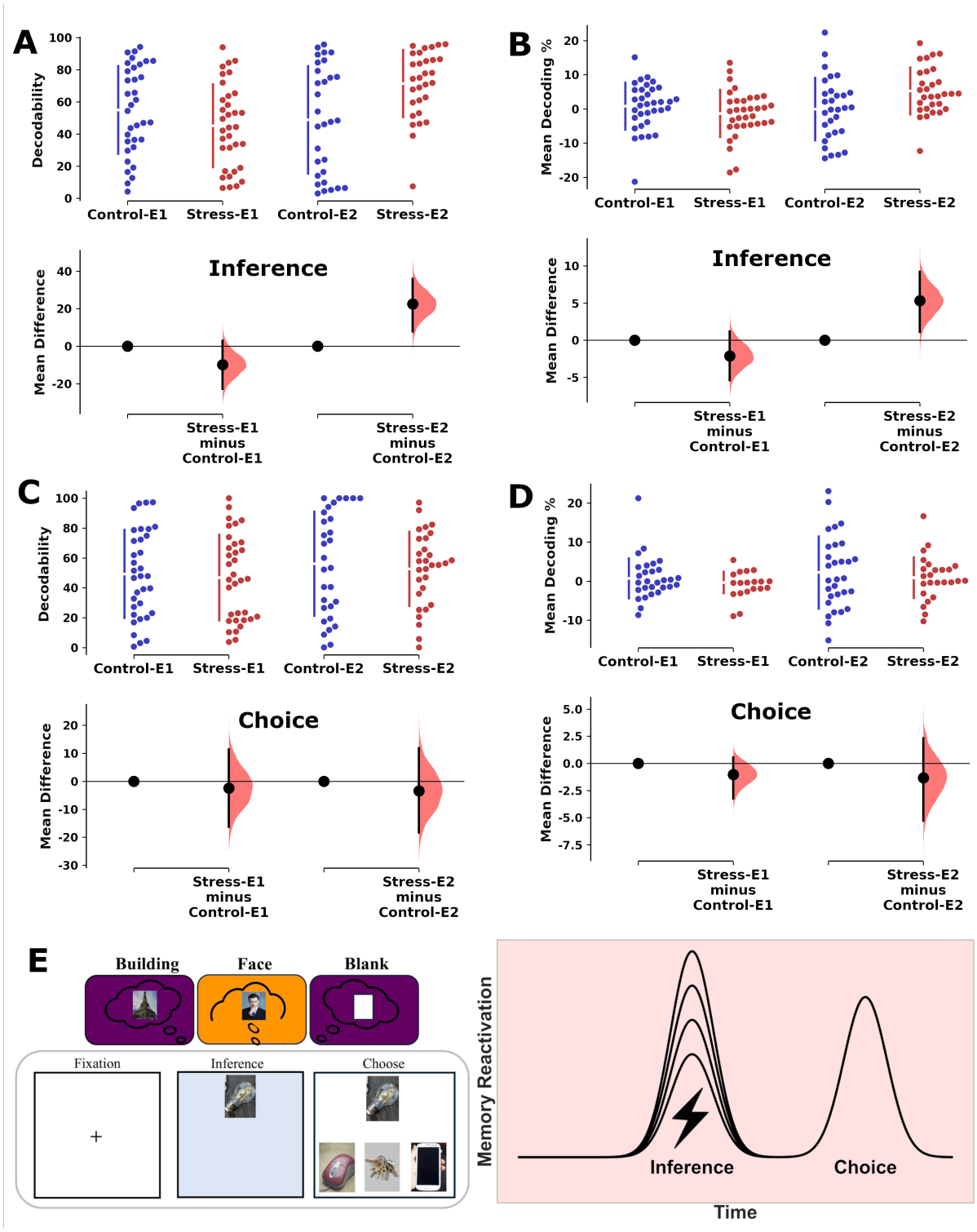

Figure 6 Retrieval practice reverses stress-induced impairment of neural reactivation during memory inference. (A) Classifier decodability for the bridge element during the inference period. In Experiment 1 (E1, no retrieval practice), the stress group exhibited lower decodability than the control group. This effect was reversed in Experiment 2 (E2, with retrieval practice), where the stress group showed significantly higher decodability than the

control group. (B) Mean decoding accuracy during the inference period showed the same pattern of results. (C, D) The interaction between stress and retrieval practice was temporally specific to the inference period, as no significant interaction was observed during the subsequent choice phase for either decodability (C) or mean decoding accuracy (D). (E) Schematic of the trial structure, illustrating the distinct inference and choice phases analyzed. These findings indicate that the modulatory effects of stress and retrieval practice on bridge element reactivation are confined to the inference process itself. E1= Experiment 1; E2= Experiment 2.

3.6 Theta oscillations analyses are consistent with behavioral and reactivation patterns

To complement our behavioral and neural reactivation findings, we performed time-frequency analyses focusing on theta oscillations to further elucidate the neural mechanisms underlying the effects of stress on memory inference. Given the established role of theta oscillations in memory integration, we hypothesized that theta power during the inference task would also be modulated by acute stress and that retrieval practice would mitigate this effect. In Experiment 1, a cluster-based permutation test comparing the control and stress groups revealed a significant stress-induced reduction in theta power (Control > Stress). This effect was isolated to a single late-emerging cluster, spanning from 1680 ms to 4000 ms relative to cue onset ($p = 0.009$; Figure 7A). The corresponding time-frequency map of T-values across an extended 2–12 Hz frequency range confirms this result, showing a concentration of positive T-values within the theta band during this significant time window, indicative of higher power in the control group (Figure 7B). In contrast, for Experiment 2, no significant clusters were found in the same comparison between control and stress groups,

suggesting that retrieval practice prevented the stress-induced suppression of theta power. To directly test the influence of retrieval practice, we performed a non-parametric cluster-based permutation test comparing participants from Experiment 2 (retrieval practice) with those from Experiment 1 (no retrieval practice). This analysis revealed that retrieval practice was associated with significantly higher theta power within an early time window, from 0 ms to 920 ms post-cue onset ($p < 0.001$; Figure 7C). Taken together, these results suggest that acute stress impairs theta-band activity associated with later stages of memory inference, and that retrieval practice can prevent this deficit by boosting theta-related processing during the initial phase of inference.

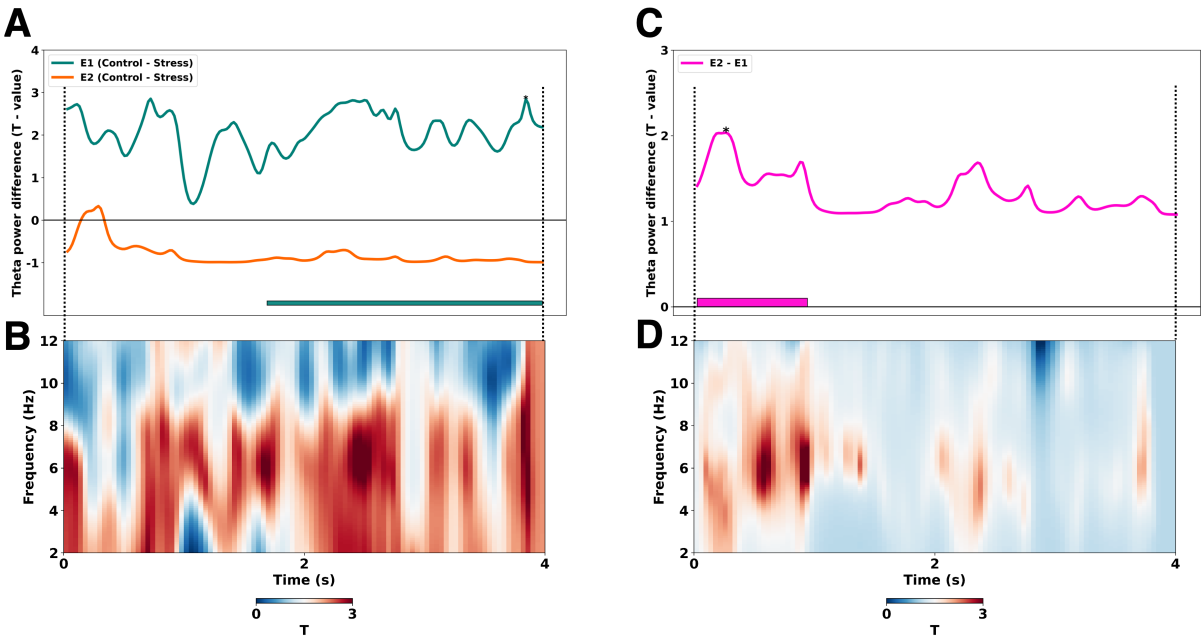

Figure7 Acute stress reduces theta power during memory inference, an effect counteracted by retrieval practice. (A) Time-resolved difference waves of theta power (Control – Stress) for Experiment 1 (E1, no retrieval practice; blue) and Experiment 2 (E2, with retrieval practice; orange). Thick horizontal bars indicate significant temporal clusters where the control group exhibited greater theta power than the stress group (cluster-based

permutation test, **p**<0.05). (B) Time-frequency representation of the t-statistic for the Control vs. Stress group contrast in E1, displayed for 2–12 Hz and 0–4 s post-cue. (C) The difference wave for the direct inter-experiment contrast (E2 – E1; magenta) reveals a significant increase in theta power attributable to retrieval practice. The thick horizontal bar marks the significant temporal cluster. (D) Corresponding time-frequency representation of the t-statistic for the E2 vs. E1 contrast. For all panels, time 0 indicates the onset of the bridge element B cue. Dashed vertical lines delineate the 4 s analysis window for memory inference. The colorbar applies to both t-statistic maps in (B) and (D). E1= Experiment 1; E2= Experiment 2.

Discussion

Our study reveals that acute stress impairs memory-based inference by disrupting the rapid neural reactivation of information that bridges related events. We demonstrate that this impairment is preventable through retrieval practice, which restores both behavioral performance and underlying neural reactivations. Using multivariate decoding of EEG data, we identified the rapid reactivation of the bridge element (B) as a critical predictor of successful inference (A-C). This reactivation process was specifically attenuated by stress but was robustly restored—and even enhanced—by retrieval practice. These findings provide a novel neuroscientific account of how memory inference is accomplished, clarify its vulnerability to stress, and highlight a powerful behavioral strategy for building stress resilience.

A central contribution of our work is the identification of rapid bridge-element reactivation as a core neural process for memory-based inference. While prior neuroimaging

research has focused on the roles of the hippocampus and prefrontal cortex in integrating memories during encoding (Zeithamova and Preston, 2010; Zeithamova et al., 2012; Backus et al., 2016; Morton et al., 2020; Schlichting et al., 2021), our work elucidates the neural dynamics of the inference phase itself. We reveal that successful A-C inference is preceded by a rapid neural reactivation of the bridging element B. This extends prior decoding studies of deliberate retrieval or imagery (Polyn et al., 2005; Chadwick et al., 2010; Xue, 2018; Dijkstra et al., 2019). as the reactivation observed here appears to be a spontaneous, goal-relevant event essential for linking discrete memories. One consideration is that participants were informed of the A-C link via B on Day 2, which could encourage a deliberate reactivation. However, as these instructions were identical across all groups, they cannot account for the observed interaction between stress and retrieval practice, nor for the trial-by-trial relationship between reactivation strength and inference success. Future research could employ more implicit paradigms to disentangle the contributions of spontaneous versus goal-directed reactivation to memory inference.

Our results extend the literature on stress-induced memory retrieval deficits (de Quervain et al., 1998; Kuhlmann et al., 2005; Smeets, 2011; Shields et al., 2017; Wolf, 2017) to the domain of more complex, constructive cognition. This result aligns with the theoretical framework proposing that stress promotes a shift away from flexible, hippocampus-dependent cognitive strategies toward more rigid, striatum-based systems (Knowlton et al., 1996; Kluen et al., 2017b, 2017a; Raio et al., 2017; Vogel et al., 2018; Schwabe et al., 2022; Zerbes et al., 2022). While EEG lacks the spatial resolution to definitively localize these

systems, our data provide crucial temporal precision, showing that stress disrupts inference at the precise moment of bridge-element reactivation. An alternative interpretation is that stress impairs the retrieval of A-C associations that were spontaneously formed during encoding (Zhou et al., 2023). However, two lines of evidence favor the online inference mechanism. First, the modulations of inference speed by stress and its reversal by retrieval practice point to an effect on the dynamic process itself. Second, the prospective, trial-by-trial prediction of inference success by B-element reactivation would be superfluous if a direct A-C association were simply being retrieved. Thus, the modulation of this specific neural signature by our experimental manipulations provides compelling evidence that acute stress targets the online inference mechanism.

Our study identifies retrieval practice (RP) as a potent countermeasure to the detrimental effects of acute stress on memory inference. Building on extensive evidence that RP strengthens memory (Karpicke and Blunt, 2011; Roediger and Butler, 2011; Liu et al., 2020; Ye et al., 2020; Yang et al., 2021; Zhuang et al., 2021; Carpenter et al., 2022), we show that practicing the AB and BC pairs rendered the novel A-C inference resilient to stress. T This generalization of the testing effect to a novel, integrative inference task contrasts with prior findings where its benefits are often contingent on congruence between practice and test formats (Smith et al., 2016; Brown et al., 2020). Notably, retrieval practice did not enhance inference performance in the non-stress condition; instead, its benefit was to build resilience against acute stress, maintaining performance at levels equivalent to the non-stress group. This specific pattern of results supports an "online" inference mechanism. If A-C associations

had been pre-compiled and stored during encoding, RP would likely have directly enhanced them, boosting performance regardless of stress. Therefore, we propose that the protective effect of RP transcends the reinforcement of individual associations. Instead, it fortifies the underlying relational memory structure, ensuring that the neural processes supporting inference—such as the rapid reactivation of the bridge element—remain robust against acute stress.

How does retrieval practice confer this stress resilience at the neural level? We propose that practicing retrieval of the AB and BC pairs serves to "tag" the entire ABC memory structure for subsequent offline consolidation (Dragoi and Tonegawa, 2011, 2013), potentially through co-reactivation during practice or subsequent sleep (Lutz et al., 2024). This mechanism is constrained by our data: retrieval practice did not significantly enhance A-C inference in the non-stress group. This specific pattern suggests that its benefit is not derived from strengthening a direct A-C link before the test. Instead, we argue that retrieval practice reinforces the relational integrity of the ABC structure, making the online reactivation of the bridge element more robust and thus more resilient to disruption by acute stress. Intriguingly, retrieval practice did more than simply restore bridge-element reactivation under stress; it upregulated this neural signal. The decoding strength for element B was highest in the RP-Stress group, significantly surpassing that of the RP-No Stress group. We interpret this heightened reactivation as a compensatory neural mechanism. When faced with the challenge of acute stress, the brain appears to amplify the critical neural process supporting inference. This interpretation aligns with broader theories of cognitive

effort, which posit that additional neural resources are recruited to preserve performance in situations such as aging (Reuter-Lorenz and Cappell, 2008), facing increased task difficulty (Gray et al., 2003), or physiological challenges such as stress (Hermans et al., 2014).

To supplement our primary multivariate decoding results from the inference phase, we conducted two additional analyses of the EEG data. First, we applied the same decoder to the choice phase following inference. While this analysis also detected neural reactivation of the bridge element that predicted subsequent choice accuracy, this reactivation was notably not modulated by acute stress. We propose that this later reactivation is cognitively distinct from that observed during the initial inference process, likely reflecting post-inferential mechanisms such as choice confirmation rather than the act of inference itself. This dissociation suggest the time-specific effect of acute stess: it selectively disrupts neural reactivation during the cognitively demanding period of memory inference. Secondly, guided by the established role of theta oscillations in human memory (Herweg et al., 2020), and particularly in associative inference (Backus et al., 2016), we analyzed how stress and retrieval practice modulated theta-band activity. Acute stress selectively impaired theta power during the later stages of memory inference. In contrast, retrieval practice not only prevented this stress-induced deficit but also enhanced theta-band activity during the initial phase of inference. The modulation of theta oscillations thus corroborates our neural reactivation findings. Crucially, these time-frequency results provide valuable temporal information, pinpointing a later disruptive effect of stress and an earlier facilitative effect of retrieval practice on the underlying neural mechanisms.

Our study is not without limitations, which highlight key directions for future investigation. First, although we used well-established stimulus categories (i.e., faces, scenes, objects), the generalizability of our findings could be further tested using materials with greater ecological validity or semantic complexity. Second, the limited spatial resolution of EEG prevented us from identifying the precise anatomical sources of the reactivation signals. Future studies employing methods with higher spatial resolution are needed to elucidate the network-level interactions between the hippocampus, prefrontal cortex, and other relevant regions during stress-impaired inference. Third, while our findings demonstrate a clear protective benefit of retrieval practice, a valuable direction for future research would be to include a re-study control group. Such a design would allow for a precise dissociation of the cognitive mechanisms attributable to active retrieval versus those stemming from mere re-exposure to the material, further clarifying the unique role of retrieval in building stress-resilient inferential memories. Nonetheless, a substantial body of literature has established the superiority of retrieval practice over restudy for enhancing long-term retention (Roediger and Butler, 2011; Roediger and Abel, 2022) and building stress-resilient memories (Smith et al., 2016), providing a strong rationale for our interpretation. Finally, pandemic-related safety protocols precluded cortisol collection, and future studies should integrate hormonal measures to provide a more comprehensive characterization of the stress response.

In conclusion, our study demonstrates that acute stress impairs memory inference by disrupting the rapid neural reactivation of the shared bridge element that links related events. We identify this rapid reactivation as a critical neural computations underlying successful

inference and pinpoint its disruption as a key mechanism through which stress impairs higher-order memory function. Critically, we show that retrieval practice can effectively buffer against this stress-induced deficit. This protection is associated with the preservation and even enhancement of the underlying rapid reactivation signals. These findings elevate the role of retrieval practice from a simple tool for enhancing memory to a proactive strategy for building cognitive resilience, rendering complex memory functions against stress.

Reference

Antony JW, Ferreira CS, Norman KA, Wimber M (2017) Retrieval as a Fast Route to Memory Consolidation. Trends Cogn Sci 21:573–576.

Backus AR, Schoffelen J-M, Szebényi S, Hanslmayr S, Doeller CF (2016) Hippocampal-Prefrontal Theta Oscillations Support Memory Integration. Current Biology 26:450–457.

Brown TI, Gagnon SA, Wagner AD (2020) Stress Disrupts Human Hippocampal-Prefrontal Function during Prospective Spatial Navigation and Hinders Flexible Behavior. Current Biology 30:1821-1833.e8.

Buchanan TW, Laures-Gore JS, Duff MC (2014) Acute stress reduces speech fluency. Biol Psychol 97:60–66.

Buchanan TW, Tranel D, Adolphs R (2006) Impaired memory retrieval correlates with individual differences in cortisol response but not autonomic response. Learning & Memory 13:382–387.

Carpenter SK, Pan SC, Butler AC (2022) The science of effective learning with spacing and retrieval practice. Nature Reviews Psychology 1:496–511.

Chadwick MJ, Hassabis D, Weiskopf N, Maguire EA (2010) Decoding Individual Episodic Memory Traces in the Human Hippocampus. Current Biology 20:544–547.

Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nat Neurosci 17:455–462.

De Quervain DJ -F., Henke K, Aerni A, Treyer V, McGaugh JL, Berthold T, Nitsch RM, Buck A, Roozendaal B, Hock C (2003) Glucocorticoid-induced impairment of declarative memory retrieval is associated with reduced blood flow in the medial temporal lobe. European Journal of Neuroscience 17:1296–1302.

de Quervain DJ-F, Roozendaal B, McGaugh JL (1998) Stress and glucocorticoids impair retrieval of long-term spatial memory. Nature 394:787–790.

Dijkstra N, Bosch SE, van Gerven MAJ (2019) Shared Neural Mechanisms of Visual Perception and Imagery. Trends Cogn Sci 23:423–434.

Dragoi G, Tonegawa S (2011) Preplay of future place cell sequences by hippocampal cellular assemblies. Nature 469:397–401.

Dragoi G, Tonegawa S (2013) Distinct preplay of multiple novel spatial experiences in the rat. Proceedings of the National Academy of Sciences 110:9100–9105.

Gagnon SA, Waskom ML, Brown TI, Wagner AD (2019) Stress Impairs Episodic Retrieval by Disrupting Hippocampal and Cortical Mechanisms of Remembering. Cerebral Cortex 29:2947–2964.

Gray JR, Chabris CF, Braver TS (2003) Neural mechanisms of general fluid intelligence. Nat Neurosci 6:316–322.

Grootswagers T, Wardle SG, Carlson TA (2017) Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. J Cogn Neurosci 29:677–697.

Hebart MN, Dickter AH, Kidder A, Kwok WY, Corriveau A, Van Wicklin C, Baker CI (2019) THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. PLoS One 14:e0223792.

Hermans EJ, Henckens MJAG, Joëls M, Fernández G (2014) Dynamic adaptation of large-scale brain networks in response to acute stressors. Trends Neurosci 37:304–314.

Herweg NA, Solomon EA, Kahana MJ (2020) Theta Oscillations in Human Memory. Trends Cogn Sci 24:208–227.

Isik L, Meyers EM, Leibo JZ, Poggio T (2014) The dynamics of invariant object recognition in the human visual system. J Neurophysiol 111:91–102.

J-F de Quervain D, Aerni A, Benno Roozendaal B (2007) Preventive Effect of β-Adrenoceptor Blockade on Glucocorticoid-Induced Memory Retrieval Deficits.

Karpicke JD, Blunt JR (2011) Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. Science (1979) 331:772–775.

Kirschbaum C, Pirke K-M, Hellhammer DH (1993) The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. Neuropsychobiology 28:76–81.

Kluen LM, Agorastos A, Wiedemann K, Schwabe L (2017a) Noradrenergic Stimulation Impairs Memory Generalization in Women. J Cogn Neurosci 29:1279–1291.

Kluen LM, Nixon P, Agorastos A, Wiedemann K, Schwabe L (2017b) Impact of Stress and Glucocorticoids on Schema-Based Learning. Neuropsychopharmacology 42:1254–1261.

Knowlton BJ, Mangels JA, Squire LR (1996) A Neostriatal Habit Learning System in Humans. Science (1979) 273:1399–1402.

Kuhlmann S, Piel M, Wolf OT (2005) Impaired Memory Retrieval after Psychosocial Stress in Healthy Young Men. The Journal of Neuroscience 25:2977–2982.

Linde-Domingo J, Treder MS, Kerrén C, Wimber M (2019) Evidence that neural information flow is reversed between object perception and object reconstruction from memory. Nat Commun 10:179.

Liu W, Kohn N, Fernández G (2020) Probing the neural dynamics of mnemonic representations after the initial consolidation. Neuroimage 221:117213.

Lu Z, Ku Y (2020) NeuroRA: A Python Toolbox of Representational Analysis From Multi-Modal Neural Data. Front Neuroinform 14.

Lutz ND, Martínez-Albert E, Friedrich H, Born J, Besedovsky L (2024) Sleep shapes the associative structure underlying pattern completion in multielement event memory. Proceedings of the National Academy of Sciences 121.

Morton NW, Schlichting ML, Preston AR (2020) Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. Proceedings of the National Academy of Sciences 117:29338–29345.

Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-Specific Cortical Activity Precedes Retrieval During Memory Search. Science (1979) 310:1963–1966.

Preston AR, Eichenbaum H (2013) Interplay of Hippocampus and Prefrontal Cortex in Memory. Current Biology 23:R764–R773.

Raio CM, Hartley CA, Orederu TA, Li J, Phelps EA (2017) Stress attenuates the flexible updating of aversive value. Proceedings of the National Academy of Sciences 114:11241–11246.

Reuter-Lorenz PA, Cappell KA (2008) Neurocognitive Aging and the Compensation Hypothesis. Curr Dir Psychol Sci 17:177–182.

Roediger HL, Abel M (2022) The double-edged sword of memory retrieval. Nature Reviews Psychology Available at: https://www.nature.com/articles/s44159-022-00115-2.

Roediger HL, Butler AC (2011) The critical role of retrieval practice in long-term retention. Trends Cogn Sci 15:20–27.

Schlichting ML, Guarino KF, Roome HE, Preston AR (2021) Developmental differences in memory reactivation relate to encoding and inference in the human brain. Nat Hum Behav 6:415–428.

Schlichting ML, Preston AR (2015) Memory integration: neural mechanisms and implications for behavior. Curr Opin Behav Sci 1:1–8.

Schwabe L, Hermans EJ, Joëls M, Roozendaal B (2022) Mechanisms of memory under stress. Neuron 110:1450–1467.

Schwabe L, Römer S, Richter S, Dockendorf S, Bilak B, Schächinger H (2009) Stress effects on declarative memory retrieval are blocked by a β-adrenoceptor antagonist in humans. Psychoneuroendocrinology 34:446–454.

Shields GS, Sazma MA, McCullough AM, Yonelinas AP (2017) The effects of acute stress on episodic memory: A meta-analysis and integrative review. Psychol Bull 143:636–675.

Smeets T (2011) Acute stress impairs memory retrieval independent of time of day. Psychoneuroendocrinology 36:495–501.

Smith AM, Floerke VA, Thomas AK (2016) Retrieval practice protects memory against acute stress. Science (1979) 354:1046–1048.

Smith AM, Thomas AK (2018) Reducing the consequences of acute stress on memory retrieval. J Appl Res Mem Cogn 7:219–229.

Vogel S, Kluen LM, Fernández G, Schwabe L (2018) Stress affects the neural ensemble for integrating new information and prior knowledge. Neuroimage 173:176–187.

Vogel S, Schwabe L (2016) Learning and memory under stress: implications for the classroom. NPJ Sci Learn 1.

von Dawans B, Fischbacher U, Kirschbaum C, Fehr E, Heinrichs M (2012) The Social Dimension of Stress Reactivity. Psychol Sci 23:651–660.

Wolf OT (2017) Stress and memory retrieval: mechanisms and consequences. Curr Opin Behav Sci 14:40–46.

Xue G (2018) The Neural Representations Underlying Human Episodic Memory. Trends Cogn Sci 22:544–561.

Yang C, Luo L, Vadillo MA, Yu R, Shanks DR (2021) Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. Psychol Bull 147:399–435.

Ye Z, Shi L, Li A, Chen C, Xue G (2020) Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations. Elife 9.

Zeithamova D, Dominick AL, Preston AR (2012) Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. Neuron 75:168–179.

Zeithamova D, Preston AR (2010) Flexible Memories: Differential Roles for Medial Temporal Lobe and Prefrontal Cortex in Cross-Episode Binding. The Journal of Neuroscience 30:14676–14684.

Zerbes G, Kausche FM, Schwabe L (2022) Stress-induced cortisol modulates the control of memory retrieval towards the dorsal striatum. European Journal of Neuroscience 55:2699–2713.

Zhou Z, Singh D, Tandoc MC, Schapiro AC (2023) Building integrated representations through interleaved learning. J Exp Psychol Gen 152:2666–2684.

Zhuang L, Wang J, Xiong B, Bian C, Hao L, Bayley PJ, Qin S (2021) Rapid neural reorganization during retrieval practice predicts subsequent long-term retention and false memory. Nat Hum Behav 6:134–145.