Brain and Cognition xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Retrieval practice improves item memory but not source memory in the context of stress

Amy M. Smith[a,*], Elizabeth Race[b], F. Caroline Davis[a,c], Ayanna K. Thomas[b]

[a] Center for Applied Brain and Cognitive Sciences, Tufts University, 200 Boston Avenue, Suite 3000, Medford, MA 02155, United States
[b] Department of Psychology, Tufts University, 490 Boston Ave, Medford, MA 02155, United States
[c] U.S. Army Natick Soldier Research, Development, and Engineering Center (NSRDEC), General Greene Avenue, Natick, MA 01760, United States

ARTICLE INFO

Keywords:
Stress
Cortisol
Retrieval practice
Testing effect
Context

ABSTRACT

Smith, Floerke, and Thomas (2016) demonstrated that learning by repeated testing, or retrieval practice, reduced stress-related memory impairment when compared to learning by repeatedly studying material. In the present experiment, we tested whether, relative to study practice, retrieval practice would improve post-stress memory by increasing access to both item and source information. Participants learned two wordlists, which were temporally segregated to facilitate distinction between the two lists. Participants returned one week later for stress induction and two memory tests. Each test featured a recognition test that was given to assess item memory accessibility, and a list-discrimination task that was given to assess source memory. Relative to study practice, successful retrieval practice during learning reduced false alarms but did not improve source memory on the post-stress test. Results are discussed as they relate to current theories surrounding stress effects and retrieval practice effects.

1. Introduction

Memory retrieval is generally impaired when preceded by a psychological stressor that causes a marked increase in the stress hormone cortisol (Buchanan, Tranel, & Adolphs, 2006; de Quervain, Roozendaal, & McGaugh, 1998; Kuhlmann, Piel, & Wolf, 2005; Shields, Sazma, McCullough, & Yonelinas, 2017). Though the majority of studies on the topic have replicated this effect, recent work has highlighted the potential for learning techniques to effectively create stress-resistant memories. Specifically, the act of taking practice tests when learning new material, also known as *retrieval practice*, has been shown to greatly improve post-stress memory performance relative to conventional study techniques (Smith et al., 2016; Smith, Davis, & Thomas, 2018). In the present experiment, we aimed to elucidate the mechanism underlying the efficacy of retrieval practice in the context of stress. Specifically, we examined the influences of stress and retrieval practice on both item and source memory, an approach that was informed by commonalities in the theories surrounding stress effects and retrieval practice effects.

Implicated in the detrimental effects of stress on memory retrieval is the physiological stress response. Cortisol that is released from the adrenal cortex during a stressful event binds to the amygdala and hippocampus (Lovallo, Robinson, Glahn, & Fox, 2010; Reul & de Kloet, 1985). When combined with increased levels of catecholamines, cortisol's occupation of these brain regions interferes with retrieval-related neural processing (de Quervain, Aerni, & Roozendaal, 2007; Roozendaal, Hahn, Nathan, de Quervain, & McGaugh, 2004). Theoretical models of memory have attempted to further specify the cognitive processes that may be disrupted by stress. One current theory posits that stress induces a "memory formation mode" in the brain, in which cortisol and catecholamines cause upregulation of neural networks involved in learning and consolidation, and downregulation of networks involved in retrieval (Schwabe, Joels, Roozendaal, Wolf, & Oitzl, 2012). Another explanation is that stress burdens the executive functions that are necessary for individuals to engage in careful, effortful recollection, thus resulting in post-stress memory impairment (Gagnon & Wagner, 2016). Finally, in their recent meta-analysis, Shields et al. (2017) suggested another, potentially complementary, hypothesis: stress may induce a mental context shift, causing a disruption to context-dependent memory. In brief, context dependence refers to the robust finding that retrieval of information is facilitated by reinstating the circumstances in which that information was learned (for a review see Smith & Vela, 2001). These circumstances can be externally manipulated, such as by changing the noise level in a room, or internally manipulated, such as through inducing different moods at times of learning and testing. The switch from a calm to a stressed mental state may help explain why psychological stress impairs memory retrieval. Further, because cortisol

* Corresponding author.
E-mail address: amy.smith@tufts.edu (A.M. Smith).

https://doi.org/10.1016/j.bandc.2018.12.005
Received 13 July 2018; Received in revised form 13 December 2018; Accepted 15 December 2018

interferes with hippocampal processing (Henckens et al., 2012) and the hippocampus is necessary for episodic, context-dependent recollection (Rugg et al., 2012), stress may result in a neural inability to reinstate one's learning context. Thus, stress-induced memory disruptions may result from changes in mood and neural changes that render contextual information inaccessible. The latter mechanism may explain why pharmacological administration of corticosterone also results in substandard memory retrieval (Het, Ramlow, & Wolf, 2005).

Studies examining the effects of pre-retrieval stress on memory have provided evidence for a context-shift mechanism. In one experiment (Schwabe & Wolf, 2009), participants underwent stress induction or a control task, and then completed a memory test in either the same context in which learning had occurred (a room scented with vanilla) or an unfamiliar context (a different room with no vanilla scent). Non-stressed participants performed similarly regardless of context changes. Among those who were stressed, memory was impaired when they were tested in an unfamiliar context but not when they were tested in the learning context. Thus, stress in combination with the external context shift yielded a context mismatch that was substantial enough to impair memory. However, providing multiple sources of external contextual support during retrieval alleviated this issue. These findings therefore suggest that stress can create a context shift that can be overridden by adequate contextual support. Additional support comes from a recent study in which acute, pre-retrieval stress impaired memory for previously-learned words as well as the images that were learned in association with each word (Goldfarb, Tompary, Davachi, & Phelps, 2018). These results suggest that stress may impair both memory for items and memory for the context associated with items during initial learning.

More support for the context-shift hypothesis comes from research examining the effects of stress on the consolidation phase of memory. In these studies, researchers induce stress immediately after initial learning and typically report memory-enhancing effects (see Shields et al., 2017). One possible explanation for this is that stress induced after encoding creates a mental context shift that isolates the encoded information from new information that is learned during the interval between encoding and retrieval. Without this isolation of the encoded information, individuals could experience retroactive interference, in which their memories for the encoded information are blocked by memories of information learned after encoding. Indeed, context disruptions have been shown to result in less retroactive interference than maintaining the same context between encoding and retrieval (Strand, 1970). Further evidence from the consolidation literature comes from a recent meta-analysis in which stress during consolidation was found to benefit memory only when the stressor and consolidation phase occurred in the original learning context (Shields et al., 2017). When a context shift occurred after encoding (e.g., participants left the laboratory), stress at consolidation no longer had a noticeable impact on performance. Thus, stress at consolidation appears to have a positive impact on memory only in circumstances in which a post-encoding context shift is not otherwise provided. When a post-encoding context shift occurs via other means, any benefits of stress are no longer observed. In summary, several studies and emerging trends in the stress-and-memory literature have begun to build a case for a context-shift mechanism as the underlying cause of the observed effects of stress on consolidation and retrieval.

The design of the present study was guided by the context-shift hypothesis because this account relies on the same memory mechanism (i.e., memory for context) as the leading theory on the efficacy of retrieval practice. The *episodic-context account* suggests that the value of retrieval practice stems from each successive retrieval attempt being made in a novel context (e.g., Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010; Karpicke & Zaromb, 2010; Whiffen & Karpicke, 2017). For example, a given retrieval attempt may occur at a different time, in a different physical location, and/or while an individual is in a different mental state than earlier attempts. Thus, during retrieval practice, each retrieval attempt updates a given memory by both reinstating the original study context and associating with the memory new contextual information from the present moment. On a final memory test, relative to study practice, retrieval practice equips an individual with a more recent memory of the original study context and additional contextual cues for guiding his or her memory search. If a studied item is not readily retrieved, these additional contextual cues may help the individual successfully remember the item.

The episodic-context account is supported by studies in which retrieval practice has been shown to improve memory for contextual information (for a review see Karpicke, Lehman, & Aue, 2014). For example, Brewer et al. (2010) had participants study two lists of words and then either engage in free recall after each list was presented or perform a time-matched distractor task. On a final test, participants were presented with the words and were asked to indicate the list (i.e., context) each item came from. Those who had engaged in retrieval practice of the lists demonstrated better list discrimination than those who had not, supporting the notion that retrieval practice updates a given memory with cues from the context in which the memory was initially acquired. Another test of the episodic-context account was carried out by Whiffen and Karpicke (2017). In their experiment, participants first studied two lists of words. They were then re-presented with the words in one mixed list and were instructed either to restudy the items or to indicate whether each item came from the first or second list. Note that in the latter group, participants incidentally engaged in retrieval practice for information regarding the context in which items were learned. On a final test of free recall, participants who performed list discrimination recalled more of the studied words than those who restudied the lists. These findings support a major prediction of the episodic-context account: retrieving contextual information about a learning event enhances subsequent memory for that event.

To summarize, current theories regarding the influence of stress on memory and the influence of retrieval practice on memory are based on the relationship between context and memory retrieval. Stress is believed to shift mental context, thereby disrupting context-dependent retrieval processes. Retrieval practice is believed to increase memory accessibility by enhancing access to contextual details that help reinstate context during retrieval. The context mechanism shared by these theories leads us to a testable hypothesis: stress disrupts memory for contextual information from the learning context, and that retrieval practice improves post-stress memory by increasing access to these contextual details.

To test this hypothesis, we used a list-discrimination task, as is commonly used in studies on context-dependent memory (e.g., Brewer et al., 2010; Chan & McDermott, 2007; Whiffen & Karpicke, 2017). Participants learned two wordlists that were separated by a distinct, 30-min interval. During learning, participants either studied the list multiple times or studied it once and then completed two free-recall tests. A week later, participants completed pre- and post-stress tests that featured two components: a recognition test that assessed memory for each item, and a list-discrimination task that assessed source memory and source-memory confidence for each item. The list-discrimination task assesses memory for episodic context: Can participants differentiate between their memories of two events that occurred at distinctly different times?

We hypothesized that both recognition accuracy and list discrimination accuracy would be most impaired for stressed individuals in the study practice (SP) group, followed by non-stressed SP, stressed retrieval practice (RP), and non-stressed RP, respectively. Our hypothesis regarding recognition accuracy was based on results from previous studies (Smith et al., 2016, 2018), and our hypothesis regarding performance on the list-discrimination task was based on the premise that stress impairs access to contextual information and that retrieval practice improves this access. Our examination of source confidence was exploratory because, to our knowledge, no other researchers have examined the influence of pre-retrieval stress on memory confidence.

We hoped that this investigation would provide initial evidence about the relationship between stress at retrieval and post-retrieval monitoring processes. Given that episodic memory retrieval is generally impaired by stress but enhanced by retrieval practice, we expected that one's confidence in their memory for source information would be reduced by stress but increased by retrieval practice.

2. Method

2.1. Design

We employed a 2 (learning strategy: SP or RP) × 2 (test timing: pre-stress or post-stress) mixed factorial design. Learning strategy was manipulated between subjects. Test timing was manipulated within subjects, such that all participants were subjected to stress induction. Non-stressed and stressed recognition performance was measured by pre- and post-stress tests.

2.2. Participants

Assuming an effect size of $\eta_p^2 = 0.04$ derived from Smith et al. (2016), a significance level of $\alpha = 0.05$, two between-subjects groups, two within-subjects measurements, and a conservative 0.70 correlation between repeated measures (in Smith et al., 2018, the correlation between Test 1 and Test 2 performance for non-stressed participants was 0.90), an a priori power analysis revealed that a total sample size of 50 participants ($n = 25$ per group) would provide 99% power to detect effects (G*Power 3.0; Faul, Erdfelder, Lang, & Buchner, 2007). In anticipation of participant error or dropout, 71 Tufts University students were recruited to participate in the experiment. Three participants were excluded from data analysis because they did not return for the second experimental session, and six participants were excluded because they demonstrated higher false alarm rates than hit rates on one or both of the recognition tests. Thus, all final analyses were conducted on 62 participants (45 women,[1] $M_{age} = 18.66$, $SD_{age} = 0.85$), all of whom reported that they had not consumed caffeine or nicotine in the 6 hr prior to the experiment. Though we did not screen for color blindness, no participants reported being colorblind when presented with color-relevant tasks. All participants were recruited through introductory psychology courses to fulfill a research participation requirement. Thirty participants were randomly assigned to the SP group and 32 participants were assigned to the RP group. All participants provided informed consent.

2.3. Materials

2.3.1. Stimuli

The stimuli consisted of three 60-item wordlists. The items from two of the wordlists served as stimuli (henceforth List 1 and List 2) and the items from the third list served as foils on the recognition test. Words were borrowed from the South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998). Each word met the following criteria: (1) non-proper noun, (2) not a homograph, (3) four to eight letters long, and (4) concreteness rating of at least 4 on a scale from 1 to 7 (7 = most concrete). Words were further compared to the valence and arousal norms established by Warriner, Kuperman, and Brysbaert (2013). Words were chosen only if they had valence ratings between 4.00 and 5.99 on a 1–9 scale (i.e., were of neutral valence) and were given arousal ratings lower than 4.00 on a 1–9 scale (i.e., were not negatively arousing). Word frequencies were determined from the Brysbaert and New (2009) norms, and word frequency was equated across the three wordlists.

2.3.2. Recognition tests

Two 90-item recognition tests were constructed (henceforth Test 1 and Test 2). Each test contained 30 items from List 1, 30 items from List 2, and 30 items from the foil list. Which 30 words from each list appeared on Test 1 and Test 2 was counterbalanced. Words were presented individually, and the presentation order was randomized for each participant. Upon presentation of each word, participants were prompted to indicate whether they had or had not learned the word during the first experimental session by pressing the L or A key, respectively. After each word was presented, and before advancing to the next word, participants were asked to indicate whether the item came from the first list they learned by pressing the '1' key, the second list they learned by pressing '2' or neither list by pressing '3'. Additionally, each list-discrimination question was accompanied by a subsequent confidence judgment. When participants indicated whether each recognized item came from List 1 or List 2, they also indicated whether their judgment was made with high or low confidence. To summarize, a recognition trial consisted of a judgment of whether the item was previously studied followed by a list-discrimination judgment and then a confidence judgment. If participants indicated that they had not previously studied an item, they advanced to the next trial.

2.3.3. State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA)

The STICSA (Grös, Antony, Simms, & McCabe, 2007) was administered to assess participants' self-reported levels of pre- and post-stress anxiety. STICSA scores range from 0 to 80 and higher scores are indicative of higher self-reported anxiety.

2.4. Procedure

Testing sessions occurred on two days with a one-week delay between Day 1 and Day 2 testing. We chose a one-week retention interval because we have found that participant retention rates are better with a one-week delay than with other common delays (e.g., 24 h). Additionally, this interval was successfully implemented in one of the studies that served as an impetus for the present experiment (Smith et al., 2018). All sessions took place in the afternoon to control for variability in diurnal cortisol secretion (e.g., Weitzman et al., 1971). Participants were tested in groups of two. In two cases in which only one participant showed up for the experiment, a research assistant served as a confederate.

Participants were first instructed that they would be presented with a series of words that they should try to remember for a later test. The words from either List 1 or List 2 (counterbalanced) were randomly presented at a rate of 2 s per word. Participants then either restudied the list twice in the same manner (SP group) or completed two time-matched (i.e., 2 min) free-recall tests for the list (RP group). During free recall, participants typed their responses and the responses were recorded by the E-Prime program. Between each study or test event, participants performed simple math problems for 30 s. During a subsequent 30-min break, participants viewed an episode of the BBC television series *Planet Earth*. This delay between learning the two wordlists has been shown to be necessary for avoiding chance-level performance on the subsequent list-discrimination task (e.g., Zeeuws, Deroost, & Soetens, 2010). Once 30 min had passed, participants learned the second wordlist in the same manner as the first list. That is, those in the study-practice group studied the 60-item list three times and those in the retrieval-practice group studied it once and then engaged in free recall for the second list during two 2-min recall periods. To promote participants' ability to discriminate between the two lists, the first list presented was referred to as the "Red" list and the second list was referred to as the "Blue" list. The words from each of the lists

[1] Women sometimes demonstrate a blunted cortisol response to psychological stress, particularly when in the follicular phase of their menstrual cycle or when taking oral contraceptives (Kajantie & Phillips, 2006). We opted to not control for these variables in order to gather a representative and generalizable pattern of data.

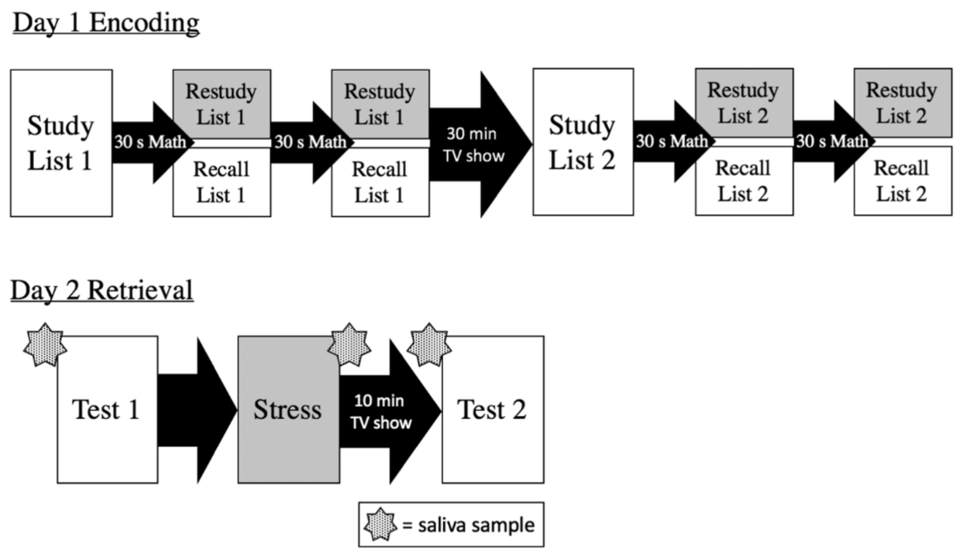

Fig. 1. Graphic representation of the experimental procedure.

were presented in those respective font colors. Last, prior to excusal on Day 1, all participants completed a STICSA. Please see Fig. 1 for a graphic depiction of the experimental procedure.

One week later, participants returned to the original testing room where they completed a second STICSA and provided a baseline saliva sample for cortisol analysis. All participants then completed Test 1. They next performed all tasks associated with the Trier Social Stress Test for Groups (TSST-G; von Dawans, Kirschbaum, & Heinrichs, 2011), which consisted of 2 min of speech preparation, 2 min each of speech delivery (4 min total), and 6 min of oral math subtraction tasks. Participants then completed the third STICSA and provided the second saliva sample, which was taken approximately 12 min after the onset of the TSST-G. During a subsequent 10-min retention interval, participants viewed part of an episode of *The Office*. Afterward, they provided the third saliva sample, which occurred approximately 25 min after the onset of the TSST-G. Last, participants completed Test 2 prior to being debriefed and excused.

All encoding tasks, videos, and recognition tests were presented using E-Prime software (Version 2.1; Schneider, Eschman, & Zuccolotto, 2001). The experimental procedure on each day lasted approximately 45 min.

2.5. Cortisol measurement and data management

Three saliva samples were collected using the passive drool method: one at baseline and one each at 12 min and 25 min after the onset of stress. Samples were stored at −20 °C until the completion of data collection, after which they were shipped to Salimetrics, LLC (Salimetrics, LLC, State College, PA) for analysis. Samples were assayed in duplicate, and the mean cortisol concentration served as the dependent measure. Cortisol concentrations were converted from µg/dL to nmol/L for consistency with the majority of human stress literature. Cortisol data for two participants (both in the SP group) were excluded from analysis because values exceeded 140 nmol/L. These values were deemed outliers given that the average cortisol concentration across all participants and samples was 5.96 nmol/L ($SD = 4.17$).

2.6. Dependent measures and statistical analyses

As outlined below, we computed five dependent measures: hits, false alarms, source memory scores, average confidence, and gamma correlations. In addition to conducting mixed-model ANOVAs on each of these measures, we conducted two-tailed, bivariate correlations (Pearson's r) to examine the relationship between participants' cortisol reactivity to stress (delta cortisol) and their memory performance. Delta cortisol was calculated by subtracting each participant's baseline cortisol concentration from the cortisol concentration of the sample taken 25 min post-stress. Alpha was set at 0.05 for all analyses.

2.6.1. Item memory

To examine item memory for previously studied items from Lists 1 and 2, we calculated hit proportions and false alarm proportions. We calculated false alarm proportions by dividing the number of non-presented foil words that each participant falsely recognized by the total number of foils presented on the recognition test. Hit proportions were calculated separately for the SP and RP groups. For those in the SP group, we divided the number of correctly recognized studied items by the total number of studied items that occurred on the recognition test for each participant. For those in the RP group, we first restricted recognition responses to items that they had accurately recalled at least once during their retrieval practice attempts on Day 1, and then performed the same calculation. This restriction on hits for the RP group was informed by previous research (Maddox & Balota, 2015; Mulligan, Susser, & Smith, 2016), and was conducted for the following reason. The benefits of retrieval practice over study practice on a final criterial test are contingent on successful retrieval during retrieval practice attempts. For instance, an individual who recalls only 2 of 60 items during the retrieval practice phase would not demonstrate exceptional memory performance when tested at a later time, whereas an individual who recalls 50 of the items likely would. Thus, for a more meaningful examination of the benefits of retrieval practice, we calculated hit proportions for individuals in the RP group after restricting recognition responses to items that they accurately recalled on Day 1. On average, this restriction resulted in eliminating 45 of the 60 previously studied items on Test 1, and 45 of the 60 items on Test 2. Though our analyses were conducted using this restriction, we present both restricted and unrestricted means and standard errors in Table 1.

2.6.2. Source memory

To examine source memory, we divided the total number of hits attributed to the correct source by the total number of hits for each participant. As with the calculation of hit proportions, source memory scores for the RP group were calculated after restricting their hits to items that they accurately recalled on Day 1. Restricting responses for those in the RP group in this manner provides a clear answer to the question of whether retrieval practice, which involves successful

Table 1
Pre-stress and post-stress performance on all item and source memory measures for the study practice (SP) and retrieval practice (RP) groups. For the RP group, metrics for both the restricted and unrestricted groups are provided. Standard errors of the mean are given in parentheses.

	Item Memory		Source Memory		
	Hits	False Alarms	Source Accuracy	Gamma Correlation	Confidence
Pre-stress					
SP	0.66 (*0.02*)	0.34 (*0.03*)	0.57 (*0.02*)	0.20 (*0.10*)	0.36 (*0.21*)
RP	0.91 (*0.02*)	0.34 (*0.03*)	0.52 (*0.03*)	0.05 (*0.07*)	0.64 (*0.22*)
$RP_{UNRESTRICTED}$	0.59 (*0.02*)	0.34 (*0.03*)	0.54 (*0.02*)	0.04 (*0.07*)	0.37 (*0.18*)
Post-stress					
SP	0.66 (*0.03*)	0.37 (*0.04*)	0.56 (*0.02*)	0.10 (*0.09*)	0.34 (*0.23*)
RP	0.91 (*0.01*)	0.30 (*0.03*)	0.55 (*0.03*)	0.26 (*0.09*)	0.63 (*0.22*)
$RP_{UNRESTRICTED}$	0.56 (*0.03*)	0.30 (*0.03*)	0.55 (*0.02*)	0.26 (*0.09*)	0.35 (*0.20*)

recollection of information as a means of learning that information, improves memory access to contextual information.

2.6.3. Confidence

The high- and low-confidence judgments were coded as binary, with 1 and 0 representing high and low confidence, respectively. We examined average confidence by calculating the proportion of high confidence judgments for each participant, and metacognitive accuracy by calculating Goodman-Kruskal gamma correlations. Gamma correlations are an analysis of the relationship between metacognitive judgments and objective memory performance (Nelson, 1984), and higher gamma values are indicative of better metacognitive accuracy. We calculated participants' gamma correlations by correlating accuracy on each list-discrimination question with each subsequent confidence rating. Because gamma correlations cannot be computed for participants who provide the same confidence rating for every response, two participants were excluded from the analysis on Test 1 and three participants were excluded on Test 2.

3. Results

3.1. Self-reported stress

We first examined whether the Day 1 manipulation (SP or RP) affected participants' self-reported levels of stress. An independent samples *t* test on average Day 1 STICSA scores revealed no difference for participants who had engaged in study practice versus retrieval practice, $t(60) = 1.59$, $p = .118$.

To test whether the TSST-G tasks increased subjective anxiety on Day 2, we conducted a 2 (learning strategy: SP or RP) × 2 (timing: pre-stress or post-stress) ANOVA on average STICSA scores. As expected, we found a main effect of timing as participants demonstrated heightened post-stress STICSA scores relative to baseline, $F(1, 60) = 25.93$, $p < .001$, $\eta_p^2 = 0.30$. The average pre-stress STICSA score was 29.00 (*SEM* = 0.86) and the average post-stress score was 32.24 (*SEM* = 1.03). No other effects were significant.

3.2. Cortisol

We next conducted a 2 (learning strategy: SP or RP) × 3 (timing: baseline, 12 min post-stress, 25 min post-stress) ANOVA on average cortisol concentrations. As shown in Fig. 2, we found a main effect of timing, $F(2, 116) = 3.75$, $p = .026$, $\eta_p^2 = 0.06$. Pairwise comparisons using a Bonferroni correction revealed a marginally significant cortisol increase from 12 min post-stress to 25 min post-stress (mean difference = -0.10, *SEM* = 0.45, $p = .010$), but all other pairwise comparisons were non-significant. Since the TSST-G stress-induction procedure has been highly effective in previous studies (Smith et al., 2016, 2018), the muted cortisol response is more likely due to unanticipated variability in the sample than to ineffective stress induction. For instance,

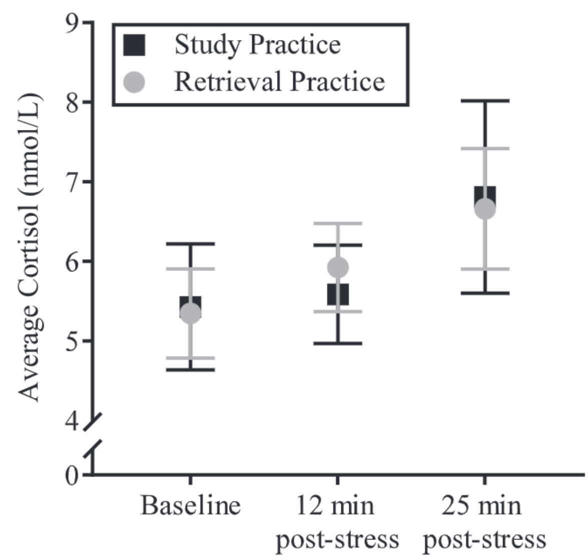

Fig. 2. Average cortisol concentrations on Day 2. Cortisol samples were taken immediately prior to the TSST-G (baseline), 12 min after the onset of the TSST-G, and 25 min after the onset of the TSST-G. Error bars represent *SEM*.

women who are in the follicular phase of their menstrual cycle or who take oral contraceptives generally demonstrate a blunted cortisol response to psychological stress (Kajantie & Phillips, 2006). It is possible that, in the present study, a greater proportion of female participants met either of those criteria than in previous experiments.

3.3. Day 1 memory performance

Table 2 displays correct recall averages for the participants who were given the RP manipulation on Day 1. Because Day 1 memory performance was not relevant to the questions posed by the present study, it was not further examined.

Table 2
Descriptive statistics for Day 1 recall for participants in the RP group (participants in the SP group did not engage in recall on Day 1).

	List 1		List 2	
	First Recall Attempt	Second Recall Attempt	First Recall Attempt	Second Recall Attempt
Mean	14.00	12.58	15.66	14.94
Standard Error	1.18	0.98	1.20	1.35
Range	7–34	6–28	7–38	4–42

3.4. Day 2 memory performance

Table 1 displays means and standard errors for all following dependent measures.

3.4.1. Item memory

A 2 (learning strategy: SP or RP) × 2 (test timing: pre-stress or post-stress) ANOVA on average hit proportions found a main effect of learning strategy, $F(1, 60) = 80.34$, $p < .001$, $\eta_p^2 = 0.57$. For items that they accurately recalled during the Day 1 learning session, individuals in the RP group demonstrated an average hit rate of 0.91, compared to an average hit rate of 0.66 in the SP group. Though this ANOVA did not reveal a specific benefit of RP over SP in the context of stress (i.e., post-stress), bivariate correlations between delta cortisol and post-stress hit proportions elucidated such a benefit. Individuals in the RP group demonstrated a positive relationship between hit rates and pre-post change in cortisol, $r(31) = 0.41$, $p = .023$, whereas those in the SP group did not, $r(27) = -0.17$, $p = .384$. These correlations suggest that a stronger stress response may improve memory for previously studied items, but only for individuals who engaged in successful retrieval practice during learning.[2]

A 2 (learning strategy: SP or RP) × 2 (test timing: pre-stress or post-stress) ANOVA on average false-alarm proportions found a significant learning strategy by test timing interaction, $F(1, 60) = 4.10$, $p = .047$, $\eta_p^2 = 0.06$. As depicted in Fig. 3, participants in the SP and RP groups demonstrated similar performance on the pre-stress test. However, those in the RP group demonstrated lower false alarms on the post-stress test compared to their pre-stress performance whereas those in the SP group demonstrated a post-stress increase in false alarms. Bivariate correlations between delta cortisol and post-stress false alarms did not reveal any associations for the RP or SP groups.

3.4.2. Source memory

We first examined whether source memory scores were significantly different from chance levels of performance (50%) to determine whether participants' source judgments were influenced by episodic memory or by simply guessing. Participants in the SP group demonstrated above chance levels of discrimination on Test 1, $t(29) = 3.14$, $p = .004$, and Test 2, $t(29) = 2.78$, $p = .009$. However, those in the RP group demonstrated chance levels of performance on both Test 1, $t(31) = 0.76$, $p = .452$, and Test 2, $t(31) = 1.50$, $p = .144$. Note that these chance levels of performance occurred for items that these participants had accurately recalled at least once during Day 1 retrieval practice.

We next conducted a 2 (learning strategy: SP or RP) × 2 (test timing: pre-stress or post-stress) mixed model ANOVA on source memory scores to determine whether learning strategy and test timing interacted to influence participants' list-discrimination abilities. This analysis did not find main effects or an interaction (all p's > 0.10). Additionally, bivariate correlations between delta cortisol and post-stress source memory scores did not reveal any associations for the RP or SP groups.

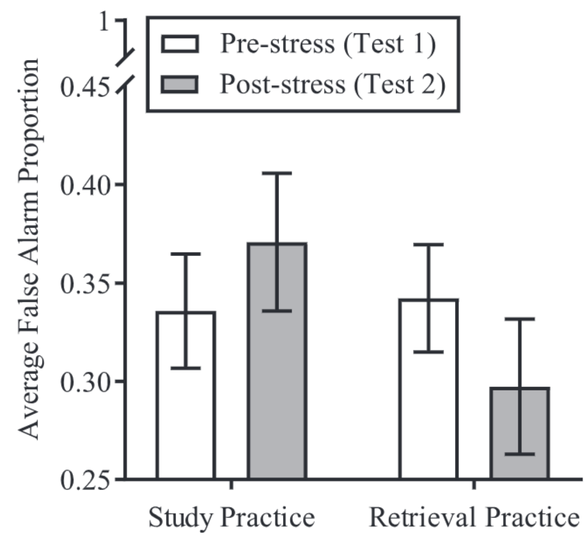

Fig. 3. Average pre-stress and post-stress false alarm proportions for participants in the SP and RP groups. Error bars represent *SEM*.

3.4.3. Confidence

We first examined changes in average confidence as a function of learning strategy and test timing in a 2 (learning strategy: SP or RP) × 2 (test timing: pre-stress or post-stress) mixed model ANOVA. We found a main effect of learning strategy, as individuals in the RP group demonstrated higher average confidence in their source memory judgments than those in the SP group, $F(1, 60) = 28.35$, $p < .001$, $\eta_p^2 = 0.32$. This pattern is not surprising, given that answers for the retrieval practice group were restricted to items that they had accurately recalled at least once during retrieval practice on Day 1. No other effects were significant.

In our analysis on gamma correlations, we first conducted one-sample t tests comparing mean gamma values to a value of 0. A gamma value of 0 demonstrates no correlation between a participant's accuracy and confidence, indicating chance-level metacognitive accuracy. The average post-stress gamma value for the RP group significantly differed from 0, $t(31) = 3.03$, $p = .005$, $d = 0.54$. However, this was not true for the post-stress gamma value for the SP group, $t(26) = 1.08$, $p = .292$, or any of the pre-stress gamma values (SP: $t(27) = 2.01$, $p = .055$; RP: $t(31) = 0.68$, $p = .499$). Thus, only the combination of stress with successful retrieval practice resulted in above-chance metacognitive accuracy.

As shown in Fig. 4, a 2 (learning strategy: SP or RP) × 2 (test timing: pre-stress or post-stress) ANOVA on gamma correlations found a significant interaction between learning strategy and test timing, $F(1, 57) = 4.98$, $p = .030$, $\eta_p^2 = 0.08$. Individuals in the SP group did not demonstrate different pre-stress and post-stress gamma values, whereas those in the RP group demonstrated lower pre-stress than post-stress gamma values. Exploratory bivariate correlations between delta cortisol and post-stress gamma values did not reveal any associations for the RP or SP groups.

4. Discussion

In the present experiment, participants learned two time- and color-segregated wordlists by either engaging in retrieval practice or restudying. A week later, they completed pre- and post-stress recognition tests with accompanying list-discrimination and confidence judgments. This paradigm allowed us to examine whether stress impairs memory for context (i.e., source memory), and whether successful retrieval practice, relative to study practice, improves memory for contextual information that is associated with the learned items. The results suggested a positive influence of successful retrieval practice on item

[2] The analyses on hit proportions were supported by analyses on d', a measure of participants' ability to discriminate between previously studied items and foils on the recognition test. A 2 (learning strategy: SP or RP) × 2 (test timing: pre-stress or post-stress) ANOVA on average d' scores found a main effect of learning strategy, $F(1, 60) = 57.48$, $p < .001$, $\eta_p^2 = 0.49$. Individuals in the RP group demonstrated higher d' scores than those in the SP group (2.34 vs. 0.94). As with the analysis on hit proportions, bivariate correlations between delta cortisol and post-stress d' scores showed that individuals in the RP group demonstrated a positive relationship between d' and change in cortisol, $r(31) = 0.48$, $p = .007$, whereas those in the SP group did not, $r(27) = 0.07$, $p = .715$.

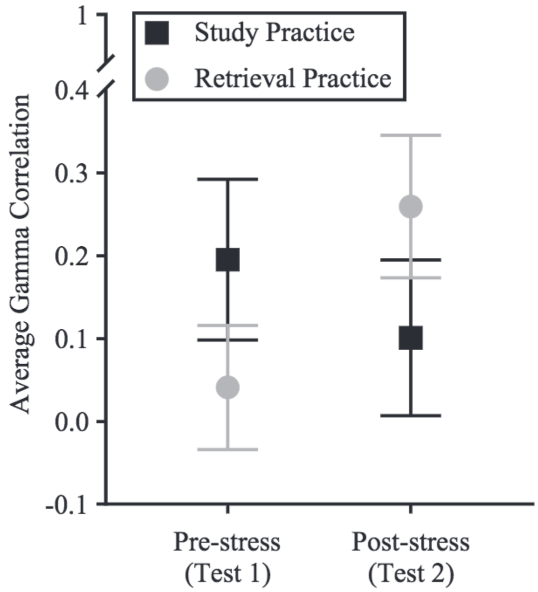

Fig. 4. Average pre-stress and post-stress gamma correlations for participants in the SP and RP groups. Correlations represent the relationship between list-discrimination accuracy and confidence in each list-discrimination judgment. Error bars represent *SEM*.

memory, but not source memory, in the presence of stress. These findings add to a growing body of evidence suggesting that retrieval practice improves post-stress memory accessibility relative to conventional learning strategies (Smith et al., 2016, 2018).

The benefit of successful retrieval practice over study practice in the context of stress was evident in our analyses on item memory as well as our exploratory metacognitive analysis. The retrieval practice group, but not the study practice group, demonstrated a positive correlation between delta cortisol and hit proportions. Thus, greater physiological reactivity to stress was associated with better item memory for individuals who engaged in successful retrieval practice during learning. Individuals in the retrieval practice group also experienced lower false alarm rates after stress induction, whereas the opposite pattern emerged for those in the study practice group. Thus, retrieval practice enhanced the detection of foil items after stress had been induced. Further, although participants' metacognitive accuracy generally was at chance levels on the source-discrimination test, the post-stress gamma value for the retrieval-practice group emerged as the only value to exceed the threshold for chance-level performance. In other words, only successful retrieval practice in combination with stress resulted in an awareness of which items were attributed to the correct source, and which were attributed to the incorrect source. The results from the list-discrimination task were less encouraging, as neither learning strategy nor test timing influenced source memory on the list-discrimination task.

The benefit of successful retrieval practice on item memory but not source memory in the presence of stress may be explained by the influence of stress on different neural networks. Hermans and colleagues (Hermans, Henckens, Joëls, & Fernández, 2014) mapped the brain regions that are upregulated and downregulated during the stress response, categorizing them as the *salience network* and *executive control network*, respectively. Effortful recollection that requires retrieval of contextual information relies on brain regions within the executive control network, explaining why stress typically impairs episodic retrieval. However, automatic and context-independent retrieval, such as semantic memory retrieval, recruits neocortical (Graham, Patterson, & Hodges, 1999; Davey et al., 2015) and striatal (Scimeca & Badre, 2012) regions that fall within the salience network. It is possible that the act of repeatedly retrieving an item from memory results in memory traces that are, like semantic memories, stored within the salience network. Evidence for this hypothesis comes from a study showing that, during final memory testing, individuals who had learned items via retrieval practice demonstrated increased activity in the ACC (part of the salience network) and decreased activity in the PFC and parietal lobes (part of the executive control network) relative to those who learned via study practice (Eriksson, Kalpouzos, & Nyberg, 2011). Thus, retrieval practice may effectively create memories that are stored within the salience network, and thus become more accessible under stress. Like semantic memories, these memories may exist independently of contextual information, explaining why stress improved item memory but not source memory in the present study. This mechanism also explains why the combination of successful retrieval practice and stress resulted in superior metacognitive awareness. If the combination of stress and retrieval practice increases the fluency of memory retrieval, it is likely that individuals are highly confident in these easily-retrieved memories, resulting in higher gamma values. Future researchers should consider further examining the brain regions that are recruited during the retrieval of practiced memories, and whether stress has upregulating effects on these regions.

Of theoretical significance is the finding that the retrieval practice group showed chance-level source discrimination performance despite having accurately recalled the items during their retrieval attempts on Day 1. This stands in contrast to our predictions that were based on the episodic-context account of retrieval practice effects (Karpicke et al., 2014). We expected that successful retrieval practice during encoding would help participants associate the episodic context (i.e., the temporally separated lists) with each to-be-remembered item, resulting in better list-discrimination performance for the retrieval practice group than the study practice group. Though previous researchers using a similar list-discrimination paradigm found this to be true (Brewer et al., 2010), a key difference in methodology could explain our discrepancy: we used a one-week retention interval, whereas Brewer et al. (2010) administered the list-discrimination test immediately after encoding. After a one-week delay, item memory (i.e., hits) may still benefit from individuals having engaged in successful retrieval practice, but source memory may no longer be above chance levels. Supporting this hypothesis is research showing that memory for items is better than memory for source, and that item and source memory decline at similar rates over a one-week retention interval (Bornstein & LeCompte, 1995; Yang et al., 2016). Given that item memory declines substantially over one week (e.g., Roediger & Karpicke, 2006), it is plausible that source memory, which is less robust than item memory to begin with, declines to chance levels after the same amount of time has passed.

This hypothesis provides a new challenge to the episodic-context account. Relative to study practice, retrieval practice has yielded better memory accessibility after retention intervals of up to several weeks in length (for a meta-analysis see Adesope, Trevisan, & Sundararajan, 2017). The episodic-context account posits that this improvement results from increased access to contextual details that are associated with each memory and help guide memory retrieval (Karpicke et al., 2014). However, our results present the possibility that memories may indeed persist over long retention intervals, but their associated contextual details may not.

The presence or absence of stress similarly did not influence performance on the list-discrimination task, standing in contrast to our predictions that were based on the context-shift account (Shields et al., 2017). We expected that stress would induce a mental context shift, as evidenced by a reduced ability to remember the episodic context in which the items were learned (i.e., the temporally-segregated lists). However, it remains possible that stress disrupts memory for other types of contextual details that aid memory retrieval, and that the present paradigm was not well suited for detecting such a disruption.

The lack of support that the present study provided for both the context-shift hypothesis of stress effects (Shields et al., 2017) and the episodic-context account of retrieval practice effects (Karpicke et al.,

2014) calls for further research examining how stress and retrieval practice influence memory for context. Though both of these theories are relatively new, evidence has accumulated in support of each. Our findings raise the question of whether the list-discrimination task employed in the present experiment was an inadequate assessment of context, or whether the two theories need further refinement. Future researchers should consider examining the influence of stress and learning strategies on memory for other types of contextual information, such as the valence, color, or modality of the learned stimuli. Additionally, the results of our list-discrimination task suggested that retrieval practice may promote memory for contextual details for only a brief period after encoding. This hypothesis would benefit from additional tests of the episodic-context account—particularly research investigating whether retrieval practice promotes memory for contextual information following retention intervals of varying lengths.

This study is not without limitations and the present results should be interpreted with caution. Because we could not counterbalance the order of stress induction in our within-subjects design, our use of repeated measures to assess recognition memory performance raises the possibility of confounding variables. For example, the post-stress effects we observed could also be explained by participant fatigue or loss of motivation. However, our conclusions about the influence of stress on memory performance are supported by the post-stress increase in cortisol that we observed, and by the wealth of literature preceding this study demonstrating deleterious effects of stress on retrieval. This study is also limited by a small sample size. We conducted our power analysis with the assumption that we would be examining interactions between learning strategy (SP or RP) and measurement timing (pre-stress or post-stress). Our sample size was sufficient for achieving optimal power but could not support analyses with additional exploratory variables such as gender or stress reactivity (cortisol responders vs. non-responders). Finally, the present study did not fully replicate our prior work (Smith et al., 2016, 2018). Specifically, we did not observe a decrease in hit rates from pre- to post-stress in the study practice group. We did, however, find that stress increased false alarm rates in the study practice group, indicating a detrimental effect of stress on item memory that parallels the free recall results from Smith et al. (2016). Stress generally impairs recognition memory performance to a lesser extent than recall performance (Gagnon & Wagner, 2016), which could explain why we did not provide a complete replication. However, the mechanisms underlying this finding are not yet understood, and future research is necessary to determine the cause of differential effects of stress on tests of recognition and recall.

Author note

The authors declare no conflicts of interest.

Funding

This research was supported by the U.S. Army Natick Soldier Research, Development, and Engineering Center (NSRDEC).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandc.2018.12.005.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659–701. https://doi.org/10.3102/0034654316689306.

Bornstein, B. H., & LeCompte, D. C. (1995). A comparison of item and source forgetting. *Psychonomic Bulletin & Review, 2*(2), 254–259. https://doi.org/10.3758/BF03210966.

Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-Foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory, 18*(4), 385–393. https://doi.org/10.1080/09658211003702163.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977.

Buchanan, T. W., Tranel, D., & Adolphs, R. (2006). Impaired memory retrieval correlates with individual differences in cortisol response but not autonomic response. *Learning & Memory, 13*, 382–387. https://doi.org/10.1101/lm.206306.

Chan, J. C. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 431–437. https://doi.org/10.1037/0278-7393.33.2.431.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavioral Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146.

Davey, J., Cornelissen, P. L., Thompson, H. E., Sonkusare, S., Hallam, G., Smallwood, J., & Jefferies, E. (2015). Automatic and controlled semantic retrieval: TMS reveals distinct contributions of posterior middle temporal gyrus and angular gyrus. *The Journal of Neuroscience, 35*(46), 15230–15239. https://doi.org/10.1523/JNEUROSCI.4705-14.2015.

de Quervain, D. J.-F., Aerni, A., & Roozendaal, B. (2007). Preventive effect of β-adrenoceptor blockade on glucocorticoid-induced memory retrieval deficits. *The American Journal of Psychiatry, 164*(6), 967–969. https://doi.org/10.1176/ajp.2007.164.6.967.

de Quervain, D. J.-F., Roozendaal, B., & McGaugh, J. L. (1998). Stress and glucocorticoids impair retrieval of long-term spatial memory. *Nature, 394*, 787–790. https://doi.org/10.1038/29542.

Eriksson, J., Kalpouzos, G., & Nyberg, L. (2011). Rewiring the brain with repeated retrieval: A parametric fMRI study of the testing effect. *Neuroscience Letters, 505*(1), 36–40. https://doi.org/10.1016/j.neulet.2011.08.061.

Gagnon, S. A., & Wagner, A. D. (2016). Acute stress and episodic memory retrieval: Neurobiological mechanisms and behavioral consequences. *Annals of the New York Academy of Sciences, 1369*(1), 55–75. https://doi.org/10.1111/nyas.12996.

Goldfarb, E. V., Tompary, A., Davachi, L., & Phelps, E. A. (2018). Acute stress throughout the memory cycle: Diverging effects on associative and item memory. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000472.

Graham, K. S., Patterson, K., & Hodges, J. R. (1999). Episodic memory: New insights from the study of semantic dementia. *Current Opinion in Neurobiology, 9*(2), 245–250. https://doi.org/10.1016/S0959-4388(99)80035-X.

Grös, D. F., Antony, M. M., Simms, L. J., & McCabe, R. E. (2007). Psychometric properties of the state-trait inventory for cognitive and somatic anxiety (STICSA): Comparison to the state-trait anxiety inventory (STAI). *Psychological Assessment, 19*(4), 369–381. https://doi.org/10.1037/1040-3590.19.4.369.

Henckens, M. J. A. G., Pu, Z., Hermans, E. J., van Wingen, G. A., Joëls, M., & Fernández, G. (2012). Dynamically changing effects of corticosteroids on human hippocampal and prefrontal processing. *Human Brain Mapping, 33*(12), 2885–2897. https://doi.org/10.1002/hbm.21409.

Hermans, E. J., Henckens, M. J. A. G., Joëls, M., & Fernández, G. (2014). Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends in Neurosciences, 37*(6), 304–314. https://doi.org/10.1016/j.tins.2014.03.006.

Het, S., Ramlow, G., & Wolf, O. T. (2005). A meta-analytic review of the effects of acute cortisol administration on human memory. *Psychoneuroendocrinology, 30*(8), 771–784. https://doi.org/10.1016/j.psyneuen.2005.03.005.

Kajantie, E., & Phillips, D. I. W. (2006). The effects of sex and hormonal status on the physiological response to acute psychosocial stress. *Psychoneuroendocrinology, 31*(2), 151–178. https://doi.org/10.1016/j.psyneuen.2005.07.002.

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language, 62*(3), 227–239. https://doi.org/10.1016/j.jml.2009.11.010.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Vol. Ed.), *The psychology of learning and motivation: vol. 61*, (pp. 237–284). San Diego, CA: Elsevier Academic Press.

Kuhlmann, S., Piel, M., & Wolf, O. T. (2005). Impaired memory retrieval after psychosocial stress in healthy young men. *The Journal of Neuroscience, 25*, 2977–2982. https://doi.org/10.1523/JNEUROSCI.5139-04.2005.

Lovallo, W. R., Robinson, J. L., Glahn, D. C., & Fox, P. T. (2010). Acute effects of hydrocortisone on the human brain: An fMRI study. *Psychoneuroendocrinology, 35*(1), 15–20. https://doi.org/10.1016/j.psyneuen.2009.09.010.

Maddox, G. B., & Balota, D. A. (2015). Retrieval practice and spacing effects in young and older adults: An examination of the benefits of desirable difficulty. *Memory & Cognition, 43*(5), 760–774. https://doi.org/10.3758/s13421-014-0499-6.

Mulligan, N. W., Susser, J. A., & Smith, S. A. (2016). The testing effect is moderated by experimental design. *Journal of Memory and Language, 90*, 49–65. https://doi.org/10.1016/j.jml.2016.03.006.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133. https://doi.org/10.1037/0033-2909.95.1.109.

Reul, J. M., & de Kloet, E. R. (1985). Two receptor systems for corticosterone in rat brain: Microdistribution and differential occupation. *Endocrinology, 117*(6), 2505–2511. https://doi.org/10.1210/endo-117-6-2505.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x.

Roozendaal, B., Hahn, E. L., Nathan, S. V., de Quervain, D. J.-F., & McGaugh, J. L. (2004). Glucocorticoid effects on memory retrieval require concurrent noradrenergic activity in the hippocampus and basolateral amygdala. *The Journal of Neuroscience, 24*(37), 8161–8169. https://doi.org/10.1523/JNEUROSCI.2574-04.2004.

Rugg, M. D., Vilberg, K. L., Mattson, J. T., Yu, S. S., Johnson, J. D., & Suzuki, M. (2012). Item memory, context memory and the hippocampus: FMRI evidence. *Neuropsychologia, 50*(13), 3070–3079. https://doi.org/10.1016/j.neuropsychologia.2012.06.004.

Schneider, W., Eschman, A., & Zuccolotto, A. (2001). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

Schwabe, L., & Wolf, O. T. (2009). The context counts: Congruent learning and testing environments prevent memory retrieval impairment following stress. *Cognitive, Affective, & Behavioral Neuroscience, 9*(3), 229–236. https://doi.org/10.3758/CABN.9.3.229.

Schwabe, L., Joels, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: An update and integration. *Neuroscience & Biobehavioral Reviews, 36*(7), 1740–1749. https://doi.org/10.1016/j.neubiorev.2011.07.002.

Scimeca, J. M., & Badre, D. (2012). Striatal contributions to declarative memory retrieval. *Neuron, 75*(3), 380–392. https://doi.org/10.1016/j.neuron.2012.07.014.

Shields, G. S., Sazma, M. A., McCullough, A. M., & Yonelinas, A. P. (2017). The effects of acute stress on episodic memory: A meta-analysis and integrative review. *Psychological Bulletin, 143*(6), 636–675. https://doi.org/10.1037/bul0000100.

Smith, A. M., Davis, F. C., & Thomas, A. K. (2018). Criterial learning is not enough: Retrieval practice is necessary for improving post-stress memory accessibility. *Behavioral Neuroscience, 132*(3), 161–170. https://doi.org/10.1037/bne0000240.

Smith, A. M., Floerke, V. A., & Thomas, A. K. (2016). Retrieval practice protects memory against acute stress. *Science, 354*(6315), 1046–1048. https://doi.org/10.1126/science.aah5067.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8*(2), 203–220. https://doi.org/10.3758/BF03196157.

Strand, B. Z. (1970). Change of context and retroactive inhibition. *Journal of Verbal Learning & Verbal Behavior, 9*(2), 202–206. https://doi.org/10.1016/S0022-5371(70)80051-2.

von Dawans, B., Kirschbaum, C., & Heinrichs, M. (2011). The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology, 36*(4), 514–522. https://doi.org/10.1016/j.psyneuen.2010.08.004.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*, 1191–1207. https://doi.org/10.3758/s13428-012-0314-x.

Weitzman, E. D., Fukushima, D., Nogeire, C., Roffwarg, H., Gallagher, T. F., & Hellman, L. (1971). Twenty-four hour pattern of the episodic secretion of cortisol in normal subjects. *The Journal of Clinical Endocrinology & Metabolism, 33*(1), 14–22. https://doi.org/10.1210/jcem-33-1-14.

Whiffen, J. W., & Karpicke, J. D. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1036–1046. https://doi.org/10.1037/xlm0000379.

Yang, J., Zhan, L., Wang, Y., Du, X., Zhou, W., Ning, X., ... Moscovitch, M. (2016). Effects of learning experience on forgetting rates of item and associative memories. *Learning & Memory, 23*(7), 365–378. https://doi.org/10.1101/lm.041210.115.

Zeeuws, I., Deroost, N., & Soetens, E. (2010). Effect of an acute d-amphetamine administration on context information memory in healthy volunteers: Evidence from a source memory task. *Human Psychopharmacology: Clinical and Experimental, 25*(4), 326–334. https://doi.org/10.1002/hup.1120.