Examining the Role of Retrieval Practice in Improving Memory Accessibility Under Stress

A dissertation by

Amy Marie Smith

In partial fulfillment of the requirements for the degree of

The Doctor of Philosophy

in

Psychology

TUFTS UNIVERSITY

May 2018

Advisor:

Dr. Ayanna Thomas

Committee Members:

Dr. Elizabeth Race, Dr. Lisa Shin, & Dr. Andrew Yonelinas

ABSTRACT

Psychological stress has been shown to impair episodic memory retrieval. The goal of this dissertation was to investigate one method of improving memory accessibility under stress, by using the learning strategy *retrieval practice* to create stress-resistant memories. In Experiment 1, I examined whether having participants learn information by using retrieval practice, which is analogous to taking practice tests, would improve post-stress memory relative to having participants learn via conventional restudying. Confirming my hypotheses, those who learned by restudying demonstrated typical stress-related memory impairment, whereas those who learned using retrieval practice were immune to the deleterious effects of stress. In Experiments 2a and 2b, I aimed to determine whether learning to criterion, in which only one successful retrieval attempt is made, would similarly buffer memory against stress, or whether multiple retrieval attempts are necessary to achieve that effect. I found that criterial learning protected memory against stress in the short term (after 24 hours), but additional retrieval practice was necessary to achieve that effect in the long term (after one week). Finally, in Experiment 3, I tested one potential theoretical mechanism underlying the efficacy of retrieval practice at bolstering memory against stress. My hypotheses were not supported; I found that retrieval practice provided some support for item memory but no support for source memory in the context of stress. Altogether, these experiments help establish and explain the ways in which retrieval practice and psychological stress interact to influence memory retrieval.

Keywords: stress, memory, retrieval, context, criterial learning

Acknowledgments

The completion of this dissertation would not have been possible without the support of many individuals. I would first like to thank my graduate advisor, Ayanna Thomas. Thank you for giving an inexperienced 22-year-old a chance at graduate school, and for providing me with the guidance and gentle nudging that I needed to achieve my best. I have benefitted immensely from your intelligence, creative mind, and unwavering support of my work. I would also like to thank my dissertation committee members, Elizabeth Race, Lisa Shin, and Andy Yonelinas. Your insights and feedback on this project have expanded my mind and greatly improved the quality of this work. Many thanks to my lab members, both past and present. Leamarie Gordon and Meeyeon Lee, thank you for being amazing mentors and friends. I wouldn't have made it through the first two years of grad school without you. Renée DeCaro and Greg Hughes, thank you for making our office a comfortable place to vent, laugh, and (sometimes) work. I wouldn't have made it through these last three years without you two. I'd also like to thank my family for their love and support—particularly their psychological support. To my amazing parents, Lora and Dave, my siblings, Biz, Danny, and Anna, and my in-laws, Carol and Antony Hughes: thank you for believing that I could do this. You convinced me that I could. Finally, my biggest thank you to Greg Hughes. Thank you for always dropping your work to help me with my own. You are the kindest, most selfless person I know, and I am so fortunate to have you in my life.

Funding Sources

This dissertation was supported in part by the U.S. Army Natick Soldier Research, Development, and Engineering Center (NSRDEC). Additional support was received from a Tufts Graduate Research Award.

Dedication

This dissertation is dedicated to my beloved dad and role model, David Alan Smith. I would never have done this without your support, and I could never have done this without your support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Examining the Role of Retrieval Practice in Improving Memory Accessibility Under Stress

CHAPTER 1: INTRODUCTION

Psychological stress often results in a paradoxical scenario: when we are under pressure to perform our best, we find ourselves performing our worst. The subjective experience of "choking under pressure" has been repeatedly substantiated in memory research. Specifically, over a dozen experiments and a recent meta-analysis have culminated in the consensus that acute incidences of stress temporarily impair memory retrieval (Shields, Sazma, McCullough, & Yonelinas, 2017). These studies adhered to a common method whereby participants learned stimuli and returned after 24 hours for stress induction and subsequent memory testing. With a primary focus on manipulating stress and observing memory performance, these studies were not expressly concerned with the quality of encoding during initial learning. Participants in these experiments either engaged in incidental encoding or were simply instructed to "memorize" stimuli. Thus, in the context of stress-and-memory paradigms, researchers have yet to manipulate the use of different encoding techniques, of which there are many of varying efficacies (see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Doing so would determine whether all memories are subject to the deleterious effects of stress, or whether post-stress memory accessibility is contingent on how information is initially learned.

The present research addresses whether post-stress memory retrieval can be improved when a highly effective learning strategy is employed during the encoding of stimuli. In a preliminary experiment using the strategy *retrieval practice*, I found evidence to suggest that this can be accomplished (Smith, Floerke, & Thomas, 2016). Retrieval practice is the act of taking practice tests to promote the learning and long-term retention of information, and consistently yields better memory on a final criterial test than simply studying (for reviews see Roediger &

Karpicke, 2006b; Roediger & Butler, 2011). In the first experiment of this dissertation, I found that the effectiveness of retrieval practice extended to contexts involving stress. Specifically, individuals who encoded stimuli under “memorize” instructions showed post-stress memory impairment, whereas those who encoded stimuli using retrieval practice did not (Smith et al., 2016).

The subsequent experiments in this dissertation further examined why retrieval practice reduces the negative effects of stress on memory retrieval. In Experiments 2a and 2b, I investigated how to optimize the use of retrieval practice for creating stress-resistant memories. I showed that a single successful retrieval practice attempt during encoding can buffer memory against stress in the short term, but multiple retrieval attempts are necessary to improve post-stress retrieval after a longer interval. This investigation has important implications for the real-world application of retrieval practice. For example, the number of retrieval attempts that a student opts to make while studying may depend on how much time will pass between her study session and the stressful exam for which she is preparing.

In the final experiment of my dissertation, I tested one theoretical mechanism that may underlie the effectiveness of retrieval practice at improving memory in the context of stress. There is a commonality in the theories surrounding stress effects and retrieval practice effects: both theories work via a context mechanism. Evidence suggests that stress creates a mental context shift (see Shields et al., 2017). When stress occurs prior to retrieval, the mismatch between the encoding context and the retrieval context could result in substandard memory performance, as context mismatches often do (for a review see Smith & Vela, 2001). Retrieval practice is believed to improve memory access to contextual information that may help reinstate context during retrieval (see Karpicke, Lehman, & Aue, 2014). Thus, I hypothesized that stress would impair memory for contextual information, and that retrieval practice would improve post-stress memory performance by improving access to contextual information. Results suggested that, in the presence of stress, retrieval practice improved item memory but not context memory relative to restudying.

The results of the experiments in this dissertation have the potential to be of both practical and theoretical importance. From a practical perspective, determining how to alleviate stress-related memory impairment promises to benefit many individuals in their everyday lives. As examples, improving post-stress memory accessibility would benefit professionals in stressful work scenarios and students taking stressful, high-stakes exams. Techniques that accomplish this would also serve to test and refine current theories surrounding the negative effects of stress on memory retrieval. Experiment 3 directly tested the context-shift hypothesis as previously described, but other theories that were not directly tested here also benefitted from this research. For instance, Schwabe and colleagues have hypothesized that the stressed brain enters a "memory formation mode," in which processes related to the learning and consolidation of information are prioritized at the expense of retrieval (Schwabe, Joels, Roozendaal, Wolf, & Oitzl, 2012). The finding that retrieval practice promotes successful post-stress retrieval helps specify the conditions under which retrieval is and is not impeded by the stress-induced memory formation mode. Stress has also been hypothesized to impair retrieval via an executive resource depletion mechanism, in which stress taxes the executive functions that are necessary for successful engagement in careful, effortful recollection (Gagnon & Wagner, 2016). Demonstrating ways in which memories can be retrieved even when executive resources are impaired also helps refine this theory.

CHAPTER 2: LITERATURE REVIEW

The Physiological Stress Response and Memory Retrieval

In the present research, psychological stress is defined as an uncontrollable and/or unpredictable threat to the physical or social self (Dickerson & Kemeny, 2004). This threat is accompanied by a subjective feeling of mental stress and a physiological response.

The detrimental effects of psychological stress on retrieval are, in part, a consequence of the human stress response. The stress response is characterized by different phases of hormone release, the first of which is initiated by the hypothalamus after a threat has been perceived (Everly & Lating, 2013). From here, preganglionic sympathetic nerves carry neural impulses to the adrenal medulla (Raven, Raven, & Chew, 2010). The stimulated adrenal medulla then serves as the postganglionic release point for epinephrine, prompting the "fight or flight" response that prepares the body to take defensive action. Though epinephrine cannot cross the blood-brain barrier to directly affect neural activity, increases in epinephrine initiate a chain of stimulation from the vagus nerve to the solitary nucleus in the medulla to the basolateral amygdala (Williams & Clayton, 2001). The basolateral amygdala modulates the learning and consolidation of information, particularly emotional information (McGaugh, Cahill, & Roozendaal, 1996). Thus, memory consolidation is often enhanced after stress (e.g., Andreano & Cahill, 2006; McCullough & Yonelinas, 2013; Smeets, Otgaar, Candel, & Wolf, 2008). During a brief period after the onset of stress (< 10 min), memory retrieval may also be enhanced (e.g., Hupbach & Fieman, 2012) or otherwise unaffected (e.g., Schönfeld, Ackermann, & Schwabe, 2014), though literature examining retrieval immediately post-stress is sparse.

The second phase of the stress response is longer lasting and occurs via a different mechanism, referred to as the hypothalamic-pituitary-adrenal (HPA) axis (Everly & Lating, 2013). While the hypothalamus activates the adrenal medulla during the first phase of the stress response, it simultaneously excretes corticotropin-releasing factor that stimulates the pituitary gland. The anterior pituitary then releases adrenocorticotropic hormone into the bloodstream, triggering the synthesis and release of the human stress hormone cortisol from the adrenal cortex. The release of cortisol from the adrenal cortex is gradual, reaching peak levels in the blood approximately 20 min after the initial detection of a threat (Kirschbaum, Pirke, & Hellhammer, 1993). The magnitude of the HPA-axis response to stress varies greatly at the individual level (e.g., Kudielka, Buske-Kirschbaum, Hellhammer, & Kirschbaum, 2004; Uhart, Chong, Oswald, Lin, & Wand, 2006), but is largest and most reliable in stress paradigms involving social evaluation (Dickerson & Kemeny, 2004; Skoluda et al., 2015).

The post-stress release of cortisol can have both desirable and undesirable consequences. Cortisol aids in the metabolism of fats, lipids, and carbohydrates to provide the body with the energy needed to deal with an imminent threat. Cortisol also crosses the blood-brain barrier to exert both positive and negative effects on the brain. In particular, cortisol binds heavily to glucocorticoid and mineralocorticoid receptors in the amygdala and hippocampus, (Lovallo, Robinson, Glahn, & Fox, 2010; Reul & de Kloet, 1985), both of which are implicated in memory formation and retrieval. The increased occupation of glucocorticoid receptors in the amygdala and hippocampus perpetuates the enhancement of memory formation that begins during the first phase of the stress response (Schwabe et al., 2012). However, the prioritization of neural pathways involved in encoding and consolidation comes at the cost of retrieval. Once cortisol levels peak after stress, memory retrieval may be impaired for over an hour (for a review see Gagnon & Wagner, 2016).

Retrieval may be further hindered by the stress response as cortisol indirectly reduces activity in the prefrontal cortex (PFC). Working-memory deficits and associated reductions in PFC activity have been observed following stress induction in a laboratory setting (Gärtner, Rohde-Liebenau, Grimm, & Bajbouj, 2014; Qin et al., 2009). In addition to consuming prefrontal cognitive resources, stress can bias attention toward threatening objects (see Christianson, 1992). All of these disruptions may serve to further impair memory retrieval.

The Effects of Stress on Memory Retrieval

To examine the effects of stress on memory retrieval, researchers have commonly used the Trier Social Stress Test (TSST) to induce acute psychological stress. The TSST is a 20-min procedure in which an individual is given 10 min to prepare a speech, must deliver the speech extemporaneously for 5 min, and then must complete 5 min of serial subtraction (e.g., serially subtract 13 from 1,022) while being videotaped and observed (Kirschbaum et al., 1993). Some adaptations to the TSST include a version for use with children in which the speech consists of having to complete a story (Buske-Kirschbaum et al., 1997), a version in which the speech and math are performed in virtual reality (Ruiz et al., 2010), and a version for group testing in which multiple participants give speeches and solve math problems in front of each other (von Dawans, Kirschbaum, & Heinrichs, 2011).

The majority of studies on the topic have indeed reported detrimental effects of stress on memory retrieval (e.g., Buchanan & Tranel, 2008; Domes, Heinrichs, Rimmele, Reichwald, & Hautzinger, 2004; Hidalgo et al., 2015; Kuhlmann, Piel, & Wolf, 2005; Merz, Wolf, & Hennig, 2010; Oei, Everaerd, Elzinga, Well, & Bermond, 2006; Schönfeld, Ackermann, & Schwabe, 2014; Schwabe & Wolf, 2009; 2014; Schwabe et al., 2009; Smith, et al., 2016; Tollenaar, Elzinga, Spinhoven, & Everaerd, 2008). However, a handful of studies have reported null or even positive findings, and these discrepant results may be explained by variations in methodology. These variations are relevant to the present research and will be revisited in later subsections.

In a standard stress-and-memory paradigm, young adult participants learn verbal or pictorial materials under either incidental or intentional study instructions. In some studies, participants also take one or more free recall tests immediately after initial encoding. Twenty-four or 48 hours later, participants return to the lab for stress induction. Between 15 and 30 min thereafter, once cortisol reaches peak post-stress levels, memory is assessed, most often via a free recall test.

In studies that have deviated from the standard methodology outlined above, null or even positive effects of stress on retrieval have been reported (Beckner, Tucker, Delville, & Mohr, 2006; Hupbach & Fieman, 2012; Li, Weerda, Milde, Wolf, & Thiel, 2014; Pulopulos et al., 2013; Schoofs & Wolf, 2009; Wolf, Schommer, Hellhammer, Reischies, & Kirschbaum, 2002). One methodological variation that may contribute to these findings is the age of the population being investigated. Studies that have examined the effects of stress on retrieval in older adults have typically found null effects of stress (Hidalgo et al., 2015; Pulopulos et al., 2013), possibly because age-related neural changes leave the older brain less sensitive to increases in cortisol (Mizoguchi et al., 2009) and/or because older adults cope with stress more effectively than younger adults (Hamarat et al., 2001).

A second influence is the timing of the memory test relative to when stress is induced. When individuals are tested 25-30 min post-stress, during the cortisol peak, studies typically report negative effects of stress on retrieval. However, when memory is tested immediately post-stress, several studies have reported null effects (Hupbach & Fieman, 2012; Schönfeld et al., 2014; Schwabe & Wolf, 2014; Smith et al., 2016; Zoladz et al., 2014), whereas only one has reported significant memory impairment (Lupien et al., 1997).

Post-stress memory performance has also been shown to differ according to the type of final memory test administered. Tests of free recall typically yield some degree of stress-related memory impairment (Buchanan & Tranel, 2008; Buchanan, Tranel, & Adolphs, 2006; Hidalgo et al., 2015; Kuhlmann et al., 2005; Oei et al., 2006; Quesada, Wiemers, Schoofs, & Wolf, 2012; Schönfeld et al., 2014; Schwabe & Wolf, 2009; Smith et al., 2016). Findings have been mixed regarding cued recall, with two studies reporting significant impairment (Lupien et al., 1997; Smeets et al., 2008), one reporting marginally significant impairment (Tollenaar et al., 2008), and one reporting null effects (Kuhlmann et al., 2005). Only one experiment used fragment completion and found null effects of stress (Lupien et al., 1997). Finally, the studies that assessed recognition performance found either significant impairment (Merz et al., 2010; Schwabe & Wolf, 2014), selective impairment for stimuli with a positive emotional valence (Domes et al., 2004; Hidalgo et al., 2015), or no effect of stress on recognition (Beckner et al., 2006; Buchanan & Tranel, 2008; Li et al., 2014; Pulopulos et al., 2013).

Finally, the use of multiple free recall tests during initial learning has likely contributed to discrepant findings in the stress-and-memory literature. Many researchers have administered tests of free or cued recall immediately after participants first learn stimuli on the first day of the experiment. The use of retrieval practice in this manner has been shown to yield highly durable long-term memories (for reviews see Roediger & Karpicke, 2006b; Roediger & Butler, 2011), and is particularly effective when multiple free-recall attempts are made. This may explain why two experiments that employed multiple free-recall tests during initial learning reported no detrimental effect of stress on a later memory test (Schoofs & Wolf, 2009; Wolf et al., 2002).

Retrieval Practice: Findings and Theories

Because of the amazing efficacy of retrieval practice at improving long-term memory relative to restudying, it is unsurprising that some researchers have not found detrimental effects of stress on memory when retrieval practice was employed during encoding (i.e., Schoofs & Wolf, 2009; Wolf et al., 2002). Retrieval practice consistently yields better memory retention than not only conventional restudying (for reviews see Roediger & Butler, 2011; Roediger & Karpicke, 2006b), but also many other highly effective learning strategies such as mental imagery (Dunlosky et al., 2013), concept mapping (Karpicke & Blunt, 2011), and the keyword mnemonic (Fritz, Morris, Acton, Voelkel, & Etkind, 2007). Further, the effect sizes in studies comparing retrieval practice to various learning strategies are often large. As examples, retrieval practice has been shown to result in roughly 100% more information remembered than restudying (e.g., Roediger & Karpicke, 2006a), and 50% more information remembered than concept mapping, which involves drawing diagrams that summarize the relationships between the various concepts being learned (Karpicke & Blunt, 2011). Of all learning strategies that are currently available, retrieval practice emerges as the best candidate for creating memories that are resistant to psychological stress. Coming full circle, this supposition is supported by the null findings reported in stress-and-memory experiments that used multiple retrieval attempts during encoding (Schoofs & Wolf, 2009; Wolf et al., 2002).

In addition to examining how retrieval practice compares to other learning strategies, an important area of literature has focused on how retrieval practice measures up to different versions of itself. Questions addressed include whether one retrieval-practice attempt provides the same benefit as more attempts, whether successful retrieval attempts are better than those that are unsuccessful, and how many retrieval attempts are necessary before any subsequent attempts begin to yield diminishing returns. Answers to these questions have emerged from studies on *criterial learning*, which involves having participants alternate between studying and testing until they reach a given number (i.e., criterion) of correct recalls. When studying, students commonly set their recall criterion to one, such that they engage in self-testing for a given set of materials only until each item has been correctly recalled just once (Kornell & Bjork, 2007; 2008; Wissman, Rawson, & Pyc, 2012). Contrary to this common practice, researchers have found that learning to a criterion of three is optimal for achieving a balance between memory durability and study efficiency (Rawson & Dunlosky, 2011). Three retrieval practice attempts during encoding yield better long-term retention than either one attempt in which only some items are remembered (Roediger & Karpicke, 2006a) or one attempt in which all items are remembered (Karpicke & Roediger, 2008). Further, these three attempts are most effective when they are all successful (Rawson & Dunlosky, 2011).

Though several theories have attempted to explain the memorial benefits of retrieval practice (for a review see Karpicke et al., 2014), the *episodic-context* account is currently the most compelling and strongly supported of them (e.g., Criss & Shiffrin, 2004; Jang & Huber, 2008; Whiffen & Karpicke, 2017). This account is based on the premise that each attempt at retrieving a desired memory happens in a novel context. 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 in 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 (Karpicke et al., 2014). On a final memory test, an individual who has engaged in retrieval practice therefore has both a more recent memory of the original study context and additional contextual cues for guiding his or her memory search than an individual who engaged in restudying. All of these cues are helpful when attempting to recall an item on a final memory test because they allow an individual to reinstate a prior context in which the item was learned or successfully retrieved.

The episodic-context account is supported by a host of studies in which retrieval practice has been shown to improve memory for contextual information (for a review see Karpicke et al., 2014). For example, Brewer, Marsh, Meeks, Clark-Foos, and Hicks (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) from which each item came. 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.

Stress and Memory: Underlying Mechanisms

Implicated in the detrimental effects of stress on memory retrieval is the physiological stress response, as the hormones released during a stressful event interfere with retrieval-related neural processing. This mechanism is complemented by theoretical models of memory that specify the cognitive mechanisms that may be disrupted by stress. Two such theories were mentioned in the previous discussion of the physiological stress response: stress may impair retrieval because the stressed brain prioritizes neural processes involved in encoding and consolidation (Schwabe et al., 2012), and/or because stress impedes executive functions that are necessary for successful retrieval (Gagnon & Wagner, 2016). 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 latter instance exemplifies the effect that stress may have on internal context. Participants may be in a calm mental state during learning, and then may experience memory impairment when in a stressed mental state at test.

Schwabe and Wolf (2009) provided early support for the context-shift hypothesis. In their experiment, 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. In sum, these findings suggest that stress may create a context shift that can be overridden by adequate contextual support.

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, in order to target the stabilizing of memory traces that occurs in the several hours after encoding. Typically, memory is enhanced when stress is induced at this phase (see Shields et al., 2017). One possible explanation for this effect is that stress induced after encoding creates a mental context shift that isolates the encoded information from anything learned during the interval between encoding and retrieval. This may be particularly important because context disruptions have been shown to result in less retroactive interference than maintaining the same context between encoding and retrieval (Strand, 1970). Without a stress-induced context shift, individuals could experience retroactive interference in which memory for the encoded information is impeded by contextually similar memories that are formed during the consolidation phase.

Though this theory has yet to be empirically tested, Shields et al. (2017) provided some support in their meta-analysis. Specifically, they found that stress during consolidation benefitted memory only when the stressor and consolidation phase occurred in the original learning context. 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. Together with Schwabe and Wolf (2009), Shields et al. (2017) 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 Interaction Between Stress and Retrieval Practice

To summarize, two well-supported conclusions have been reached in distinct areas of research: psychological stress impairs memory retrieval, and retrieval practice is a powerful tool for improving long-term memory retention. Perhaps inadvertently, several researchers have combined these topics by administering memory tests during the encoding phase of their stress-and-memory experiments (e.g., Buchanan & Tranel, 2008; Kuhlmann, Piel, & Wolf, 2005; Oei et al., 2006; Schoofs & Wolf, 2009; Wolf et al., 2002). A few of these studies have reported null effects (Schoofs & Wolf, 2009; Wolf et al., 2002), contesting the broad consensus that stress impairs memory retrieval. However, at the time of this dissertation, no researchers had attempted to manipulate retrieval practice at encoding to determine whether it could be used to create stress-resistant memories.

In addition to the mentioned null-effect studies, the notion that retrieval practice strengthens memory against stress is supported by commonality in the theories surrounding stress effects and retrieval practice effects: both theories work via a context mechanism. 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. Together, these models suggest that retrieval practice may improve post-stress memory accessibility by providing the stressed brain with the tools to reinstate mental context when stress has otherwise disrupted it. Thus, the idea that retrieval practice may create memories that are accessible under stress is supported from both empirical and theoretical sources.

CHAPTER 3: OVERVIEW OF THE PRESENT RESEARCH

In the present research, I examined the interaction between retrieval practice and psychological stress as it relates to the dependent outcome of memory retrieval. Theoretical and empirical evidence suggests that retrieval practice should result in better post-stress memory performance than does conventional studying. In Experiment 1, I tested this hypothesis by manipulating the use of retrieval practice at encoding in an otherwise typical stress-and-memory paradigm. The results of this experiment were as predicted: participants who learned via restudying demonstrated the typical stress-related memory impairment, whereas those who learned via retrieval practice were immune to the deleterious effects of stress. In Experiments 2a and 2b, I replicated and expanded upon the Experiment 1 findings. Specifically, I aimed to determine whether learning to a criterion of one correct recall would similarly buffer memory against stress, or whether multiple retrieval attempts are necessary to achieve that effect. The retention interval between encoding and retrieval was varied in Experiments 2a and 2b (24 hr and one week, respectively). Finally, in Experiment 3, I tested the hypothesis that retrieval practice improves post-stress memory by increasing access to contextual details from the learning context. Participants learned two distinct lists of words and engaged in either retrieval practice or restudying of the lists. A subsequent list-discrimination test allowed me to examine whether stressed individuals who learned via retrieval practice had better memory access to temporal contextual details than those who learned via restudying.

CHAPTER 4: EXPERIMENT 1

Experiment 1 addressed two key limitations of previous studies investigating the effects of psychological stress on memory retrieval. First, I aimed to determine whether the use of retrieval practice at encoding could improve post-stress memory accessibility relative to having participants use a conventional rereading technique. Previous studies either have used retrieval practice at encoding (e.g., Schoofs & Wolf, 2009; Wolf et al., 2002) or have had participants simply study or "memorize" stimuli (e.g., Buchanan & Tranel, 2008; Domes et al., 2004). However, none have manipulated these learning strategies at encoding and compared subsequent post-stress memory performance. Second, few studies have examined memory retrieval immediately following the onset of stress when stress may actually improve performance. The present study aimed to address these limitations by manipulating retrieval practice versus restudying at encoding, and assessing memory both immediately after stress and after the typical 25-min delay.

Experiment 1 Method

Design

The experiment employed a 2 (learning strategy: retrieval practice vs. study practice) x 2 (TSST-G group: stressed vs. non-stressed) between-subjects factorial design.

Participants

One hundred twenty Tufts University students participated in the experiment (72 females, $M_{\text{age}} = 20.08$, $SD_{\text{age}} = 3.95$). Some participants were recruited through introductory psychology courses to fulfill a research participation requirement, and some were recruited from across the Tufts campus and received $20 for their participation. Thirty participants were randomly assigned to each of four groups: non-stressed study practice (SP), non-stressed retrieval practice (RP), stressed SP, and stressed RP. All participants provided informed consent.

Materials

Stimuli. The stimuli consisted of 30 nouns presented as words (15 neutral, 15 negative) and 30 nouns presented as images (15 neutral, 15 negative). All 60 stimuli were chosen from the Snodgrass and Vanderwart (1980) pictures that have been normed according to emotional valence and arousal. The words presented were the nouns associated with each of 30 images. All images and words were semantically distinct. Because preliminary analyses showed that memory performance on all memory measures did not differ according to the emotionality of the stimuli, this aspect of the design will not be further discussed.

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 0-80 and higher scores are indicative of higher self-reported anxiety.

The Trier Social Stress Test for Groups (TSST-G) Testing Room. Stress was induced using the TSST-G, which is a modified version of the TSST that accommodates group testing (von Dawans et al., 2011). In the testing room, participants were seated at four partitioned desks that were numbered 1-4 to facilitate participants being called on by the experimenter. When called on, participants stood and faced the front of the room where the two experimenters stood and took notes using clipboards. A camcorder was mounted on a tripod to the left of the experimenters and the camcorder appeared to be recording as they gave speeches and solved math problems.

Empatica E4 Wristbands. Heart rate was measured continuously on the second day of the experiment via Empatica E4 wristbands (see www.empatica.com). The E4 features a large button on the watch face, which participants were instructed to press at various points throughout the experiment to mark the onset and offset of different phases of the procedure.

Procedure

Stressed Group. Testing sessions occurred on two consecutive days between 3:30 p.m. and 5:30 p.m. to control for variability in diurnal cortisol secretion. Morning testing was avoided because cortisol levels are naturally elevated in the morning (Weitzman et al., 1971), which could influence memory performance (e.g., de Kloet, Oitzl, & Joëls, 1999).

Four participants per session partook in the experiment. On Day 1, participants first began the encoding task, which was presented using E-Prime software (Version 2.1; Schneider, Eschman, & Zuccolotto, 2001). Participants were instructed that they would see several words and images, and that they would be given 10 s per item to type a sentence that included each given item. They were then presented with either the 30 words or 30 images at a rate of 2 s per item. Whether words or images were studied first was counterbalanced. After each item was presented, participants had 10 s to enter a sentence before the program advanced to the next item. Next, participants completed simple multiplication and division problems using pen and paper for 1 min (e.g., 15 x 7). Those in the SP group then restudied the 30 items at a rate of 4 s per item, whereas those in the RP group were given 2 min to freely recall as many items as they could remember. On this test and all subsequent recall tests in this experiment, participants were not given feedback as to the correctness of their answers. This procedure (item presentation and sentence generation, math, and restudy or free recall) was then repeated for the 30 items of the other item-type. During a subsequent 5-min retention interval, participants worked on a Sudoku puzzle. SP participants then engaged in restudy of all 30 words followed by all 30 images at a rate of 3 s per item, whereas RP participants were given 3 min for free recall. During free recall, RP participants were instructed to type as many words and images as they could remember, and to record a “W” next to items that had been presented as words and an “I” next to items that had been presented as images. Another 5-min retention interval followed, in which participants worked on a new Sudoku puzzle. SP participants then engaged in a final round of restudy (60 items, 3 s per item) and RP participants were given 3 min for free recall.

After encoding, participants filled out the first iteration of the STICSA. Participants were then paid (when applicable) and excused. The Day 1 experimental procedure lasted approximately 45 min.

On Day 2, participants returned to the original testing room 24 hr after the first session. The undergraduate experimenter first asked them to complete the STICSA for a second time. She then fastened an Empatica E4 wristband around each participant’s non-dominant wrist and instructed participants to sit still for one min while the devices collected a baseline measure of heart rate. A graduate-student experimenter then entered the room, dressed in business attire. She gave each participant a blank sheet of paper, and instructed them to prepare a speech in which they would be applying for the position of a teaching assistant in a class of their choice. After 2 min, an experimenter abruptly took participants’ notes away and instructed them that they would give their speeches extemporaneously. The graduate experimenter turned on the video camera and told participants that they would be video recorded during their speeches for the purpose of coding their non-verbal behavior at a later time. The graduate experimenter then called on participants in random order to deliver 1-min speeches. During each speech, the experimenters took notes on a clipboard and withheld verbal and non-verbal feedback.

After giving their speeches, participants were given Test 1, a pen-and-paper free recall test in which they were asked to recall *either* the words *or* the images that they had learned the previous day. Initial recall of words or images was counterbalanced. Participants were given 2.5 min for Test 1.

Participants were next called on randomly to orally subtract numbers in the teens from four-digit numbers (e.g., 4,866 - 19). Each participant was called on multiple times during the 6-min subtraction phase. The TSST-G stress induction took approximately 15 min, including the speech preparation, individual speeches, Test 1, and subtraction task.

After the math portion of the TSST-G, participants completed the STICSA for the third and final time. During a subsequent 10-min resting period leading up to the final memory test, participants viewed part of an episode of the NBC television series *The Office*. Afterward, an experimenter gave participants Test 2, prompting them to recall items of the item-type (words or images) that had not been assessed on Test 1. After 2.5 min had passed, the experimenter collected the tests.

Following Test 2, participants were paid (when applicable) and debriefed. The experimenter explained to participants that they had not actually been videotaped and that the experimenters were not judging their non-verbal behavior. The Day 2 experimental procedure lasted approximately 45 min.

Non-stressed Group. The procedure for the non-stressed group followed the same protocol as discussed for the stress group, with the exception that the non-stressed group did not receive the TSST-G stress manipulation on Day 2. The TSST-G protocol for non-stressed participants followed von Dawans et al. (2011) and was designed to mimic the procedure for stressed participants without the components of socio-evaluative threat and unpredictability. In place of the 2-min speech preparation phase and the 4-min speech phase, participants in the non-stressed group sat and silently read a chapter from a biology textbook for 6 min. In place of the 6-min oral math subtraction task, participants in the non-stressed group solved math subtraction problems with pen and paper for 6 min. They were given as much time as they needed to complete each problem and were told that their answers would not be graded. The two experimenters were present in the room during these tasks but wore casual clothing and did not observe or question the participants. All other experimental procedures were identical to those of the stressed group.

Experiment 1 Results

Physiological Arousal

To determine whether the TSST-G induced a physiological stress response, blood volume pulse and inter-beat interval (IBI) were examined as measures of heart rate variability. Blood volume pulse was measured in number of beats per min, and IBI was measured in number of milliseconds between heart beats. The MATLAB Kubios software package (see http://kubios.uef.fi/) was used to calculate each participant's average pulse and IBI over the span of the 1-min baseline measurement and over the span of the 12-min TSST-G task. The 12-min TSST-G measurement did not include the 2.5-min memory test (Test 1) that occurred in the middle of the TSST-G on Day 2, since the test was not part of the stress manipulation.

One Empatica E4 wristband was unpredictably faulty, resulting in physiological data for only 104 of 120 participants. Thus, the following analyses were conducted on 58 participants who completed the non-stressed TSST-G tasks, and 46 participants who completed the stress-induction tasks.

Paired-samples *t* tests were used to compare mean pulse and IBI during the stressed and non-stressed TSST-G tasks to mean activity during the baseline measurement on Day 2. As expected, stressed participants demonstrated post-stress increases in pulse, $t(45) = 7.53$, $p < .001$, $d = 0.82$, and decreases in IBI, $t(45) = 5.05$, $p < .001$, $d = 0.38$, relative to baseline, whereas non-stressed participants did not show changes in pulse, $t(57) = 1.77$, $p = .08$, or IBI, $t(57) = 0.21$, $p = .84$. Pulse and IBI averages are reported in Table 1.

Self-reported Stress

Because the act of taking tests may be stressful for some participants, I first examined whether the Day 1 manipulation (retrieval practice vs. study practice) affected participants' subsequent self-reported levels of stress. An independent samples *t* test revealed no difference in STICSA scores for participants who had engaged in retrieval practice versus study practice, $t(118) = 0.56$, $p = .58$.

To test whether the TSST-G tasks increased subjective anxiety on Day 2, paired-samples *t* tests were used to compare pre- and post-TSST-G STICSA scores. As expected, stressed participants demonstrated heightened post-stress STICSA scores relative to baseline, $t(59) = 3.30$, $p < .001$, $d = 0.27$, whereas non-stressed participants did not, $t(59) = 0.54$, $p = .60$. Average STICSA scores are reported in Table 1.

Table 1. Experiment 1: Average STICSA scores, pulse, and IBI as a function of TSST-G group (stressed vs. non-stressed). Standard errors of the mean are given in parentheses.

	Non-stressed		Stressed	
Measure	Baseline	During TSST-G Task	Baseline	During TSST-G Task
STICSA Score	28.6 (*0.83*)	29.0 (*0.95*)	31.6 (*1.13*)	33.9 (*1.08*)
Pulse (*bpm*)	75.0 (*1.55*)	75.9 (*1.69*)	73.1 (*1.93*)	84.7 (*2.11*)
IBI (*ms*)	883.7 (*21.05*)	889.7 (*22.13*)	929.7 (*25.67*)	866.4 (*22.18*)

Day 1 Memory Performance

Table 2 displays correct recall averages for the participants who were given the RP manipulation on Day 1. On the final two recall tests (T_2 and T_3) on Day 1, participants were instructed to recall both words and images and indicate the source from which each item came (i.e., the word list or the image list). On average, participants made 0.4 (*SEM* = 0.09) source misattributions on T_2 and 0.4 (*SEM* = 0.10) source misattributions on T_3. Because Day 1 memory performance was not relevant to the questions posed by the present study, it was not further examined.

Table 2. Experiment 1: Average number of items correctly recalled on Day 1 for participants given the RP manipulation (participants given the SP manipulation did not engage in recall on Day 1). Standard errors of the mean are given in parentheses. *Note:* Words and images were tested in separate blocks on the first test (T_1). On the last two tests (T_2 and T_3), participants recalled words and images together (see *Materials and Methods* for more detail).

	T_1		T_2		T_3	
	Words	Images	Words	Images	Words	Images
Negative	6.6 (*0.26*)	6.9 (*0.32*)	4.8 (*0.25*)	6.1 (*0.28*)	5.3 (*0.27*)	6.4 (*0.29*)
Neutral	5.9 (*0.30*)	5.7 (*0.24*)	4.3 (*0.29*)	5.0 (*0.29*)	4.5 (*0.28*)	5.7 (*0.30*)
Total	12.5 (*0.43*)	12.4 (*0.47*)	9.0 (*0.44*)	11.0 (*0.47*)	9.8 (*0.46*)	12.1 (*0.49*)

Day 2 Memory Performance: Approach to Data Analysis

In the analyses on Day 2 memory performance, memory for words and images was not examined separately. To control for picture superiority effects, I counterbalanced whether words or images were recalled during the first memory test, with the other item-type being recalled during the second test. Because words and images were recalled by an equal number of participants on each memory test, I collapsed across item type when examining means. Furthermore, gender differences have been reported in both the stress and memory literature (e.g., Buchanan & Tranel, 2008; Schwabe & Wolf, 2014), as well as in the broader memory literature (e.g., Herlitz, Nilsson, & Bäckman, 1997; Lewin, Wolgers, & Herlitz, 2001), with women outperforming men on tests of verbal memory. To control for these gender differences, gender was included as a covariate in all following analyses.

Day 2: Recall During the Immediate Stress Response

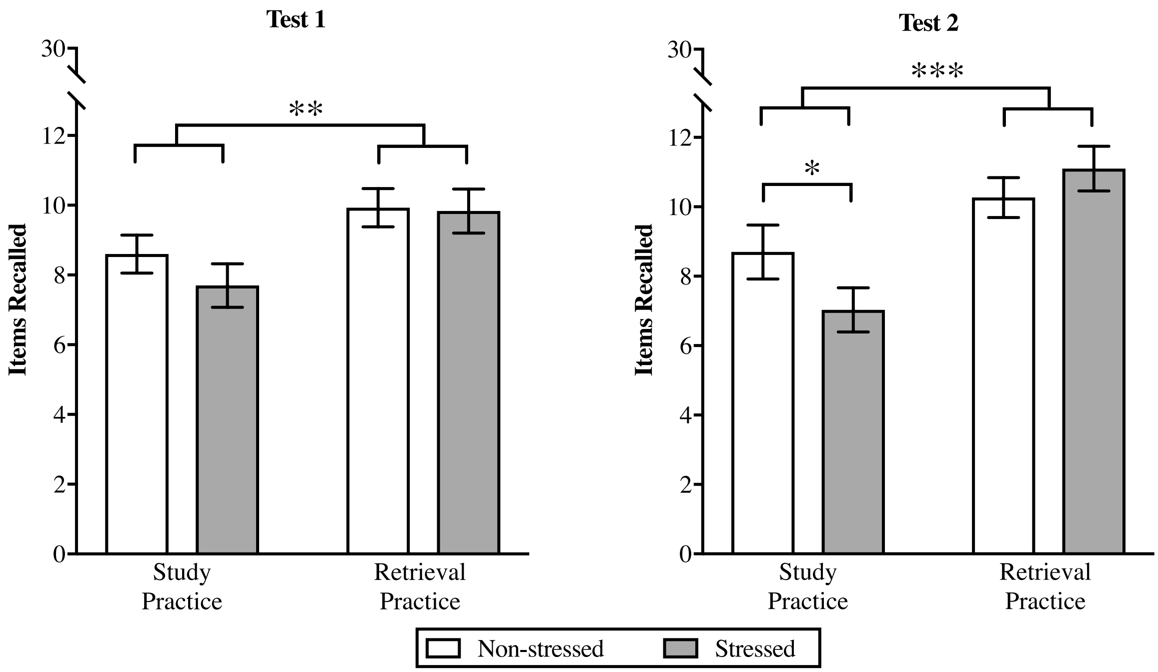
A two-way analysis of covariance (ANCOVA) was conducted to determine the effects of learning strategy (RP vs. SP) and TSST-G group (stressed vs. non-stressed) on Test 1 recall, controlling for gender. The analysis returned a main effect of learning strategy, as those who learned via RP recalled more items than those who learned via SP (9.88 vs. 8.15 items), F (1, 115) = 7.60, p = .007, η_p^2 = .06. Confirming the prediction that gender would act as a covariate, there was also a significant main effect of gender, with females recalling more items on average than males (9.58 vs. 8.17 items), F (1, 115) = 4.12, p = .045, η_p^2 = .04. All other effects, including the main effect of the TSST-G manipulation, were non-significant (all p's > .10). Average Test 1 recall is displayed in Figure 1.

Day 2: Recall During the Delayed Stress Response

Planned independent samples *t* tests were conducted to compare Test 2 recall performance for stressed versus non-stressed participants, all of whom learned via SP. This analysis served to replicate the several previous studies that found impaired memory performance for stressed participants when retrieval was assessed after a post-stress delay. Indeed, stressed SP participants recalled significantly fewer items than non-stressed SP participants on Test 2, $t\ (58) = 1.71$, $p = .047$, $d = 0.44$.

A two-way ANCOVA was then conducted to determine the effects of learning strategy (RP vs. SP) and TSST-G group (stressed vs. non-stressed) on Test 2 recall, again controlling for gender. Most notably, the analysis revealed a significant learning strategy by TSST-G group interaction, $F\ (1, 115) = 4.24$, $p = .042$, $\eta_p^2 = .04$. As shown in Figure 1, stress impaired recall performance for individuals who learned via SP, but did not affect memory for those who learned via RP. This analysis also uncovered a main effect of learning strategy, as those who learned via RP recalled more items than those who learned via SP (10.68 vs. 7.88 items), $F\ (1, 115) = 16.65$, $p < .001$, $\eta_p^2 = .13$. Last, once again confirming the influence of gender as a covariate, females recalled more items than males (9.86 vs. 8.42 items), $F\ (1, 115) = 3.20$, $p = .08$, $\eta_p^2 = .03$. Average Test 2 recall is displayed in Figure 1.

Figure 1. Experiment 1: Number of items accurately recalled on Test 1 and Test 2 on Day 2. Error bars represent *SEM*. *$p < .05$, ** $p < .01$, *** $p < .001$



Day 2: Source Misattributions

Two exploratory two-way ANCOVAs (controlling for gender) were conducted to determine whether learning strategy (RP vs. SP) and TSST-G group (stressed vs. non-stressed) affected participants' source misattributions on Test 1 and Test 2 (e.g., incorrectly reporting a word as an image). All main effects and interactions were non-significant (all *p*'s > .10).

Experiment 1 Discussion

Though a strong testing effect emerged on both Test 1 and Test 2, only Test 2 performance was affected by stress. This pattern of results was anticipated, as stress-related memory impairment is typically observed 20-30 min after the onset of stress but not necessarily during the first phase of the stress response (Schönfeld et al., 2014; Schwabe & Wolf, 2014). As hypothesized, on Test 2, participants who engaged in study practice at encoding demonstrated typical stress-related memory deficits, whereas those who engaged in retrieval practice were immune to these deleterious effects. Not only did stressed individuals in the retrieval-practice group outperform those in the study-practice group, they demonstrated memory performance comparable to that of individuals who engaged in retrieval practice but were not stressed. Thus, retrieval practice served to completely inoculate memory against psychological stress, as if stress had not even been present for those individuals. The analyses on heart rate suggest that the TSST-G manipulation successfully induced stress, supporting the conclusion that retrieval practice, and not insufficient stress induction, was responsible for the observed results.

These initial findings are consistent with both prior empirical evidence and theories about mechanisms underlying the effects of stress on memory. In several previous experiments on the topic, all participants engaged in retrieval practice during encoding on the first day of the experiment. In two such studies, stress did not have the usual impairing effects on retrieval when memory was tested the next day (Schoofs & Wolf, 2009; Wolf et al., 2002). Because retrieval practice was not experimentally manipulated in those experiments, researchers could not determine whether the use of this strategy at encoding had created stress-resistant memories. The results of the present Experiment 1 strongly suggest that retrieval practice was, indeed, the reason for the null findings reported in these studies.

As discussed in the introduction, the theories surrounding stress effects and retrieval practice effects share a common mechanism. A growing body of evidence suggests that psychological stress causes a mental context shift, which disrupts context-dependent retrieval (see Shields et al., 2017). Retrieval practice is believed to enhance memory access to contextual details that help reinstate context during retrieval (Karpicke et al., 2014). In the presence of stress, retrieval practice may improve access to the contextual cues that individuals need to reinstate mental context and successfully retrieve information. The finding that retrieval practice bolstered memory against the deleterious effects of stress is consistent with this hypothesis. However, a more direct measure of memory for contextual details would provide stronger support for this idea.

In addition to not providing a direct test of the context-shift hypothesis, the results from Experiment 1 do not speak to whether multiple recall attempts during encoding are necessary for creating stress-resistant memories, or whether one retrieval attempt is sufficient. This question has applied relevance, as self-report (Kornell & Bjork, 2007; Wissman, Rawson, & Pyc, 2012) and experimental (Kornell & Bjork, 2008) evidence suggests that most students who use self-testing (e.g., flashcards) will only test themselves until they can accurately recall each desired item once. Thus, determining whether one successful retrieval attempt while studying can buffer memory against stress to the same degree as three retrieval attempts would provide useful information for learners who already default to a criterial learning strategy. Further, knowing whether criterial learning of this nature effectively creates stress-resistant memories would help optimize the time that individuals spend studying and self-testing in preparation for a stressful event. In Experiments 2a and 2b, I aimed to determine whether learning to a criterion of one correct recall could bolster memory against stress, or whether additional retrieval practice would be necessary to achieve that goal.

CHAPTER 5: EXPERIMENTS 2A AND 2B

Experiments 2a and 2b addressed a key limitation of Experiment 1: whether multiple (i.e., three) retrieval practice attempts are needed to buffer memory against psychological stress, or whether one successful attempt is sufficient. In two experiments, participants engaged in a criterial learning task followed by either additional restudying or additional retrieval practice. Thus, participants either successfully recalled information one time (study-practice group) or up to four times (retrieval-practice group). After a 24-hr (Experiment 2a) or one-week (Experiment 2b) delay, participants returned to the lab for stress induction followed by cued recall testing. The key difference between the two experiments was the retention interval between encoding and post-stress retrieval. The advantages of retrieval practice over restudying have been shown to become increasingly pronounced as a function of the amount of time that has passed between encoding and retrieval (e.g., Thompson, Wenger, & Bartling, 1978; Wheeler, Ewers, & Buonanno, 2003). In Experiment 2a, I aimed to test whether the benefits of additional, post-criterial-learning retrieval practice would emerge after the typical 24-hour delay. In Experiment 2b, I aimed to determine whether these benefits would be more pronounced after a longer delay of one week.

Another difference between experiments 2a and 2b concerns the timing of the final memory test. Motivated by the finding that retrieval is sometimes enhanced during the first phase of the physiological stress response (Hupbach & Fieman, 2012), in Experiment 2a, post-stress memory performance was tested both immediately after the onset of stress and during the cortisol peak 30 min later. In Experiment 2b, for reasons discussed in the method section, memory performance was measured only during the cortisol peak.

The criterial learning task used in the present studies was modelled after Karpicke and Roediger (2008), who found that participants who engaged in criterial learning followed by additional studying (CL_S) recalled 36% of stimuli one week later whereas those for whom criterial learning was followed by additional retrieval practice (CL_R) recalled 80% of stimuli. Based on their findings, I hypothesized that CL_R participants would outperform all CL_S participants. Based on the results from Experiment 1, I further hypothesized that stressed and non-stressed CL_R participants would demonstrate similar memory performance, but that stressed CL_S participants would demonstrate lower cued recall than those in the non-stressed CL_S group. I expected that this pattern of results would emerge in both experiments but would be more pronounced after the one-week delay in Experiment 2b than after the 24-hr delay in Experiment 2a.

Experiment 2a Method

Design

The experiment employed a 2 (learning strategy: CL_S or CL_R) x 2 (TSST-G group: non-stressed or stressed) between-subjects factorial design.

Participants

Assuming an effect size of $\eta_p^2 = .04$ derived from Experiment 1, a significance level of $\alpha = .05$, and four between-subjects groups, an a priori power analysis revealed that a total sample size of 191 participants (approximately $n = 47$ per group) would provide a minimum acceptable value of 80% power to detect effects (G*Power 3.0; Faul, Erdfelder, Lang, & Buchner, 2007). However, the experiment was terminated part-way through data collection when preliminary analyses revealed null effects backed by strong statistical support (see the *Results* section for more detail). Thus, 133 Tufts University students participated in the experiment. Twenty participants were excluded from data analysis for the following reasons: experimenter error (n = 4), prior knowledge of Swahili (n = 1), failure to return for the second experimental session (n = 3), or perfect memory performance (n = 12). As such, all final analyses were conducted on 113 participants (75 females, M_{age} = 18.96, SD_{age} = 1.41), all of whom reported that they had not consumed caffeine or nicotine in the 6 hr prior to the experiment. Participant recruitment and compensation were identical to Experiment 1. Twenty-eight participants were randomly assigned to the non-stressed CL_S group, 27 to the non-stressed CL_R group, 28 to the stressed CL_S group, and 30 to the stressed CL_R group. All participants provided informed consent.

Materials

Stimuli. The learning materials consisted of 40 Swahili nouns, each paired with their English translation (e.g., *bustani – garden*). The stimuli were chosen because they had been used in a prior study on criterial learning and did not result in ceiling memory performance (Karpicke & Roediger, 2008). A full list of the stimuli can be found in Karpicke and Roediger's (2008) supplemental materials.

Anxiety questionnaire. As in Experiment 1, the STICSA was administered to assess participants' self-reported levels of pre- and post-stress anxiety.

Procedure

Testing sessions occurred on two consecutive days between 3:30 p.m. and 5:30 p.m. to control for variability in diurnal cortisol secretion (Weitzman et al., 1971). Stressed participants were tested four at a time according to stress induction protocol, and non-stressed participants were tested in groups of 2-4.

On Day 1, participants first began the criterial learning task, which was presented using E-Prime software (Version 2.1; Schneider et al., 2001). This task consisted of four study/test trials, the nature of which differed depending on whether a given participant was randomly assigned to the CL_S or CL_R group. Prior to the first trial, participants in the CL_S group were instructed that they would complete study/test trials in which they would learn Swahili words, and that every item they correctly recalled on each test would be removed from all subsequent tests. Participants in the CL_R group were similarly instructed that they would complete study/test trials, but that every item they correctly recalled on each test would be removed from all subsequent study periods. A final note in the instructions encouraged participants to provide correct answers on each test so as to finish the experiment faster.

The first of the four study/test trials was the same for the CL_S and CL_R groups. Participants viewed 40 Swahili-English word pairs, presented in random order for 5 s each. A 30-s retention interval followed, in which participants completed simple math problems (e.g., 12 x 9). Afterward, participants took a cued recall test, in which they were presented with all 40 Swahili words and were asked to type each English translation. The Swahili words were presented individually and in alphabetical order. Participants were given 8 s to enter a response before the program advanced to the next Swahili cue word. This study-math-test procedure was then repeated for three additional trials, but these trials differed according to the CL_S or CL_R group. During each study period, participants in the CL_S group restudied all 40 word pairs whereas those in the CL_R group only restudied word pairs that they had not yet correctly recalled on any previous cued recall test. During each test period, participants in the CL_S group were prompted to recall only words that they had not yet correctly recalled on any previous cued recall test, whereas those in the CL_R group were prompted to recall all 40 English target words. The program ended once all four study/test trials had been completed. In cases where CL_S participants correctly recalled all 40 items before the fourth trial, no subsequent cued recall tests were given and thus they engaged in back-to-back study periods. When CL_R participants correctly recalled all 40 items before the fourth trial, no subsequent restudying was required and thus they completed back-to-back cued recall tests.

After encoding, participants filled out the first STICSA. Participants were then paid (when applicable) and excused. The Day 1 experimental procedure lasted approximately 45 min.

Twenty-four hr after encoding, participants returned to the original testing room where they completed the second STICSA and provided a baseline saliva sample for cortisol analysis. Participants then began the procedure associated with either the control or stress version of the TSST-G (von Dawans et al., 2011). As described in Experiment 1, participants in the stress groups first prepared and performed a short speech whereas participants in the control groups read silently from a textbook. After the speech or reading task, all participants provided the second saliva sample for cortisol analysis.

All participants were next given Test 1 to assess memory performance during the immediate stress response. They were presented with a list of 20 of the 40 Swahili cue words on a single sheet of paper and were instructed to fill in the English translations in the blank spaces next to each word. Participants were given 2.5 min to complete Test 1.

The TSST-G procedure then continued, as participants in the stress groups performed oral math subtraction for 6 min whereas those in the control groups solved the same math problems using pen and paper (see Experiment 1 Method for more detail). The TSST-G procedure, including Test 1, lasted for 15 min for both the stress and control groups.

After the TSST-G, participants completed the third STICSA. During a subsequent 10-min resting period leading up to the final memory test, participants viewed part of an episode of the NBC television series *The Office*. Afterward, participants provided the third saliva sample as a measure of the delayed stress response in which cortisol reaches peak post-stress levels. They then completed Test 2, which was identical to Test 1 but assessed participants' memory for the 20 Swahili words that were not previously tested. Participants were again given 2.5 min to complete the test. Tests 1 and 2 were counterbalanced, such that participants were presented with the 20-item tests in one of two orders.

Following Test 2, participants were paid (when applicable) and debriefed. Those in the stress groups were told that they had not actually been videotaped and that the experimenters were not judging their non-verbal behavior. The Day 2 experimental procedure lasted approximately 45 min.

Cortisol Measurement and Data Management

Three saliva samples were collected using the passive drool method: one at baseline and one each at 10 min and 30 min after the onset of the stress or control task. Samples were stored at -20ºC 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.

Researchers have reported variability in cortisol reactivity to the TSST (e.g., Buchanan & Tranel, 2008; Schönfeld et al., 2014), with individuals referred to as "responders" showing increased post-stress cortisol levels relative to baseline, and "non-responders" showing no change or decreased post-stress cortisol. As is common practice in research on stress and memory, in the subsequent analyses on Day 2 memory performance, analyses were conducted across all participants and then were repeated after eliminating non-responder participants in the stress groups. There were eight non-responders (seven females) in the stressed CL_S group and three non-responders (all females) in the stressed CL_R group, reducing the sample size in each of these groups to 20 and 27, respectively.

Experiment 2a Results

Self-reported Stress

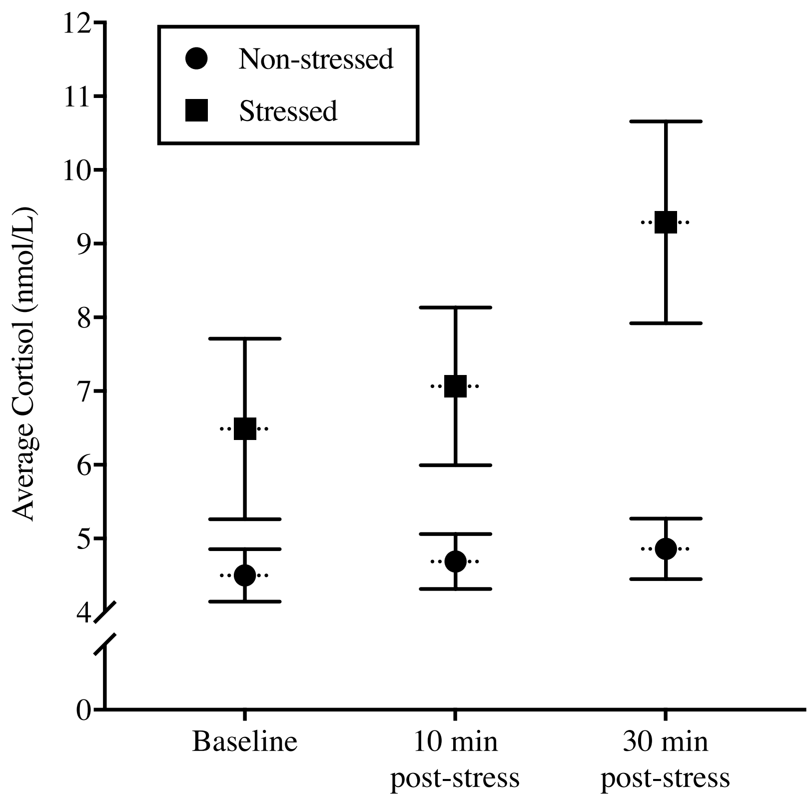
Because the act of taking tests may be stressful for some participants, the first analysis on STICSA scores compared self-reported level of stress on Day 1 for individuals in the CL_S and CL_R groups. An independent samples *t* test on average Day 1 STICSA scores revealed no difference between participants who had engaged in additional retrieval practice versus additional study practice, $t(111) = 1.07$, $p = .286$.

To test whether the TSST-G tasks increased subjective anxiety on Day 2, paired-samples *t* tests were used to compare pre- and post-TSST-G STICSA scores. As expected, stressed participants demonstrated heightened post-stress STICSA scores relative to baseline, $t(54) = 5.67$, $p < .001$, $d = 0.83$, whereas non-stressed participants did not, $t(54) = 0.84$, $p = .404$.

Cortisol

Cortisol reactivity to the TSST-G tasks was next examined. A 3 (time: baseline, 10 min post-stress, 30 min post-stress) x 2 (TSST-G group: stressed or non-stressed) mixed model ANOVA revealed a significant interaction, $F(2, 214) = 4.22$, $p = .016$, $\eta_p^2 = .04$. Pairwise comparisons using a Bonferroni correction showed that, as expected, stressed participants demonstrated increased cortisol from baseline to 30 min post-stress (mean difference = 2.85, *SEM* = 0.54, $p < .001$) and from 10 min post-stress to 30 min post-stress (mean difference = 2.37, *SEM* = 0.80, $p = .011$), whereas non-stressed participants demonstrated no differences in cortisol across the three measurements (*p*'s > .10). Thus, the TSST-G manipulation successfully induced a physiological stress response for individuals in the stress groups. Average cortisol concentrations are shown in Figure 2.

Figure 2. Experiment 2a: Average cortisol concentrations for stressed and non-stressed participants on Day 2. Cortisol samples were taken immediately prior to the TSST-G (baseline), 10 min after the onset of the TSST-G, and 30 min after the onset of the TSST-G. Error bars represent *SEM*.

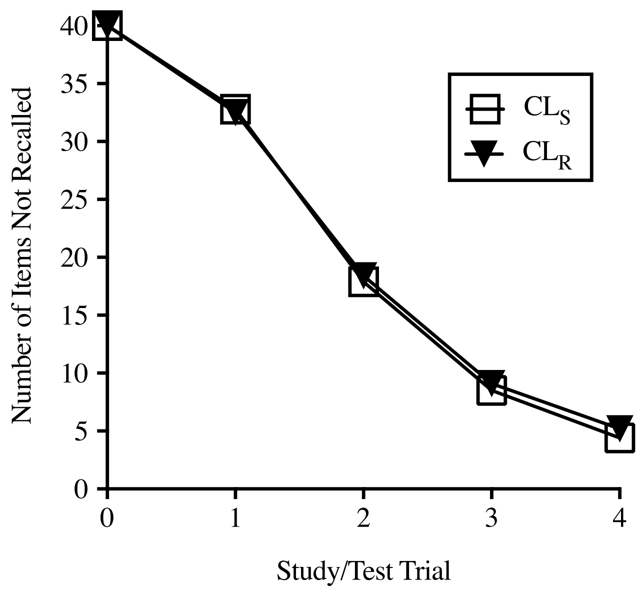


Day 1 Criterial Learning

As shown in Figure 3, there were no differences in the learning curves for participants who engaged in additional study practice or additional retrieval practice during the criterial learning task on Day 1. This was confirmed by a 2 (learning strategy: CL_S, CL_R) x 4 (study/test

Trial: 1, 2, 3, 4) mixed ANOVA with study/test trial as a within-subjects variable, in which there was no main effect of learning strategy on how many items participants correctly remembered on each cued recall test, $F(1, 111) = 0.11$, $p = .746$. By the end of the task, participants had accurately recalled an average of 35 out of 40 items at least once, or approximately 88%.

Figure 3. Experiment 2a: Average number of English word targets that had not been correctly recalled on any previous study/test trial.



Day 2 Memory Performance

As in Experiment 1, gender was included as a covariate in all of the following omnibus analyses.

Test 1. A two-way ANCOVA was conducted to determine the effects of learning strategy (CL_S vs. CL_R) and TSST-G group (stressed vs. non-stressed) on Test 1 cued recall, controlling for gender. All main effects and the interaction were nonsignificant (all p's > .10). When this analysis was repeated after eliminating cortisol non-responders, all effects were again nonsignificant (all *p*'s > .10). Average Test 1 cued recall performance is displayed in Table 3.

Test 2. Test 2 cued recall performance was examined using identical analyses. Again, all main effects and interactions were nonsignificant for the analysis on all participants and the analysis excluding non-responders (all *p*'s > .10). Test 2 cued recall performance is also displayed in Table 3.

The null effect found for Test 2 performance stands in contrast to the Experiment 1 results, which featured a significant interaction between learning strategy and TSST-G group on the second memory test. To further support the null finding reported here, the Test 2 data were examined by estimating Bayes factors for both main effects and the interaction. Bayes factors have emerged as a way to support the conclusion that null results are sound evidence in favor of the null hypothesis and are not the result of data insensitivity (Dienes, 2014). Unlike frequentist approaches, such as power estimates that are based on pre-existing values and assumptions (e.g., alpha values and estimated effect sizes), Bayes factors rely only on the observed data to determine whether evidence favors the null or alternative hypothesis. Bayes factors range from 0 to ∞, with values close to 0 being strong evidence in favor of the null hypothesis, values much greater than 1 being strong evidence in favor of the alternative hypothesis, and values around 1 suggesting that the data are insensitive (Dienes, 2014). Bayes factors were analyzed using the statistical software program JASP (2017), assuming the default recommended fixed-effect and random-effect prior values. The estimated Bayes factors (BF_{10}) for the effects of learning strategy, TSST-G group, and the interaction were 0.20, 0.20, and 0.31, respectively. All three values are considered strong evidence that the data are more likely to occur under the null than the alternative hypothesis (Jeffreys, 1939/1961), further supporting the null findings of the previous ANCOVA.

Table 3. Experiment 2a: Average cued recall performance on Test 1 and Test 2. Standard errors of the mean are given in parentheses.

Group	Test 1	Test 2
Non-stressed		
CL_S	16.0 (*0.77*)	15.4 (*0.65*)
CL_R	15.3 (*0.92*)	15.1 (*0.95*)
Stressed		
CL_S	15.3 (*0.92*)	14.9 (*0.92*)
CL_R	16.5 (*0.62*)	15.3 (*0.71*)

Experiment 2a Discussion

Regardless of whether participants engaged in additional studying or retrieval practice after criterial learning on Day 1, or whether they were subjected to stress induction or a non-stressful task on Day 2, participants in Experiment 2a demonstrated similar final cued recall performance. These findings stand in contrast to the majority of studies examining the effects of stress on memory, as stressed individuals did not demonstrate poorer memory performance than those who were not stressed. The present cortisol results suggest that these discrepant findings are not due to an ineffective stress-induction procedure; the majority of stressed participants demonstrated post-stress increases in cortisol, and stressed participants as a group showed a strong post-stress increase in cortisol. Instead, these results further support the notion that engaging in retrieval, during learning, increases memory accessibility during later retrieval attempts. Specifically, an initial successful retrieval attempt during learning may be sufficient for creating stress-resistant memories, regardless of whether individuals perform additional studying or retrieval attempts. These results further refine the Experiment 1 findings, in which three retrieval attempts during learning resulted in stress-resistant memories, but simply restudying material during learning yielded the typical stress-related memory impairment. When just a 24-hour delay separates encoding and retrieval, only one successful retrieval attempt may be necessary in order to buffer memory against stress.

Because cued recall performance was generally high in all groups, the question remains whether the benefits of additional retrieval practice after learning to criterion would emerge after a longer delay when more forgetting has occurred. In fact, in several studies, the benefits of retrieval practice over restudying did not emerge until 48 hr or even 1 week after initial learning (for a review see Roediger & Karpicke, 2006b). Thus, in Experiment 2b, I aimed to determine whether learning to criterion would continue to buffer memory against stress after a one-week delay, or whether additional retrieval practice would provide a benefit over additional restudying.

Experiment 2b Method

Design

The experiment employed a 2 (learning strategy: CL_S or CL_R) x 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 underwent TSST-G stress induction. Non-stressed and stressed cued recall performance was measured by pre- and post-TSST-G tests.

Participants

Assuming an effect size of $\eta_p^2 = .04$ derived from Experiment 1, a significance level of $\alpha = .05$, two between-subjects groups, two within-subjects measurements, and a conservative .70

correlation between repeated measures (in Experiment 2a, the correlation between Test 1 and Test 2 performance for non-stressed participants was .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 et al., 2007). Sixty-five Tufts University students were initially recruited to participate in the experiment. However, seven participants were excluded from data analysis because they did not return for the second experimental session, and two participants were excluded because they did not recall any items on one or both of the cued recall tests on Day 2. Thus, all final analyses were conducted on 56 participants (35 females, M_{age} = 19.54, SD_{age} = 1.38), all of whom reported that they had not consumed caffeine or nicotine in the 6 hr prior to the experiment. Participant recruitment and compensation were identical to Experiment 1. Twenty-seven participants were randomly assigned to the CL_S group and the other 29 were assigned to the CL_R group. All participants provided informed consent.

Materials

Materials included the same Swahili-English word pairs and STICSA that were used in Experiment 2a.

Procedure

Testing sessions occurred on two days between 3:30 p.m. and 5:30 p.m. with a one-week delay between Day 1 and Day 2 testing. All participants were tested two at a time. In a few cases in which only one participant showed up for the experiment, a research assistant served as a confederate in the experiment.

The Day 1 procedure (i.e., criterial learning and subsequent studying or retrieval practice) was identical to that of Experiment 2a. One week later, on Day 2, participants returned to the original testing room, where they completed a second STICSA and provided a baseline saliva sample for cortisol analysis. All participants were then given 2.5 min to complete Test 1, which was the same Test 1 used in Experiment 2a. They then performed all tasks associated with TSST-G stress induction, which is described in detail in the Experiment 1 method. Briefly, this 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. As in Experiments 1 and 2a, participants next viewed part of an episode *The Office* during a 10-min retention interval. Afterward, they provided the third saliva sample and were given 2.5 min to complete Test 2, which was the same Test 2 used in Experiment 2a. Last, participants were paid (when applicable) and debriefed. The Day 2 experimental procedure lasted approximately 45 min.

Note that in Experiment 2b a cued recall test was not administered immediately after the onset of stress as in Experiment 2a. Because assessing memory during the first phase of the stress response was not a primary goal of this experiment, the two memory tests were administered in a different way in Experiment 2b: pre- and post-stress. This allowed us to manipulate stress within-subjects, thereby decreasing error variance.

Cortisol Measurement and Data Management

Saliva samples for cortisol analysis were collected and handled as outlined in the Experiment 2a method. As in Experiment 2a, I examined Day 2 memory performance for all participants and then repeated these analyses after eliminating non-responders. There were six non-responders (four females) in the CL_S group and 10 non-responders (seven females) in the CL_R group, reducing the sample size in each of these groups to 21 and 19 respectively.

Experiment 2b Results

Self-reported Stress

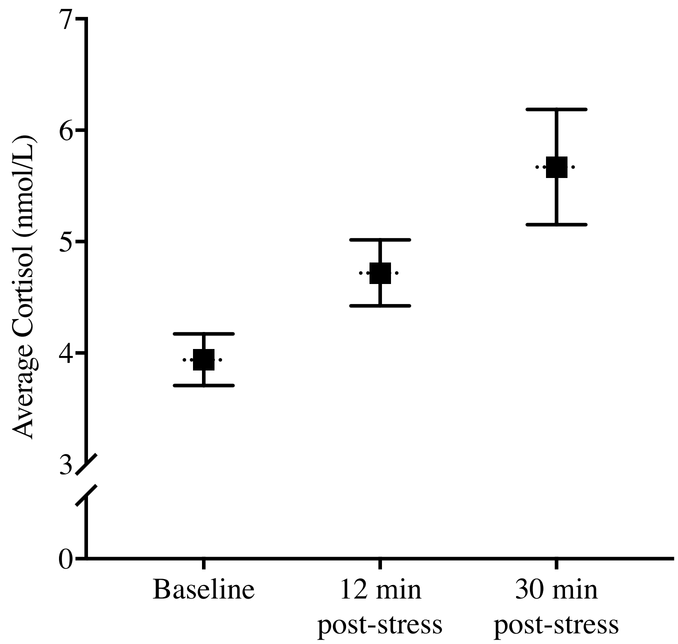
As in Experiment 2a, I first examined whether the Day 1 manipulation (CL_S or CL_R) affected participants' subsequent self-reported levels of stress. An independent samples *t* test on average Day 1 STICSA scores revealed no difference for participants who had engaged in additional retrieval practice versus additional study practice, $t(54) = 0.52$, $p = .609$.

To test whether the TSST-G tasks increased subjective anxiety on Day 2, I collapsed across learning group (CL_S or CL_R) and conducted a paired-samples *t* test comparing pre- and post-stress STICSA scores. As expected, participants demonstrated heightened post-stress STICSA scores relative to baseline, $t(55) = 2.67$, $p = .005$, $d = 0.37$. The average pre-stress STICSA score was 30.79 (*SEM* = 1.20) and the average post-stress score was 33.18 (*SEM* = 1.42).

Cortisol

To examine cortisol reactivity to the TSST-G, I collapsed across learning group (CL_S or CL_R) and conducted a one-way repeated measures ANOVA comparing the baseline cortisol measurement to that taken 12 and 30 min post-stress. As shown in Figure 4, this analysis revealed a main effect of timing, $F(2, 106) = 12.47$, $p < .001$, $\eta_p^2 = .19$. Pairwise comparisons using a Bonferroni correction revealed significant cortisol increases from baseline to 12 min post-stress (mean difference = 0.75, *SEM* = 0.21, *p* = .002), from baseline to 30 min post-stress (mean difference = 1.71, *SEM* = 0.46, *p* = .001), and from 12 min post-stress to 30 min post-stress (mean difference = 0.96, *SEM* = 0.32, *p* = .012).

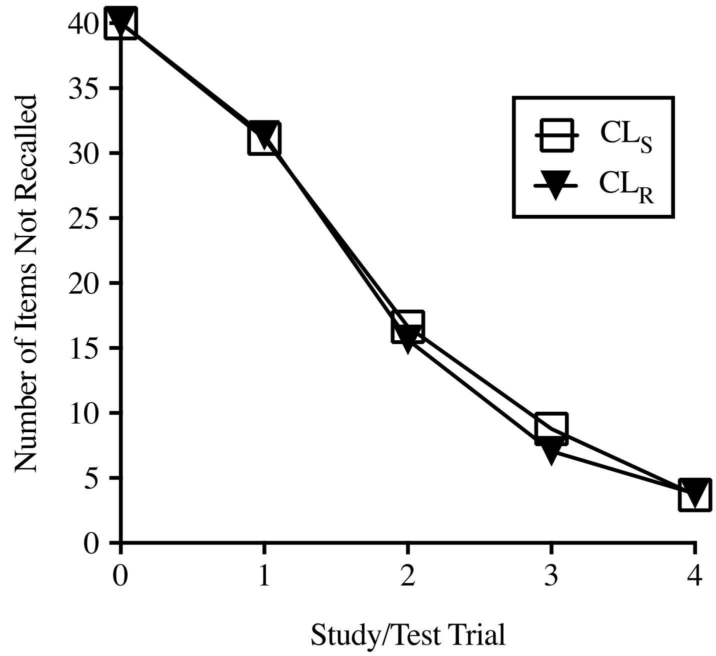
Figure 4. Experiment 2b: Average cortisol concentrations on Day 2. Cortisol samples were taken immediately prior to the TSST-G (baseline), 12 min after the onset of the TSST-G, and 30 min after the onset of the TSST-G. Error bars represent *SEM*.



Day 1 Criterial Learning

As shown in Figure 5, there were no differences in the learning curves for participants who engaged in additional study practice or additional retrieval practice during the criterial learning task on Day 1. This was confirmed by a 2 (learning strategy: CL_S, CL_R) x 4 (study/test Trial: 1, 2, 3, 4) mixed ANOVA, with study/test trial as a within-subjects variable, in which there was no main effect of learning strategy on how many items participants correctly remembered on each cued recall test, $F(1, 54) = 0.11$, $p = .740$. By the end of the task, participants had accurately recalled an average of 36.3 out of 40 items, or approximately 91%, at least once.

Figure 5. Experiment 2b: Average number of English word targets that had not been correctly recalled on any previous study/test trial.



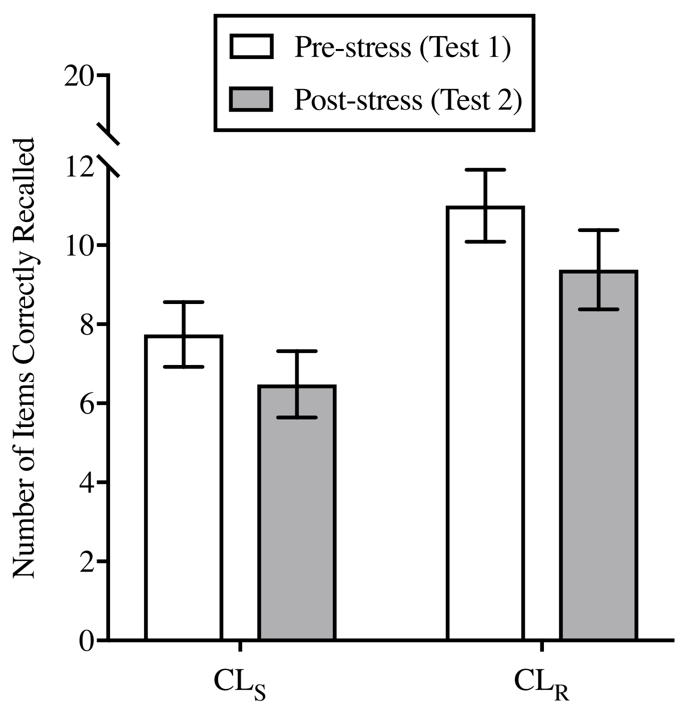
Day 2 Memory Performance

In contrast to Experiments 1 and 2a, gender was not included as a covariate in the analyses on memory performance. The sample size in the present experiment was less than half of the size recruited for the previous two experiments, resulting in a particularly low number of male participants (11 males in the CL_S group, 10 males in the CL_R group). Thus, in order to achieve optimal statistical power, I excluded gender as a covariate.

I first conducted a 2 (learning strategy: CL_S or CL_R) x 2 (test timing: pre-stress or post-stress) ANOVA to determine whether cued recall performance differed according to learning strategy or conditions of stress. This analysis revealed a main effect of stress, as participants recalled fewer words on the post-stress test (M = 9.4) than on the pre-stress test (M = 8.0), $F(1, 54) = 22.21$, $p < .001$, $\eta_p^2 = .29$. Replicating the robust testing effect, a main effect of learning strategy was found, as those in the CL_R group recalled more words when averaged across the two tests (*M* = 10.2) than those in the CL_S group (*M* = 7.1), $F(1, 54) = 6.20$, $p = .016$, $\eta_p^2 = .10$. The interaction between learning strategy and test timing was not significant ($p > .10$).

Repeating this analysis after excluding non-responders, I again found a main effect of test timing, as participants recalled fewer words on the post-stress test (*M* = 9.1) than on the pre-stress test (*M* = 7.7), $F(1, 37) = 17.72$, $p < .001$, $\eta_p^2 = .32$. However, this analysis did not uncover a main effect of learning strategy, $F(1, 37) = 2.31$, $p = .137$, likely because of a lack of statistical power resulting from the reduced sample size. I also did not find an interaction between learning strategy and test timing in the analysis on cortisol responders. Average cued recall performance in Experiment 2b is displayed in Figure 6.

Figure 6. Experiment 2b: Average cued recall performance on Test 1 and Test 2. Error bars represent *SEM*.



Experiment 2b Discussion

In Experiment 2b, the retention interval between encoding and retrieval was increased from 24 hours to one week, in order to determine whether criterial learning would continue to sufficiently support longer-term memory accessibility in the presence of stress. In the context of this one-week delay, some of the hypotheses regarding the effects of the stress and criterial-learning manipulations on episodic retrieval were borne out. Consistent with Karpicke and Roediger (2008), participants who engaged in additional retrieval practice after learning to criterion recalled more items than those who engaged in additional restudying. Further, replicating numerous studies (for reviews see Gagnon & Wagner, 2016; Shields et al., 2017), participants demonstrated lower post-stress than pre-stress recall performance. However, in contrast to Experiment 1, there was no interaction between learning strategy and test timing. Though retrieval practice improved memory relative to study practice, performance was negatively affected by stress in both groups. Thus, retrieval practice again emerged as the superior learning strategy for battling the negative effects of stress, but it did not completely buffer memory against the deleterious effects of stress as was observed in Experiment 1.

The lack of an interaction between learning strategy and test timing may be explained by two methodological differences between Experiment 1 and Experiment 2b. The first pertains to the stimuli used, as participants in Experiment 1 learned simple English nouns and images, whereas those in Experiment 2b learned Swahili-English word pairs. Foreign language learning requires the binding of cues with targets—a task for which we have less prior knowledge and fewer prior associations available than for memorizing words in our native language. Retrieval practice may serve to increase automaticity in responding when stimuli are relatively easy to learn (Experiment 1), but may not have this effect when the stimuli require new associative infrastructure (Experiment 2b). Though researchers have debated the effectiveness of retrieval practice with increasingly complex stimuli (Karpicke & Aue, 2015; van Gog & Sweller, 2015), under standard laboratory conditions, retrieval practice does appear to be efficacious with stimuli of varying complexities (e.g., Eglington & Kang, 2016; Martin, Nguyen, & McDaniel, 2016; Tempel, Loran, & Frings, 2015). The present results and those of Experiment 1 may lend further insight into this debate. Specifically, stress may serve to expose memorial weaknesses that are not evident under non-stressful laboratory conditions. It is possible that retrieval practice creates better memory infrastructure for stimuli of low complexity (e.g., English nouns) than those of higher complexities (e.g., foreign language learning), but that this limitation is only evident when memories are subjected to stress. This hypothesis would benefit from a direct manipulation of stimulus complexity in the context of a paradigm such as that used in the present experiments.

An alternative, but not mutually exclusive, explanation for the lack of an interaction in Experiment 2b relates to the cognitive demand of the cued recall test. Retrieving foreign vocabulary words is simply a more cognitively demanding task than retrieving individual words in one's native tongue. Further, post-stress memory performance has been shown to differ according to the level of cognitive demand that a test places on the test-taker. For example, tests of free recall tend to expose detrimental effects of stress more often than less cognitively demanding recognition tests (e.g., Buchanan & Tranel, 2008). Thus, the post-stress memory impairment observed in both learning groups in Experiment 2b may be attributable to the use of a memory test that was more demanding than that used by Smith et al. (2016). This explanation is well supported by a wealth of studies demonstrating the impairing effects of stress on executive functions such as working memory and attention (for a review see Gagnon & Wagner, 2016). Because executive functions provide crucial support for retrieval processes (see Gagnon

& Wagner, 2016; Levy & Anderson, 2002), more effortful retrieval tasks may be more vulnerable to stress. Because recalling Swahili-English word pairs may be more difficult than recalling English nouns, this retrieval task may be more vulnerable to stress regardless of how information was learned. To further test this hypothesis, future researchers may consider manipulating cognitive load during retrieval in a stress-and-memory paradigm.

Finally, the within-subjects manipulation of stress in Experiment 2b may also have contributed to the null interaction between test timing and learning strategy. Relative to between-subjects designs, like those employed in Experiments 1 and 2a, within-subjects manipulations reduce error variance and provide better control for individual differences in performance. Thus, if stress does indeed impair memory retrieval for individuals who engaged in retrieval practice at encoding, such an effect may be better detected using a within-subjects experimental design such as that used in Experiment 2b.

In addition to the theoretical contributions of the present experiments, the results of Experiments 2a and 2b begin to shed light on the practical implications of using retrieval practice to support post-stress memory. As mentioned, college students who use self-testing to study for exams often stop studying once they have learned to a criterion of one (Kornell & Bjork, 2007; 2008; Wissman et al., 2012). The results of Experiment 2a suggest that this approach may provide sufficient memory support for students cramming for a stressful next-day exam. However, for those studying for exams that are several days or weeks away, such as high-stakes standardized tests, the results of Experiment 2b suggest that additional retrieval practice may yield the best memory access. This advice is consistent with the broader retrieval-practice literature, in which learning to a criterion of three (i.e., three successful retrieval attempts) is recommended for optimizing long-term memory and time spent studying (Rawson & Dunlosky, 2011).

The results of Experiments 1, 2a, and 2b have helped establish and specify the ways in which retrieval practice and psychological stress interact to influence memory retrieval. However, the cognitive mechanism underlying the efficacy of retrieval practice as a post-stress memory enhancer has yet to be determined. An emerging theory suggests that retrieval practice may improve post-stress memory accessibility by providing the stressed brain with the tools to reinstate mental context when stress has otherwise disrupted it (see Chapter 2: Literature Review for elaboration on this). Thus, in Experiment 3, I tested this hypothesis in a first step toward determining why retrieval practice is so effective in the context of stress.

CHAPTER 6: EXPERIMENT 3

In Experiment 3, I investigated whether stress impairs access to contextual information from the learning context, and whether retrieval practice improves post-stress memory by increasing access to these contextual details. To examine memory for contextual information, I 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). In such a paradigm, two lists of stimuli are learned during two encoding sessions that are separated by a distinct interval (e.g., 30 min). During each session, participants either study the list multiple times or study it once and then complete one or more free recall tests. On a final criterial memory test, participants are asked to remember the items learned, as well as to discriminate the list from which list each item came. This task is a test of memory for episodic context: Can participants differentiate between their memories of two events that occurred at distinctly different times? With regard to memory performance, I hypothesized that participants who engaged in retrieval practice at encoding would continue to outperform those who engaged in study practice on the post-stress test. My 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. I hypothesized that discrimination 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.

Experiment 3 Pilot Method

In Experiment 3, participants learned two lists of words on the first day of testing and returned one week later for pre-stress and post-stress recognition tests involving list discrimination. Before I conducted the experiment, it was important to consider whether pre/post testing of this nature could potentially cause interference effects. For example, retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994) or part-set cueing effects (Slamecka, 1968) could occur, both of which would manifest as an impaired ability to retrieve items on the post-stress test due to interference from having recollected items on the pre-stress test (e.g., Gómez-Ariza, Lechuga, & Pelegrina, 2005). Interference of this nature would be unfortunate, as the results would mimic the anticipated effects of stress on the pre/post memory tests. To ensure that any differences in post-stress memory performance were due to the influence of stress and not interference effects, the following experiment served as a pilot study for the stimuli and recognition tests that were used in Experiment 3.

Design

The experiment employed a repeated measures design. The independent variable was the timing of the recognition test: pre- or post-retention interval.

Participants

The sample size was calculated using an a priori power analysis. Because the methods used in this experiment differ greatly from studies examining retrieval-induced forgetting and part-set cueing effects, the effect sizes from those experiments were not used to inform the present study. Instead, a medium effect size of $\eta_p^2 = .06$ was assumed, along with a significance level of $\alpha = .05$, two within-subjects measurements, and a conservative .70 correlation between repeated measures. Based on these specifications, a sample size of 21 participants was recommended so as to provide a minimum acceptable value of 80% power (G*Power 3.0; Faul et al., 2007). Data collection was terminated once useable data from 21 participants were acquired, which occurred after 26 Tufts University students had participated in the experiment (18 females, $M_{age} = 20.95$, $SD_{age} = 2.16$). The five participants who were not included in the analyses demonstrated statistically equal hit rates and false alarm rates, indicating that they were responding based on a pre-established criterion instead of their recollection for the learned material. Participants either received credit toward a research-participation requirement, or were compensated $20 for their time. All participants provided informed consent.

Materials

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 1-7 (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 neutrally valenced) and were given arousal ratings lower than 4.00 on a 1-9 scale (i.e., were not negatively arousing). Lastly, word frequencies were determined from the Brysbaert and New (2009) norms, and the three wordlists were constructed with word frequency equated across them.

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 A or L key. 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.

Procedure

Two experimental sessions took place exactly one week apart. After providing informed consent at the first experimental session, participants began the wordlist encoding procedure. Participants were instructed that they would be presented with a series of words that they should try to remember for a later test. Using E-Prime software (Version 2.1; Schneider et al., 2001), the words from either list 1 or list 2 (counterbalanced) were randomly presented at a rate of 2 s per word. The list was then re-presented two consecutive times in this manner. Between presentations of the list, participants performed simple math problems for 30 s. After the third presentation of the first list, participants viewed an episode of the BBC television series *Planet Earth* for 30 min. 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 (i.e., three presentations separated by 30 s of math). 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 were presented in those respective font colors. After learning the lists, participants were paid (when applicable) and excused. The Day 1 experimental procedure lasted approximately 45 min.

Seven days after encoding, participants returned to the original testing room and completed Test 1, which was self-paced. During a subsequent break, participants watched a 22-min episode of the NBC television series *The Office.* They then completed Test 2, which was also self-paced. Participants were then paid (when applicable) and debriefed. The Day 2 procedure lasted approximately 35 min.

Experiment 3 Pilot Results

Recognition Performance

Hits. Hit proportions were calculated by dividing the number of correctly recognized studied items by the total number of studied items that occurred on the recognition test for each participant. A paired samples *t* test showed no difference in participants' hit proportions on Test 1 (M = .69) versus Test 2 (M = .73), $t(20) = 1.51$, $p = .148$.

False alarms. False alarm proportions were calculated 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. A paired samples *t* test showed no difference in participants' false alarm proportions on Test 1 (M = .31) versus Test 2 (M = .33), $t(20) = 0.75$, $p = .460$.

Zhang and Mueller's A. I next examined memory accuracy via Zhang and Mueller's (2005) non-parametric A statistic. The A statistic is an average of the maximum-area and minimum-area proper ROC curves through each observed point, where each point is a participant's average proportion of hits and proportion of false alarms plotted in a Cartesian plane. Higher A values indicate relatively high hit rates and relatively low false alarm rates, providing a measure of one's ability to discriminate between previously studied words and foils. A paired samples *t* test showed no difference in participants' A scores on Test 1 (M = .75) versus Test 2 (M = .76), $t(20) = 1.40$, $p = .177$.

List Discrimination

To examine list-discrimination performance, source memory scores were calculated for each participant by dividing the total number of hits attributed to the correct source by the total number of hits (e.g., Foley, Johnson, & Raye, 1983; Johnson, Nolde, & Leonardis, 1996; Zeeuws et al., 2010). A paired samples *t* test showed no difference in participants' source memory scores for Test 1 (*M* = .58) versus Test 2 (*M* = .54), $t(20) = 2.03$, $p = .065$. Further, source memory scores were significantly different from chance levels of performance (50%) on both Test 1, $t(20) = 3.06$, $p = .003$, and Test 2, $t(20) = 1.77$, $p = .046$, suggesting that participants' source judgements were influenced by episodic memory and not simply by guessing.

In summary, participants' performance did not differ on Test 1 versus Test 2 across all memory measures, suggesting that the recognition tests that were constructed for Experiment 3 were not subject to interference effects. Further, after a one-week delay between encoding and retrieval, participants demonstrated typical hit and false alarm rates, as well as an ability to use episodic memory to discriminate between the two lists presented on Day 1. Therefore, the stimuli used in this pilot experiment were deemed appropriate for use in Experiment 3.

Experiment 3 Method

Design

The experiment employed a 2 (learning strategy: SP or RP) x 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 TSST-G stress induction. Non-stressed and stressed recognition performance was measured by pre- and post-stress tests.

Participants

An a priori power analysis was conducted using the same calculations as in Experiment 2b. This analysis recommended a sample size of 50 participants (*n* = 25 per group) to attain 99% power to detect the interaction reported in Experiment 1 (G*Power 3.0; Faul et al., 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 females, 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. 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.

Materials

Stimuli. The three wordlists described in the Experiment 3 pilot study served as stimuli.

Recognition tests. The two recognition tests are described in the Experiment 3 pilot. The only change to these tests was the addition of a confidence judgment that accompanied each list-discrimination question. 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.

Anxiety questionnaire. As in the previous experiments, the STICSA was administered to assess participants' self-reported levels of pre- and post-stress anxiety.

Procedure

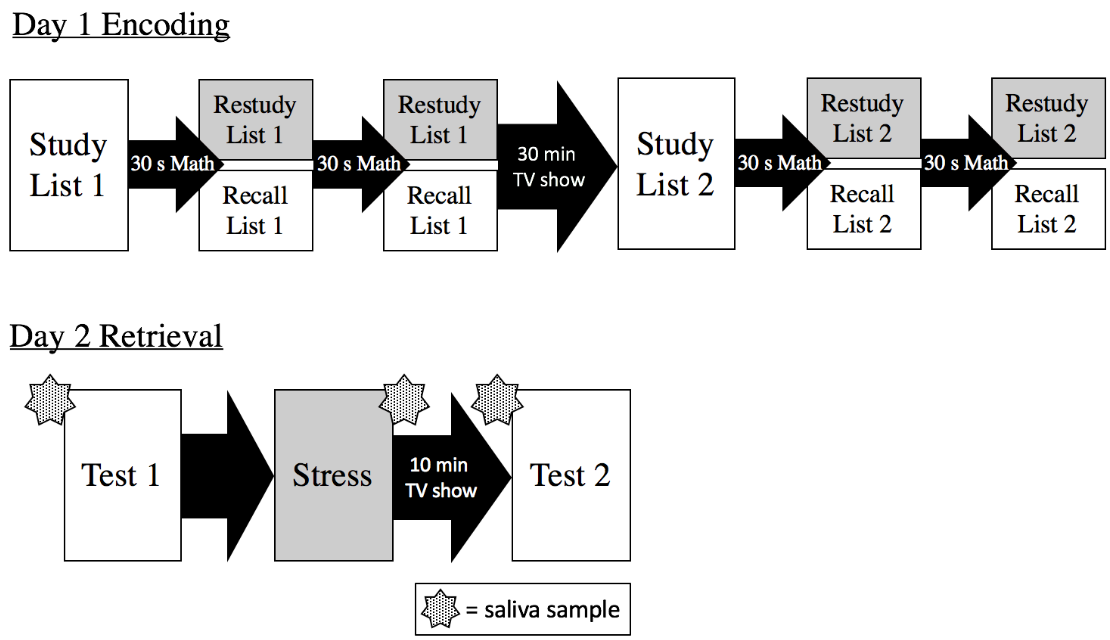
Testing sessions occurred on two days with a one week delay between Day 1 and Day 2 testing. All sessions took place in the afternoon. Participants were tested two at a time. In two cases in which only one participant showed up for the experiment, a research assistant served as a confederate.

For participants in the SP group, the Day 1 encoding procedure was exactly as described in the Experiment 3 pilot study. Participants in the RP group underwent the same task, with one change. Whereas those in the SP group studied each wordlist three consecutive times, individuals in the RP group studied each list once and then completed two consecutive free recall tests for the list. Prior to excusal on Day 1, all participants completed a STICSA. Please see Figure 7 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 TSST-G stress induction, consisting 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. During a subsequent 10 min retention interval, participants viewed part of an episode of *The Office*. Afterward, they provided the third saliva sample and completed Test 2. Participants were then debriefed and excused.

All encoding tasks, videos, and recognition tests were presented using E-Prime software (Version 2.1; Schneider et al., 2001). The Day 2 experimental procedure lasted approximately 45 min.

Figure 7. Graphic representation of the Experiment 3 procedure.



Cortisol Measurement and Data Management

Saliva samples for cortisol analysis were collected and handled as outlined in the Experiment 2a method. Cortisol data for two participants were excluded from analysis because values exceeded the normal range of cortisol concentrations (> 28 nmol/L; Laudat et al., 1988), indicating error in data processing. As in Experiments 2a and 2b, analyses on memory accuracy (i.e., hits, false alarms, and A scores) were repeated after eliminating cortisol non-responders. There were 13 non-responders (8 females) in the SP group and 11 non-responders (7 females) in the RP group, reducing the sample size in each of these groups to 17 and 21 respectively.

Experiment 3 Results

Self-reported Stress

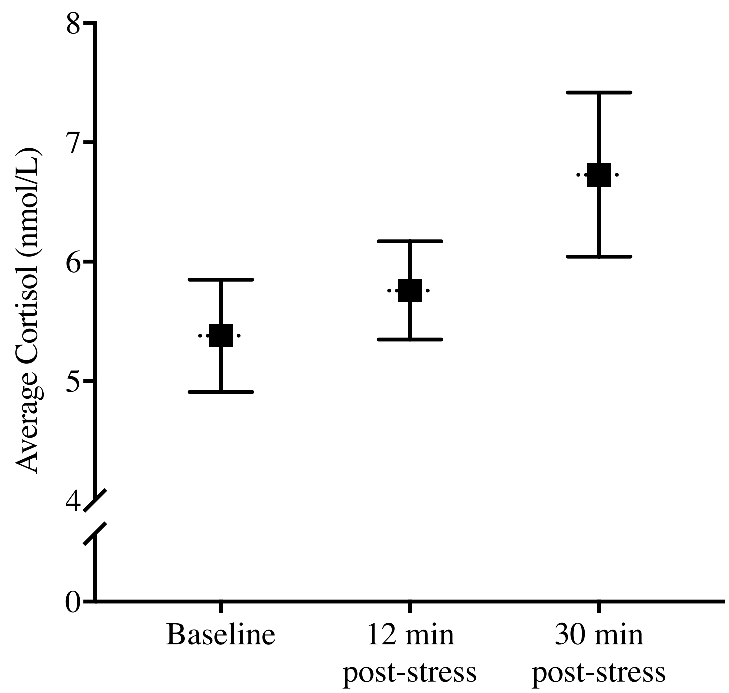
I first examined whether the Day 1 manipulation (SP or RP) affected participants' self-reported levels of stress. An independent samples *t* test on average Day 1 STICSA scores revealed no difference for participants who had engaged in study practice versus retrieval practice, $t(60) = 1.59$, $p = .118$.

To test whether the TSST-G tasks increased subjective anxiety on Day 2, I collapsed across learning group and conducted a paired-samples *t* test comparing pre- and post-stress STICSA scores. As expected, participants demonstrated heightened post-stress STICSA scores relative to baseline, $t(61) = 5.14$, $p < .001$, $d = 0.74$. The average pre-stress STICSA score was 29.00 (*SEM* = 0.86) and the average post-stress score was 32.24 (*SEM* = 1.03).

Cortisol

As in Experiment 2b, I collapsed across learning group and conducted a one-way repeated measures ANOVA comparing the baseline cortisol measurement to that taken 12 and 30 min post-stress. As shown in Figure 8, I found a main effect of timing, $F(2, 118) = 3.78$, $p = .026$, $\eta_p^2 = .06$. Pairwise comparisons using a Bonferroni correction revealed a marginally significant cortisol increase from 12 min post-stress to 30 min post-stress (mean difference = -0.10, *SEM* = 0.45, $p = .010$), but all other pairwise comparisons were non-significant. Thus, the cortisol response to stress was markedly lower than in Experiments 2a and 2b. Since the present experiment and Experiment 2b implemented the same stress-induction procedure, the discrepancy in cortisol reactivity is more likely due to unanticipated variability in the sample than to ineffective stress induction. For instance, 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 the previous experiments.

Figure 8. Experiment 3: 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 30 min after the onset of the TSST-G. Error bars represent *SEM*.



Day 1 Memory Performance

Table 4 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 4. Experiment 3: 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

Day 2 Recognition Performance

Please refer to Table 5 for means and standard errors for all of the dependent measures reported in the following analyses. As in Experiment 2b, gender was not included as a covariate in the follow analyses because of the low number of male participants (eight males in the SP group, nine males in the RP group).

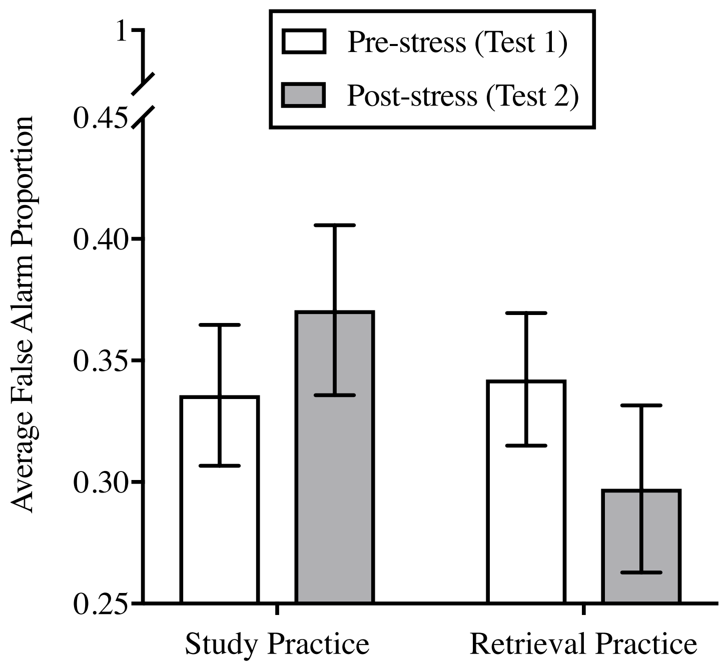
Hits. I first conducted a 2 (learning strategy: SP or RP) x 2 (test timing: pre-stress or post-stress) mixed model ANOVA on average hit proportions to determine whether accurate recognition of previously studied items differed according to learning strategy or conditions of stress. This analysis found a main effect of learning strategy, as participants in the SP group unexpectedly demonstrated higher hit proportions than those in the RP group (M_{SP} = .66 vs. M_{RP} = .57), $F(1, 60) = 6.75$, $p = .012$, $\eta_p^2 = .10$. No other main effects or interactions were significant (p's > .10). After eliminating cortisol non-responders, this analysis again found a main effect of learning strategy, $F(1, 34) = 5.91$, $p = .021$, $\eta_p^2 = .15$, but no other significant effects.

The benefits of retrieval practice over study practice on a final criterial test are contingent on successful retrieval during the 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. With that in mind, the relatively low hit rates I observed in the RP group could be due to low rates of recall during Day 1 retrieval practice. As shown in Table 4 above, Day 1 recall was low, as participants in the RP group recalled approximately 15 out of the 60 possible words during each retrieval attempt. For a more meaningful examination of the benefits of retrieval practice, I repeated the ANOVA on hit proportions after restricting recognition responses for those in the RP group 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. The ANOVA on restricted hits again found a main effect of learning strategy, $F(1, 60) = 80.34$, $p < .001$, $\eta_p^2 = .57$, but the effect was in the opposite direction as before. For items that they accurately recalled at least once during the Day 1 encoding session, individuals in the RP group demonstrated an average hit rate of .91, compared to the average hit rate of .66 in the SP group. No other main effects or interactions were significant (*p*'s > .10).

False alarms. I next conducted a 2 (learning strategy: SP or RP) x 2 (test timing: pre-stress or post-stress) mixed model ANOVA on average false-alarm proportions. This analysis found a significant learning strategy by test-timing interaction, $F(1, 60) = 4.10$, $p = .047$, $\eta_p^2 = .06$. As depicted in Figure 9, 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. No other main effects or interactions were significant.

Repeating this analysis after eliminating cortisol non-responders, I found a marginal learning strategy by test timing interaction that was consistent with the previous analysis, $F(1, 34) = 3.77$, $p = .061$, $\eta_p^2 = .10$. No other effects were significant.

Figure 9. Experiment 3: Average pre-stress and post-stress false alarm proportions for participants in the SP and RP groups. Error bars represent *SEM*.



Zhang and Mueller's A. A 2 (learning strategy: SP or RP) x 2 (test timing: pre-stress or post-stress) mixed model ANOVA on A scores found a significant main effect of test timing, $F(1, 60) = 4.11$, $p = .047$, $\eta_p^2 = .06$. Participants demonstrated higher pre-stress ($M = .60$) than post-stress ($M = .53$) A scores, suggesting that stress impaired participants' ability to discriminate between studied items and foils. A marginally significant main effect of learning strategy was also found, $F(1, 60) = 3.00$, $p = .090$, as participants in the RP group ($M = .51$) trended toward lower A scores than those in the SP group ($M = .63$). These main effects were even more pronounced when this analysis was repeated on cortisol responders. A main effect of test timing showed cortisol responders demonstrated the same pattern of higher pre-stress (*M* = .60) than post-stress (*M* = .48) A scores, $F(1, 34) = 4.99$, $p = .032$, $\eta_p^2 = .13$. A main effect of learning strategy showed that cortisol responders in the RP group (*M* = .47) had lower A scores than those in the SP group (*M* = .65), $F(1, 34) = 4.40$, $p = .044$, $\eta_p^2 = .12$.

Grier's *B"*. The pattern of low hits and low post-stress false alarms observed in the RP group suggest that these individuals may have adopted a bias toward "no" responses on the recognition test. To examine this potential response bias, I calculated the nonparametric statistic Grier's *B"* (Grier, 1971). Grier's *B"* ranges from -1 (bias in favor of a "yes" response) to 1 (bias in favor of a "no" response), with 0 signifying no response bias. A 2 (learning strategy: SP or RP) x 2 (test timing: pre-stress or post-stress) mixed model ANOVA on Grier's *B"* scores revealed a learning strategy by test timing interaction, $F(1, 60) = 4.47$, $p = .039$, $\eta_p^2 = .07$. As displayed in Table 5, prior to stress induction, individuals in the SP and RP groups demonstrated similar *B"* values. However, after stress induction, those in the RP group demonstrated higher *B"* values than those in the SP group. That is, individuals in the RP group were more biased toward providing "no" responses after stress induction than those in the SP group.

This analysis also found a marginally significant main effect of learning strategy, $F(1, 60) = 3.11$, $p = .083$, as those in the RP group trended toward more conservative responding (*M* = .12) than those in the SP group (*M* = .03). This finding was further supported by one-sample *t* tests comparing mean *B"* values to a fixed value of 0 (i.e., no response bias). Individuals in the SP group demonstrated *B"* values that were statistically similar to 0 on both the pre-stress, $t(29) = 0.76$, $p = .454$, and post-stress test, $t(29) = 0.08$, $p = .940$, suggesting that they were not biased toward "yes" or "no" responses. However, those in the RP group demonstrated *B"* values that were significantly higher than 0 on both the pre-stress, $t(31) = 2.50$, $p = .018$, $d = 0.44$, and post-stress tests, $t(31) = 3.68$, $p = .001$, $d = 0.65$, suggesting that those in the RP group were indeed biased toward "no" responses, particularly after stress induction.

List Discrimination

Source memory. As described in the experiment 3 pilot, source memory scores were calculated by dividing the total number of hits attributed to the correct source by the total number of hits for each participant. I first examined whether source memory scores were significantly different from chance levels of performance (50%) to ensure that participants' source judgements were influenced by episodic memory and not by simply guessing. Indeed, participants in both the SP and RP groups demonstrated above chance levels of discrimination on Test 1 (SP: $t(29) = 3.14$, $p = .004$; RP: $t(31) = 2.18$, $p = .037$) and Test 2 (SP: $t(29) = 2.78$, $p = .009$; RP: $t(31) = 2.50$, $p = .018$). However, the RP group demonstrated chance levels of performance on both Test 1, $t(31) = 0.76$, $p = .452$, and Test 2, $t(31) = 1.50$, $p = .144$, when items were restricted to those that had been recalled at least once during Day 1 retrieval practice.

I next conducted a 2 (learning strategy: SP or RP) x 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 found no main effects or an interaction (all p's > .10). After eliminating cortisol non-responders, this analysis again found no significant effects.

As with the analysis on hit proportions, I repeated this analysis after restricting responses for those in the RP group to items that had been accurately recalled at least once during their retrieval practice attempts on Day 1. Limiting responses in this manner provides a clearer answer to the question of whether retrieval practice, which involves successful recollection of information as a means of learning that information, improves memory access to contextual information. Despite restricting responses in this fashion, the ANOVA on source memory scores again found no significant effects (all *p*'s > .10).

Table 5. Experiment 3: Average pre-stress and post-stress performance on all memory measures and confidence measures. In addition to displaying means for the SP and RP groups, means are displayed for the $RP_{Restricted}$ group, referring to the restriction of hits to those that were accurately recalled at least once during retrieval practice on Day 1. Standard errors of the mean are given in parentheses.

	Memory Performance					Source Confidence	
Group	Hits	False Alarms	*A* Score	Grier's *B"*	Source Memory Score	Proportion of "High" Judgments	Gamma
Pre-stress							
SP	.66 (*.02*)	.34 (*.03*)	.66 (*.05*)	.04 (*.05*)	.57 (*.02*)	.36 (*.04*)	.20 (*.10*)
RP	.59 (*.02*)	.34 (*.03*)	.55 (*.06*)	.09 (*.03*)	.54 (*.02*)	.37 (*.03*)	.04 (*.07*)
$RP_{Restricted}$	.91 (*.02*)	---	---	---	.52 (*.03*)	.63 (*.04*)	.05 (*.07*)
Post-stress							
SP	.66 (*.03*)	.37 (*.04*)	.59 (*.06*)	.00 (*.06*)	.56 (*.02*)	.34 (*.04*)	.10 (*.09*)
RP	.56 (*.03*)	.30 (*.03*)	.47 (*.06*)	.16 (*.04*)	.55 (*.02*)	.36 (*.03*)	.26 (*.09*)
$RP_{Restricted}$	.91 (*.01*)	---	---	---	.55 (*.03*)	.63 (*.04*)	.26 (*.09*)

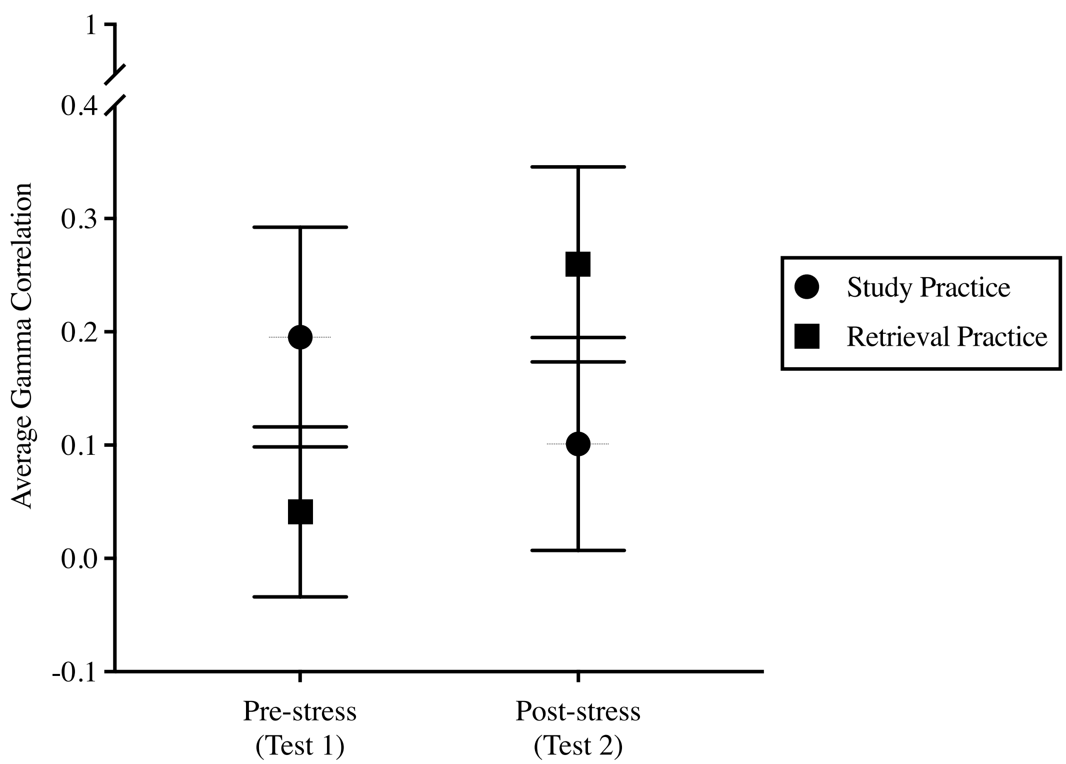
Confidence. Recall that participants gave a "high confidence" or "low confidence" rating for each list-discrimination judgment. These judgments were coded as binary, with 1 and 0 representing high and low confidence, respectively. To examine whether confidence varied as a function of learning strategy and test timing, I conducted 2 (learning strategy: SP or RP) x 2 (test timing: pre-stress or post-stress) ANOVAs on (1) the proportion of high confidence ratings each participant provided, and (2) 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. I 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.

The analysis on average proportions of high confidence ratings found neither significant main effects nor an interaction (all p's > .10). After restricting responses for participants in the RP group as was done for hit proportions and source memory, the 2 x 2 ANOVA found a main effect of learning strategy, $F(1, 60) = 28.35$, $p < .001$, $\eta_p^2 = .32$. Compared to participants in the SP group, those in the RP group were more confident in their list-discrimination judgments for items that they accurately recalled during retrieval practice on Day 1 ($M_{SP} = .35$ versus $M_{RP} = .63$).

As shown in Figure 10, the analysis on gamma correlations found a significant interaction between learning strategy and test timing, $F(1, 57) = 5.13$, $p = .027$, $\eta_p^2 = .08$. Individuals in the SP group did not demonstrate different pre-stress and post-stress gamma values, whereas those in the RP group demonstrated significantly lower pre-stress than post-stress gamma values. All other effects were non-significant. When this analysis was repeated after restricting responses for the RP group to items that had been previously recalled, I again found an interaction between learning strategy and test timing, $F(1, 57) = 4.98$, $p = .030$, $\eta_p^2 = .08$, which was characterized by the same pattern of gamma values.

Last, I 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, $t(31) = 2.98$, $p = .006$, $d = 0.53$, and the average post-stress gamma for the restricted RP group, $t(31) = 3.03$, $p = .005$, $d = 0.54$, significantly differed from 0. However, this was not true for any of the pre-stress gamma values (SP: $t(27) = 2.01$, $p = .055$; RP: $t(31) = 0.55$, $p = .585$; $RP_{Restricted}$: $t(31) = 0.68$, $p = .499$) or the post-stress gamma value for the SP group, $t(26) = 1.08$, $p = .292$. Thus, only the combination of stress with retrieval practice resulted in above-chance metacognitive accuracy.

Figure 10. Experiment 3: 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*.



Experiment 3 Discussion

In Experiment 3, I used a recognition test with accompanying list-discrimination and confidence judgments to examine whether stress would impair memory for the learning context (i.e., the list on which each item was presented), and whether retrieval practice would improve post-stress memory by increasing access to contextual details. Consistent with findings from Experiments 1 and 2b, I hypothesized that participants who engaged in retrieval practice at encoding would demonstrate better memory accuracy those who engaged in study practice on the final recognition test. I predicted that list discrimination would be most impaired for stressed individuals in the SP group, followed by non-stressed SP, stressed RP, and non-stressed RP, respectively.

The results from the yes/no recognition test in Experiment 3 were generally consistent with my predictions. First, consistent with Experiment 1, Experiment 2b, and previous research examining the effects of stress on recognition (Schwabe & Wolf, 2014), stress impaired retrieval. This was evidenced by lower post-stress than pre-stress A scores, indicating a reduced ability to discriminate between studied items and foils after stress was induced.

Second, the results suggest a positive influence of retrieval practice in the context of stress. Although initial inspection of the data showed that the RP group had lower hit rates than the SP group, this effect was reversed after restricting responses for those in the RP group to items that had been accurately recalled during their Day 1 retrieval practice attempts. Since individuals in the RP group only recalled approximately 25% of items during Day 1 retrieval practice, restricting their Day 2 recognition responses in this manner provided a more meaningful answer regarding whether successful retrieval practice improved recognition memory relative to study practice. Indeed, when retrieval practice was successful, participants demonstrated better memory accuracy on the recognition test. Further support for a benefit of retrieval practice over study practice in the context of stress was evident in participants' false-alarm proportions. Whereas the SP and RP groups had similar false-alarm rates on the pre-stress test, individuals in the RP group demonstrated lower false alarm rates on the post-stress test. Together, the false-alarm rates and restricted hit rates for the RP group suggest that successful retrieval practice improves recognition performance.

The patterns observed in the unrestricted hit rates and false-alarm rates for individuals in the RP group were indicative of a conservative response bias, therefore prompting the post-hoc analysis on Grier's *B"*. As suspected, participants in the RP group demonstrated a bias toward responding "no" to items on the recognition test, thus explaining their low hit rates relative to the SP group. Further, this bias was even more pronounced on the post-stress recognition test, as was also evident by their low post-stress false-alarm rates.

Because response bias has not previously been examined in the context of experiments involving stress or retrieval practice, the present results cannot be related to prior literature on the topic. However, a consideration of the dual-process model of memory may help explain the observed shifts in response bias. The dual-process view defines two memory processes: *recollection*, which is characterized by conscious retrieval of an item from memory, and *familiarity*, which is characterized by a feeling of having encountered an item before but an inability to recall it (Jacoby, 1991). For example, one may recall that the capital of Australia is Canberra (recollection), or one may feel that they know the capital of Australia but would need to see it on a list to remember it (familiarity). Whereas tests of free and cued recall rely more heavily on recollection, recognition tests capitalize on both memory processes. In the present study, the RP group had substantially less exposure than the SP group to the wordlists during

Day 1 encoding, which may have reduced their sense of familiarity with the items on the Day 2 recognition test. Further, because members of the RP group recalled so few items during Day 1 encoding, they likely did not benefit from the improvement in recollection that retrieval practice typically provides (Chan & McDermott, 2007). Thus, a lack of familiarity, coupled with a lack of recollection for the majority of the items on the recognition test, may explain why individuals in the RP group adopted a conservative response bias. Further, the presence of stress may have caused participants in the RP group to further doubt their memory for the items, causing them to be even more conservative.

Complementing the finding that stress and retrieval practice combined to reduce false alarm rates is the finding that the combination of stress and retrieval practice improved gamma correlations. Whereas individuals in the SP group showed no difference in their pre- and post-stress gamma scores, those in the RP group showed a substantial improvement from pre-stress to post-stress. Further, across all groups and conditions, the post-stress gamma value for the RP group was the only gamma value to differ from 0. These results indicate that the only experimental conditions that resulted in metacognitive accuracy on the list-discrimination task were the combination of retrieval practice on Day 1 and the presence of stress. This finding is consistent with the hypothesis that retrieval practice, combined with stress, encourages more-cautious, deliberate remembering.

Although the results from the yes/no recognition test and gamma analyses cast retrieval practice in a positive light, the results from the list-discrimination task and accompanying confidence ratings were not as encouraging. In contrast to my hypotheses, retrieval practice did not improve source memory on the list-discrimination task relative to study practice. Even more surprising, when items were restricted to those that had been previously recalled for the RP group, the RP group demonstrated even worse performance—their source memory scores were at chance levels. Despite their poor performance, the RP group gave a greater proportion of high-confidence judgments on these restricted items than the SP did across all items. An initial impression is that the pattern exhibited by the RP group stands in contrast to several studies showing that individuals are more confident in their memory for repeatedly studied items than for repeatedly tested items (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Kornell & Son, 2009; Roediger & Karpicke, 2006a). However, one must consider that these high-confidence judgments were made for items that had been accurately recalled during Day 1 retrieval practice. The act of successfully retrieving these items on Day 1 may have caused an unwarranted increase in their source-memory confidence. That is, because participants could easily and accurately remember the previously recalled items on the recognition test, they may have incorrectly assumed that they could also accurately remember the source of these items. These results thus put forth an interesting hypothesis: that metacognitive judgments for contextual, non-criterial information (e.g., source memory) are heavily biased by criterial recollection (e.g., item memory). This idea warrants further investigation, particularly given that such a phenomenon could have important real-world implications. For instance, an individual may accurately recognize a wanted suspect in a news announcement (i.e., item) and thus be highly confident that she saw the suspect at the grocery store (i.e., source) when in fact she saw the suspect at the bank.

Because stress did not influence performance on the list-discrimination task, the results of the present experiment do not provide support for the context-shift explanation for the effects of stress on retrieval. This hypothesis is based on the notion that pre-retrieval stress induces a mental context shift (Shields et al., 2017), thus impairing memory for contextual details associated with the information that was learned during encoding. By this account, the detrimental effects of stress on memory retrieval could be the result of inaccessibility to contextual information that is crucial for successful retrieval. In the present study, I expected to see a stress-induced context shift manifest as a reduced ability to remember the episodic context in which the items were learned (i.e., the temporally separated lists). The lack of an effect of stress on source memory may indicate that a stress-induced context shift is not the mechanism underlying stress effects on memory. However, the temporal context associated with a memory is just one of infinitely many contextual details that may accompany a memory. It remains possible that stress disrupts memory for other types of contextual details that aid in memory retrieval, and the present paradigm was not well suited for detecting such a disruption.

Performance on the list-discrimination task was also not differentially affected by the learning strategies that were manipulated during encoding, and thus the present results do not support the episodic-context account of retrieval practice effects. Briefly, this hypothesis dictates that retrieval attempts associate a given memory with new contextual information, including information about the time, mental state, and physical location in which the attempt was made (Karpicke et al., 2014). Retrieval practice thus creates contextually rich memories that are easier to recall than memories that have been subjected to restudying. In the present study, I expected that 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 RP group than the SP group. Though previous researchers using a similar list-discrimination paradigm found this to be true (Brewer et al., 2010), the RP group in the present study did not demonstrate superior source memory, even when examining list discrimination for items that participants had accurately retrieved during Day 1 encoding.

However, a key difference in methodology could explain this discrepancy: I used a one-week retention interval, whereas Brewer et al. (2010) administered the list-discrimination test immediately after encoding. Thus, the present results suggest that retrieval practice may improve memory for temporal context only in the short term. 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 a one-week retention interval (e.g., Roediger & Karpicke, 2006a), 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.

As a final note, in thinking about the potential real-world implications of this experiment, the low hit rates demonstrated by the RP group could potentially dissuade learners and educators from using retrieval practice as a learning tool. These results may be particularly disheartening considering that they are consistent with a growing consensus that the benefits of retrieval practice over study practice are less reliably demonstrated on recognition tests than on recall tests (see Karpicke et al., 2014). Some studies have demonstrated no advantage of retrieval practice over study practice on recognition tests (Darley & Murdock, 1971; Glover, 1989), and one study even reported results that are similar to the hit rates observed in the present experiment (Hogan & Kintsch, 1971). In this set of experiments by Hogan and Kintsch (1971), participants engaged in either four study trials or one study trial followed by three free recall attempts. After a 48-hr delay, those in the study-practice group demonstrated higher accuracy on a two-alternative forced-choice recognition test than those in the retrieval-practice group. Though this pattern of findings is discouraging, it speaks more broadly to an issue of the way in which retrieval practice is being employed at encoding. In reality, retrieval practice results in the best subsequent memory performance when it occurs in spaced intervals and is combined with interleaved re-exposures to stimuli (see Putnam, Sungkhasettee, & Roediger, 2016). Thus, the protocol adopted in the present study should not be taken at face value as the most efficacious retrieval-practice regimen.

CHAPTER 7: GENERAL DISCUSSION

In the present dissertation, I sought to determine how learning information through the use of retrieval practice reduces or eliminates the negative influence of psychological stress on memory retrieval. Experiment 1 established for the first time that retrieval practice can, indeed, bolster memory against stress. In Experiments 2a and 2b, I determined that multiple retrieval attempts during learning, as opposed to just one attempt, produced the best chances of remembering information under stress. Finally, in Experiment 3, I unexpectedly found that retrieval practice induced a bias toward rejecting items on a final recognition test, and that stress further amplified this conservative bias. The results of these four experiments share common themes, provide insight into current theories on both stress effects and retrieval practice effects, and pose important questions for future research in this area.

Common Themes Across the Present Experiments

Across all three experiments, several commonalities emerged. Of primary importance, all four experiments provided support for the notion that retrieval practice is a more effective learning strategy than study practice for improving post-stress memory accessibility. However, the benefits of retrieval practice over restudying became less and less pronounced with each subsequent experiment. Experiment 1 provided the most promising results, as stressed individuals who learned using study practice experienced the typical detrimental effects of stress on free recall, whereas stressed individuals in the retrieval-practice group demonstrated no post-stress memory impairment. In Experiments 2a and 2b, retrieval practice was again shown to be preferable to study practice in the context of stress, but only when a one-week delay separated encoding and retrieval. Further, in Experiment 2b, retrieval practice did not completely inoculate memory against stress, as had been observed in Experiment 1. Last, the results of Experiment 3 showed that the combination of stress and retrieval practice resulted in the lowest false alarm rates and the highest gamma values, although the benefits of retrieval practice did not materialize on the list-discrimination task. Thus, the results of the present dissertation suggest that retrieval practice helps negate the detrimental effects of stress on memory retrieval across a variety of stimuli and different retention intervals. However, there are clear limitations to retrieval practice, as discussed below.

Another important theme across these experiments was a relationship between the difficulty of the learning materials and the effectiveness of retrieval practice at helping participants learn the materials and avoid stress-related memory issues. Experiment 1 presented two simple sets of stimuli: 30 nouns of mixed valence, and 30 images of mixed valence. Not only are nouns and images easy to distinguish from each other, but including valenced items provides additional memorability. In Experiment 2b, in which the benefits of retrieval practice were less pronounced than in Experiment 1, participants were tasked with learning Swahili vocabulary words. Learning vocabulary in a foreign language requires the binding of a foreign cue with a familiar target—which inherently requires more effort than studying simple nouns and images in one's native language. Last, in Experiment 3, in which retrieval practice had the least obvious impact on post-stress memory accuracy, the stimuli consisted of two 60-item wordlists of neutrally-valenced nouns. These lists were twice as long as those used in Experiment 1, and required binding of item information with source information. Although researchers have made a strong case for the efficacy of retrieval practice with stimuli of varying complexities (Karpicke & Aue, 2015), the pattern of results observed in the present set of experiments suggests that, in the presence of stress, the effectiveness of retrieval practice may diminish as a function of stimulus difficulty. Further, this may be due to retrieval practice having a lesser effect when stimuli require the binding of information (Experiments 2b and 3) than when stimuli can be learned as individual units.

The results of the present experiments also highlight how the length of the retention interval between encoding and retrieval influences the interaction between retrieval practice and stress. The importance of the retention interval was most pronounced in Experiments 2a and 2b, in which a single retrieval attempt was found to be more than sufficient for promoting post-stress memory accessibility after 24 hours (Experiment 2a), but was substantially less effective than three retrieval attempts when a full week had passed since encoding (Experiment 2b). When compared to the findings of Brewer et al. (2010), the results of Experiment 3 similarly elucidate the influence of a longer retention interval. Brewer et al. (2010) used a list-discrimination paradigm that was similar to that used in Experiment 3, but tested source memory performance immediately after encoding. The results of the two studies conflict: Brewer et al. (2010) reported better source memory for individuals who had learned using retrieval practice versus study practice, whereas no group differences were found in Experiment 3. Altogether, these findings begin to suggest that retrieval practice may be most beneficial in the short term for bolstering memory against stress. Retrieval practice may also be effective only in the short term for purposes of increasing memory access to the temporal context associated with to-be-remembered items. The latter 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, the results of Experiment 3 present the possibility that memories may persist over long retention intervals but their associated contextual details may not.

As a final note, a common trend in physiological arousal also emerged across Experiments 2a, 2b, and 3, in which pre- and post-stress cortisol levels were measured. In all three experiments, the majority of cortisol non-responders were female participants. This pattern of findings can likely be explained by the fact that women demonstrate a blunted cortisol response to stress during particular phases of the menstrual cycle or when taking oral contraceptives (Kajantie & Phillips, 2006). Further, evidence is accumulating in support of the hypothesis that female hormones or hormonal contraceptives interact with the HPA-axis stress response to blunt the secretion of cortisol and consequently reduce the negative impact of stress on retrieval (Hidalgo et al., 2015; Shields et al., 2017). However, given that eliminating cortisol non-responders from analyses in this dissertation had little-to-no impact on the results, the present findings do not support this hypothesis.

Revisiting Findings and Theories in Stress Research

The results of the present experiments were generally consistent with broad trends across studies examining the influence of stress on memory retrieval. Post-stress memory performance has been shown to differ according to the type of final memory test administered, with detrimental effects of stress emerging on tests of free recall (Buchanan & Tranel, 2008; Buchanan, Tranel, & Adolphs, 2006; Hidalgo et al., 2015; Kuhlmann et al., 2005; Oei et al., 2006; Quesada, Wiemers, Schoofs, & Wolf, 2012; Schönfeld et al., 2014; Schwabe & Wolf, 2009; Smith et al., 2016) and cued recall (Lupien et al., 1997; Smeets et al., 2008; Tollenaar et al., 2008), but often not on recognition tests (Beckner et al., 2006; Buchanan & Tranel, 2008; Li et al., 2014; Pulopulos et al., 2013). Consistent with these trends, stressed individuals in the study-practice group demonstrated poor free-recall performance in Experiment 1, participants demonstrated impaired cued recall performance after stress was induced in Experiment 2b, and no main effects of stress were found for hit rates or false alarm rates on the recognition test in Experiment 3. With regard to the latter experiment, stress did, however, reduce participants' ability to discriminate between studied and non-studied items on the recognition test. This is consistent with findings from Schwabe and Wolf (2014), who used *d'* to measure sensitivity and found impaired performance for stressed individuals.

Experiments 1 and 2a also yielded findings that are consistent with studies in which researchers chose to administer multiple recall tests during initial learning, in order to promote memorability of stimuli after a 24-hour retention interval. Two experiments that used this approach reported null effects of stress on memory (Schoofs & Wolf, 2009; Wolf et al., 2002), potentially because their use of retrieval practice at encoding created stress-resistant memories. In the two experiments in this dissertation that similarly featured a 24-hr retention interval, retrieval practice of this nature yielded similar results; Experiment 1 revealed no memory impairment for stressed individuals who learned using retrieval practice, and Experiment 2a showed no memory impairment for stressed individuals who engaged in either one or three retrieval attempts during learning. Thus, the results of the present dissertation provide more evidence for the hypothesis that the null findings reported by Schoofs and Wolf (2009) and Wolf et al. (2002) were due to the memory-bolstering influence of retrieval practice.

In addition to contributing to the broad patterns in data that are observed in stress-and-memory studies, the results from the present experiments provide insight for current theories on the topic. First, an emerging theory about stress effects, and the theory that was tested in Experiment 3, is the context-shift theory. The idea that stress creates a mental context shift was proposed and supported by a recent meta-analysis examining how various memory processes are affected by stress (Shields et al., 2017). Based on the support for the context-shift hypothesis that was provided by the meta-analysis and prior literature (e.g., Schwabe & Wolf, 2009), I expected that stress would impair memory for both to-be-remembered items and the context in which they were learned (i.e., the lists they came from) in Experiment 3. The results did not support the context-shift account, but, as argued in the Experiment 3 discussion section, do not eliminate the possibility that there is some truth to the context-shift theory.

Another 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 et al., 2012). By this account, tests of free recall, which require conscious recollection and hippocampal support (Eichenbaum, Yonelinas, & Ranganath, 2007; Yonelinas et al., 2002), are expected to be impaired by stress. However, recognition tests, which can be completed with familiarity-based judgments and do not rely on the hippocampus (Eichenbaum et al., 2007; Yonelinas et al., 2002), should be less affected by stress. Though the results from the free recall (Experiment 1) and recognition (Experiment 3) tests used in the present research broadly support this theory, the finding that retrieval practice resulted in no stress-related changes in free recall performance (Experiment 1) presents a new challenge. Retrieval practice has been shown to improve hippocampus-dependent recollection, but not familiarity (Chan & McDermott, 2010). Based on this evidence, individuals in the retrieval-practice group may have been able to recruit the hippocampus to aid in the retrieval of items, even when they were stressed. Thus, the memory-formation-mode theory may need to be refined to explain conditions under which hippocampus-dependent episodic memory retrieval is and is not impaired by stress.

The findings from the present dissertation can also be related to the executive-resource hypothesis recently proposed by Gagnon and Wagner (2016). This hypothesis postulates that stress burdens the executive functions that are necessary for individuals to engage in careful, effortful recollection, thus explaining why tests of recall are more vulnerable to stress than tests of recognition (Gagnon & Wagner, 2016). Though no direct measures of executive function were included in the present studies, the detrimental effects of stress were indeed most apparent on the memory tests that inherently require more executive resources. However, as with the memory-formation-mode model, the executive-resource hypothesis may fall short in explaining why retrieval practice helped reduce post-stress memory impairment. If retrieval practice strengthens recollection (Chan & McDermott, 2007) and recollection requires executive resources (see Gagnon & Wagner, 2016), then stress should also impair memory performance when information was learned via retrieval practice. Though this logic was borne out in Experiment 2b, Experiments 1 and 2a were less supportive.

Importantly, the challenges that the present results pose to the memory-formation-mode model and executive-resource hypothesis are contingent on the finding that retrieval practice enhances recollection but not familiarity relative to study practice (Chan & McDermott, 2007). Because the Chan and McDermott (2007) experiments featured very short retention intervals (< 15 min), it remains to be determined whether retrieval practice and study practice differentially influence recollection and familiarity over longer retention intervals, such as the 24-hr and one-week intervals used in the present studies. If retrieval practice *does* increase familiarity relative to study practice after a long retention interval, the theories in their current form would account for the results of this dissertation. Specifically, if retrieval practice enhances familiarity in the long term, then it would follow that stressed individuals would be able to access these memories by engaging in automatic, less effortful retrieval via neural routes that bypass the hippocampus. Thus, both the executive-resource and memory-formation-mode hypotheses could account for the finding that memories resulting from retrieval practice are still accessible during a psychological stress response.

Revisiting Findings and Theories in Retrieval Practice Research

The testing effects demonstrated in the present experiments are consistent with those reported across the broader retrieval-practice literature. Testing effects are typically more pronounced on tests of free and cued recall than on recognition tests (Darley & Murdock, 1971; Glover, 1989; Karpicke et al., 2014), as was observed in Experiments 1, 2b, and 3, respectively. Additionally, the results of Experiment 3 mimicked those of Hogan and Kintsch (1971), who used a similar study/test paradigm and also found higher hit rates for individuals who were exposed to four study trials than for those who engaged in one study trial followed by three recall attempts.

Whereas the testing effects observed for recall and recognition were consistent with previous retrieval-practice research, the list-discrimination results from Experiment 3 were less supportive of the episodic-context account, which is the leading theory on the topic. This account proposes that retrieval practice increases access to contextual details that are associated with to-be-remembered information, and that these contextual cues help promote memory access to that information (Karpicke et al., 2014). Though Karpicke et al. (2014) presented a strong argument for this mechanism, Experiment 3 showed that retrieval practice did not improve memory for contextual details associated with to-be-remembered information. Specifically, retrieval practice did not improve participants' ability to remember which of the two temporally-segregated lists each recognized item originated from. This finding is considered in more detail in the Experiment 3 discussion section.

Though the results of Experiment 3 were not consistent with the episodic-context account, the findings from all experiments in this dissertation are broadly consistent with a *transfer-appropriate processing* (TAP) interpretation of retrieval-practice effects. In the context of learning and memory, TAP refers to the principle that memory performance is best when the mental processes that are necessary for completion of a memory test are the same processes that are engaged during initial learning (Morris, Bransford, & Franks, 1977). Researchers have hypothesized that retrieval practice is efficacious because it mimics the cognitive processes that are needed for retrieval (e.g., Roediger & Karpicke, 2006a; 2006b). Supporting this hypothesis, a recent meta-analysis found that testing effects were stronger when there was a match between practice and final test formats than when the formats differed (Adesope et al., 2017). The results of the present dissertation may lend further support to a TAP interpretation, since testing effects were more pronounced when the practice and final tests were both free recall (Experiment 1) or both cued recall (Experiment 2b) than when the practice tests involved free recall and the final test was a recognition test (Experiment 3). But what are the cognitive processes that are involved in creating the match between encoding and retrieval? These mental processes may relate to the influence of retrieval practice on recollection and familiarity. Because retrieval practice strengthens recollection but not familiarity (Chan & McDermott, 2007), the benefits of retrieval practice may be more apparent on final tests that rely on recollection (i.e., free recall, cued recall) than on those that can be completed based on a sense of familiarity (i.e., recognition tests). In other words, retrieval practice may be preferable to study practice only when individuals must engage in conscious, careful, and effortful remembering on a final test.

Future Directions

The results of the present studies point to several directions for future research. Here I will summarize the suggestions for future work that have been made throughout this dissertation. The results of Experiments 2b and 3 were particularly thought-provoking. In Experiment 2b, all participants, regardless of learning strategy, demonstrated post-stress memory impairment. These findings stood in contrast to those of Experiment 1, and I offered a few testable mechanisms that may account for this discrepancy. The first was that retrieval practice may be more efficacious for the learning of stimuli of low complexity (English nouns) than of higher complexity (Swahili-English word pairs), but that this limitation is evident only when these memories are subjected to stress. Future researchers may consider directly manipulating stimulus complexity in a stress-and-memory paradigm to test this hypothesis. Doing so would also contribute valuable evidence to one side of the ongoing debate regarding the efficacy of retrieval practice with complex materials (see Karpicke & Aue, 2015; van Gog & Sweller, 2015).

The second hypothesis to emerge from Experiment 2b was based on the premise that the cued recall test in Experiment 2b was more cognitively demanding than the free recall test administered in Experiment 1. Memory tests that require more effort may be more vulnerable to stress, such that even retrieval practice may not be able to inoculate memory against the deleterious effects of stressful circumstances. This hypothesis would benefit from intentional manipulations of cognitive demand in the context of paradigms similar to those used in the present studies.

The Experiment 3 findings also pose several questions for future research. First, the novel finding that retrieval practice induced a response bias that was further exacerbated by stress calls for more research examining the roles that stress and retrieval practice play in changing how individuals approach a memory test. This finding is particularly provocative, as it questions the ways in which researchers have conceptualized common patterns across studies examining the effects of stress on retrieval. In particular, the finding that stress tends to impair recall but not recognition has been interpreted as evidence that stress hormones interfere with brain regions involved in the former task but not the latter (see Gagnon & Wagner, 2016). However, this pattern of results could also be attributed to a stress-induced conservative response bias. If stress induces a response bias but does not necessarily interfere with memory accessibility, it would follow that tests in which responses can be withheld (i.e., free and cued recall) would be affected whereas forced-response tests (i.e., recognition) would not.

The failure of Experiment 3 to validate either the context-shift hypothesis of stress effects (Shields et al., 2017) or the episodic-context account of retrieval practice effects (Karpicke et al., 2014) also 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. Thus, the lack of support provided by the present research raises the question of whether the list-discrimination task employed in Experiment 3 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 the 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.

Experiment 3 also revealed the curious finding that, despite demonstrating chance-level source-memory scores, individuals in the retrieval-practice group reported higher confidence in their source judgments than those in the study-practice group. As hypothesized in the Experiment 3 discussion section, this may indicate that metacognitive judgments for contextual, non-criterial information (e.g., source memory) are heavily biased by criterial recollection (e.g., item memory). A direct test of this hypothesis would be both practically and theoretically important (see the Experiment 3 discussion section for more detail).

Last, several of the points raised in the general-discussion section refer to the finding that retrieval practice enhances recollection but not familiarity (Chan & McDermott, 2007). Though Chan and McDermott (2007) demonstrated this finding across three experiments and several dependent measures, additional research into the influence of retrieval practice on recollection and familiarity could help further strengthen their results and support the hypotheses that rely on those results.

Conclusions

Psychological stress has been shown to impair episodic memory retrieval (see Shields et al., 2017). However, the results of this dissertation present evidence that memory is not uniformly disrupted by stress. Rather, memories that have been reinforced through the use of retrieval practice may still be accessible in the presence of a physiological stress response. This improvement in memory accessibility may be most evident on tests involving free recall, followed by tests of cued recall and recognition, respectively. Further, retrieval practice may be most efficacious in the context of stress when the final test occurs after a shorter (24 hr) versus a longer (one week) retention interval. Though the present dissertation also aimed to determine the mechanism underlying the efficacy of retrieval practice at improving post-stress memory, the results did not support the hypothesized mechanism. Instead, the findings suggested an unexpected influence of retrieval practice and stress on response bias. These findings call for further research examining the theoretical mechanisms that have been proposed to explain effects due to testing and stress. In addition to helping refine current theory on the topic, a complete understanding of the interaction between stress and retrieval practice would help inform learners who wish to combat the detrimental effects of stress on memory.

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

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*(7), 861-876. https://doi.org/10.1002/acp.1391

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1063-1087. https://doi.org/10.1037/0278-7393.20.5.1063

Andreano, J. M., & Cahill, L. (2006). Glucocorticoid release and memory consolidation in men and women. *Psychological Science, 17*(6), 466-470. https://doi.org/10.1111/j.1467-9280.2006.01729.x

Beckner, V. E., Tucker, D. M., Delville, Y., & Mohr, D. C. (2006). Stress facilitates consolidation of verbal memory for a film but does not affect retrieval. *Behav Neurosci, 120*(3), 518-527. https://doi.org/10.1037/0735-7044.120.3.518

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. (2008). Stress and emotional memory retrieval: effects of sex and cortisol response. *Neurobiol Learn Mem, 89*(2), 134-141. https://doi.org/10.1016/j.nlm.2007.07.003

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*(3), 382-387. https://doi.org/10.1101/lm.206306

Buske-Kirschbaum, A., Jobst, S., Wustmans, A., Kirschbaum, C., Rauh, W., & Hellhammer, D. (1997). Attenuated free cortisol response to psychosocial stress in children with atopic dermatitis. *Psychosomatic Medicine, 59*(4), 419-426. https://doi.org/10.1097/00006842-199707000-00012

Christianson, S. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin, 112*, 284-309. https://doi.org/10.1037/0033-2909.112.2.284

Criss, A. H., & Shiffrin, R. M. (2004). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review, 111*(3), 800-807. https://doi.org/10.1037/0033-295X.111.3.800

Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology, 91*(1), 66-73. https://doi.org/10.1037/h0031836 de Kloet, E. R., Oitzl, M. S., & Joëls, M. (1999). Stress and cognition: Are corticosteroids good or bad guys? *Trends in Neurosciences, 22*(10), 422-426. https://doi.org/10.1016/S0166-2236(99)01438-1

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*(3), 355-391. https://doi.org/10.1037/0033-2909.130.3.355

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 17. https://doi.org/10.3389/fpsyg.2014.00781

Domes, G., Heinrichs, M., Rimmele, U., Reichwald, U., & Hautzinger, M. (2004). Acute stress impairs recognition for positive words--association with stress-induced cortisol secretion. *Stress: The International Journal on the Biology of Stress, 7*(3), 173-181. https://doi.org/10.1080/10253890412331273213

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58. https://doi.org/10.1177/1529100612453266

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30*, 123-152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Everly, G. S., Jr., & Lating, J. M. (2013). A clinical guide to the treatment of the human stress response (3rd ed.) Springer Science + Business Media, New York, NY. https://doi.org/10.1007/978-1-4614-5538-7

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

Foley, M. A., Johnson, M. K., & Raye, C. L. (1983). Age-related changes in confusion between memories for thoughts and memories for speech. *Child Development, 54*(1), 51-60. https://doi.org/10.2307/1129860

Fritz, C. O., Morris, P. E., Acton, M., Voelkel, A. R., & Etkind, R. (2007). Comparing and combining retrieval practice and the keyword mnemonic for foreign vocabulary learning. *Applied Cognitive Psychology, 21*, 499–526. https://doi.org/10.1002/acp.1287

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

Gärtner, M., Rohde-Liebenau, L., Grimm, S., & Bajbouj, M. (2014). Working memory-related frontal theta activity is decreased under acute stress. *Psychoneuroendocrinology, 43*, 105-113. https://doi.org/10.1016/j.psyneuen.2014.02.009

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*(3), 392-399. https://doi.org/10.1037/0022-0663.81.3.392

Gómez-Ariza, C. J., Lechuga, M. T., Pelegrina, S., & Bajo, M. T. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition, 33*(8), 1431-1441. https://doi.org/10.3758/BF03193376

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*(6), 424-429. https://doi.org/10.1037/h0031246

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

Hamarat, E., Thompson, D., Zabrucky, K. M., Steele, D., Matheny, K. B., & Aysan, F. (2001). Perceived stress and coping resource availability as predictors of life satisfaction in young, middle-aged, and older adults. *Experimental Aging Research, 27*(2), 181-196. https://doi.org/10.1080/036107301750074051

Herlitz, A., Nilsson, L., & Bäckman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25*(6), 801-811. https://doi.org/10.3758/BF03211324

Hidalgo, V., Pulopulos, M. M., Puig-Perez, S., Espin, L., Gomez-Amor, J., & Salvador, A. (2015). Acute stress affects free recall and recognition of pictures differently depending on age and sex. *Behav Brain Res, 292*, 393-402. https://doi.org/10.1016/j.bbr.2015.07.011

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning & Verbal Behavior, 10*(5), 562-567. https://doi.org/10.1016/S0022-5371(71)80029-4

Hupbach, A., & Fieman, R. (2012). Moderate stress enhances immediate and delayed retrieval of educationally relevant material in healthy young men. *Behavioral Neuroscience, 126*(6), 819. https://doi.org/10.1037/a0030489

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513-541. https://doi.org/10.1016/0749-596X(91)90025-F

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 112–127. https://doi.org/10.1037/0278-7393.34.1.112

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, *7*(1), 2-9. https://doi.org/10.7771/1932-6246.1167

JASP Team (2017). JASP (Version 0.8.2)[Computer software].

Johnson, M. K., Nolde, S. F., & De Leonardis, D. M. (1996). Emotional focus and source monitoring. *Journal of Memory and Language, 35*(2), 135-156. https://doi.org/10.1006/jmla.1996.0008

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., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*, 772–775. https://doi.org/10.1126/science.1199327

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), The Psychology of Learning and Motivation (vol. 61; pp. 237-284). Elsevier Academic Press, San Diego, CA. https://doi.org/10.1016/b978-0-12-800283-4.00007-1

Karpicke, J., & Roediger, H. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Kirschbaum, C., Pirke, K., & Hellhammer, D. H. (1993). The "trier social stress test": A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*(1-2), 76-81. https://doi.org/10.1159/000119004

Kornell, N., & Bjork, R. A. (2008). Optimising self-regulated study: The benefits--and costs--of dropping flashcards. *Memory, 16*(2), 125-136. https://doi.org/10.1080/09658210701763899

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*(2), 219-224. https://doi.org/10.3758/BF03194055

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17*(5), 493-501. https://doi.org/10.1080/09658210902832915

Kudielka, B. M., Buske-Kirschbaum, A., Hellhammer, D. H., & Kirschbaum, C. (2004). Differential heart rate reactivity and recovery after psychosocial stress (TSST) in healthy children, younger adults, and elderly adults: The impact of age and gender. *International Journal of Behavioral Medicine, 11*(2), 116-121. https://doi.org/10.1207/s15327558ijbm1102_8

Kuhlmann, S., Piel, M., & Wolf, O. T. (2005). Impaired memory retrieval after psychosocial stress in healthy young men. *J Neurosci, 25*(11), 2977-2982. https://doi.org/10.1523/JNEUROSCI.5139-04.2005

Laudat, M. H., Cerdas, S., Fournier, C., Guiban, D., Guilhaume, B., & Luton, J. P. (1988). Salivary cortisol measurement: A practical approach to assess pituitary-adrenal function. *Journal of Clinical Endocrinology and Metabolism, 66*(2), 343-360. https://doi.org/10.1210/jcem-66-2-343

Lewin, C., Wolgers, G., & Herlitz, A. (2001). Sex differences favoring women in verbal but not in visuospatial episodic memory. Neuropsychology, 15(2), 165-173. https://doi.org/10.1037/0894-4105.15.2.165

Li, S., Weerda, R., Milde, C., Wolf, O. T., & Thiel, C. M. (2014). Effects of acute psychosocial stress on neural activity to emotional and neutral faces in a face recognition memory paradigm. *Brain Imaging and Behavior, 8*(4), 598-610. https://doi.org/10.1007/s11682-013-9287-3

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

Lupien, S. J., Gaudreau, S., Tchiteya, B. M., Maheu, F., Sharma, S., Nair, N. P., Hauger, R. L., McEwen, B. S., & Meaney, M. J. (1997). Stress-induced declarative memory impairment in healthy elderly subjects: Relationship to cortisol reactivity. *The Journal of Clinical Endocrinology and Metabolism, 82*(7), 2070-2075. https://doi.org/10.1210/jc.82.7.2070

McCullough, A. M., & Yonelinas, A. P. (2013). Cold-pressor stress after learning enhances familiarity-based recognition memory in men. *Neurobiology of Learning and Memory, 106*, 11-17. https://doi.org/10.1016/j.nlm.2013.06.011

McGaugh, J. L., Cahill, L., & Roozendaal, B. (1996). Involvement of the amygdala in memory storage: Interaction with other brain systems. *Proceedings of the National Academy of Sciences of the United States of America, 93*(24), 13508-13514. https://doi.org/10.1073/pnas.93.24.13508

Merz, C. J., Wolf, O. T., & Hennig, J. (2010). Stress impairs retrieval of socially relevant information. *Behav Neurosci, 124*(2), 288-293. https://doi.org/10.1037/a0018942

Mizoguchi, K., Ikeda, R., Shoji, H., Tanaka, Y., Maruyama, W., & Tabira, T. (2009). Aging attenuates glucocorticoid negative feedback in rat brain. *Neuroscience, 159*(1), 259-270. https://doi.org/10.1016/j.neuroscience.2008.12.020

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16*(5), 519-533. https://doi.org/10.1016/S0022-5371(77)80016-9

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

Oei, N. Y., Everaerd, W. T., Elzinga, B. M., van Well, S., & Bermond, B. (2006). Psychosocial stress impairs working memory at high loads: an association with cortisol levels and memory retrieval. *Stress, 9*(3), 133-141. https://doi.org/10.1080/10253890600965773

Pulopulos, M. M., Almela, M., Hidalgo, V., Villada, C., Puig-Perez, S., & Salvador, A. (2013). Acute stress does not impair long-term memory retrieval in older people. *Neurobiol Learn Mem, 104*, 16-24. https://doi.org/10.1016/j.nlm.2013.04.010

Putnam, A. L., Sungkhasettee, V. W., & Roediger, H. L. III (2016). Optimizing learning in college: Tips from cognitive psychology. *Perspectives on Psychological Science, 11*(5), 652-660. https://doi.org/10.1177/1745691616645770

Qin, S., Hermans, E. J., van Marle, Hein, J. F., Luo, J., & Fernández, G. (2009). Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. *Biological Psychiatry, 66*(1), 25-32. https://doi.org/10.1016/j.biopsych.2009.03.006

Quesada, A. A., Wiemers, U. S., Schoofs, D., & Wolf, O. T. (2012). Psychosocial stress exposure impairs memory retrieval in children. *Psychoneuroendocrinology, 37*(1), 125-136. https://doi.org/10.1016/j.psyneuen.2011.05.013

Raven, J. H., Raven, P., & Chew, S. (2010). *The endocrine system* (2nd ed.). Edinburgh: Churchill Livingstone.

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General, 140*(3), 283-302. https://doi.org/10.1037/a0023956

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., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20-27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006a). 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

Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181-210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Ruiz, A. S., Peralta-Ramirez, M., Garcia-Rios, M., Mu-oz, M. A., Navarrete-Navarrete, N., & Blazquez-Ortiz, A. (2010). Adaptation of the trier social stress test to virtual reality: Psycho-physiological and neuroendocrine modulation. *Journal of Cybertherapy and Rehabilitation, 3*(4), 405-415.

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

Schönfeld, P., Ackermann, K., & Schwabe, L. (2014). Remembering under stress: different roles of autonomic arousal and glucocorticoids in memory retrieval. *Psychoneuroendocrinology, 39*, 249-256. https://doi.org/10.1016/j.psyneuen.2013.09.020

Schoofs, D., & Wolf, O. T. (2009). Stress and memory retrieval in women: no strong impairing effect during the luteal phase. *Behav Neurosci, 123*(3), 547-554. https://doi.org/10.1037/a0015625

Schwabe, L., Joels, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: an update and integration. *Neurosci Biobehav Rev, 36*(7), 1740-1749. https://doi.org/10.1016/j.neubiorev.2011.07.002

Schwabe, L., & Wolf, O. T. (2009). The context counts: congruent learning and testing environments prevent memory retrieval impairment following stress. *Cogn Affect Behav Neurosci, 9*(3), 229-236. https://doi.org/10.3758/CABN.9.3.229

Schwabe, L., & Wolf, O. T. (2014). Timing matters: temporal dynamics of stress effects on memory retrieval. *Cogn Affect Behav Neurosci, 14*(3), 1041-1048. https://doi.org/10.3758/s13415-014-0256-0

Schwabe, L., Romer, S., Richter, S., Dockendorf, S., Bilak, B., & Schachinger, H. (2009). Stress effects on declarative memory retrieval are blocked by a beta-adrenoceptor antagonist in humans. *Psychoneuroendocrinology, 34*(3), 446-454. https://doi.org/10.1016/j.psyneuen.2008.10.009

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

Skoluda, N., Strahler, J., Schlotz, W., Niederberger, L., Marques, S., Fischer, S., Thoma, M. V., Spoerri, C., Ehlert, U., & Nater, U. M. (2015). Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity. *Psychoneuroendocrinology, 51*, 227-236. https://doi.org/10.1016/j.psyneuen.2014.10.002

Slamecka, N. J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology, 76*(4), 504-513. https://doi.org/10.1037/h0025695

Smeets, T., Otgaar, H., Candel, I., & Wolf, O. T. (2008). True or false? memory is differentially affected by stress-induced cortisol elevations and sympathetic activity at consolidation and retrieval. *Psychoneuroendocrinology, 33*(10), 1378-1386. https://doi.org/10.1016/j.psyneuen.2008.07.009

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

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental

Psychology: Human Learning and Memory,6(2), 174-215. https://doi.org/10.1037/0278-7393.6.2.174

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

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory, 4*(3), 210-221. https://doi.org/10.1037/0278-7393.4.3.210

Tollenaar, M. S., Elzinga, B. M., Spinhoven, P., & Everaerd, W. A. (2008). The effects of cortisol increase on long-term memory retrieval during and after acute psychosocial stress. *Acta Psychol (Amst), 127*(3), 542-552. https://doi.org/10.1016/j.actpsy.2007.10.007

Uhart, M., Chong, R. Y., Oswald, L., Lin, P., & Wand, G. S. (2006). Gender differences in hypothalamic-pituitary-adrenal (HPA) axis reactivity. *Psychoneuroendocrinology, 31*(5), 642-652. https://doi.org/10.1016/j.psyneuen.2006.02.003

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

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*(6), 571-580. https://doi.org/10.1080/09658210244000414

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

Williams, C. L., & Clayton, E. C. (2001). Contribution of brainstem structures in modulating memory storage processes. In P. E. Gold, & W. T. Greenough (Eds.), Memory consolidation: Essays in honor of James L. McGaugh (pp. 141-163). American Psychological Association, Washington, DC. https://doi.org/10.1037/10413-008

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2012). How and when do students use flashcards? *Memory, 20*(6), 568-579. https://doi.org/10.1080/09658211.2012.687052

Wolf, O. T., Schommer, N. C., Hellhammer, D. H., Reischies, F. M., & Kirschbaum, C. (2002). Moderate psychosocial stress appears not to impair recall of words learned 4 weeks prior to stress exposure. *Stress, 5*(1), 59-64. https://doi.org/10.1080/102538902900012332

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

Yonelinas, A. P., Kroll, N. E. A., Quamme, J. R., Lazzara, M. M., Sauvé, M., Widaman, K. F., & Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience, 5*(11), 1236-1241. https://doi.org/10.1038/nn961

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

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika, 70*(1), 1-10. https://doi.org/10.1007/s11336-003-1119-8

Zoladz, P. R., Kalchik, A. E., Hoffman, M. M., Aufdenkampe, R. L., Burke, H. M., Woelke, S. A., . . . Talbot, J. N. (2014). Brief, pre-retrieval stress differentially influences long-term memory depending on sex and corticosteroid response. *Brain and Cognition, 85*, 277-285. https://doi.org/10.1016/j.bandc.2014.01.010