**Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

__________________________ ____________________

Andrew Cawley-Bennett Date

The Role of Memory Retrieval in Retrieval Practice.

By

Andrew T.J. Cawley-Bennett
Doctor of Philosophy

Graduate Division of Psychology

Joseph Manns
Advisor

Patricia Bauer
Committee Member

Daniel Dilks
Committee Member

Stephan Hamann
Committee Member

Robert Hampton
Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D., MPH.
Dean of James T. Laney School of Graduate Studies

Date

The Role of Memory Retrieval in Retrieval Practice.

By

Andrew T.J. Cawley-Bennett
M.A., Emory University, 2020

Advisor: Joseph R. Manns, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2023

Abstract

The Role of Memory Retrieval in Retrieval Practice.
By Andrew T.J. Cawley-Bennett

Retrieval practice involves intervening memory tests to enhance retention of information on a later test. The memory benefit from practice tests compared to restudying information is known as the retrieval practice effect. Retrieval practice has been studied for over a century, and retrieval practice effects are robust. However, the precise role of recall in retrieval practice remains uncertain. Relative to a restudy control condition, memory testing during retrieval practice involves recall but can also differentially engage participants' attention or depth of processing with the test stimuli. The possibility of increased participant involvement while testing brings up a potential issue, suggesting that the effects of retrieval practice might not solely be due to memory retrieval. In this dissertation, I compared a cued-recall retrieval practice condition, using word pairs (e.g., APPLE-WAGON; APPLE-?????), to a restudy control condition (e.g., APPLE-WAGON; APPLE-WAGON). The final memory test involved an old/new recognition memory paradigm, testing either the cue word (APPLE) or the target word (WAGON) from each word pair. The rationale of testing individual words was to differentiate the general memory testing influences on cue words from the specific contribution of memory recall for target words, relative to restudy control words. Across two behavioral experiments and one fMRI experiment, the results consistently showed a retrieval practice effect for both cue and target words. However, signal detection theory-based analyses revealed distinct memory effects for cue and target words. Retrieval practice cue words exhibited a modest yet consistent memory benefit from testing, while retrieval practice target words showed a larger and more variable benefit. The fMRI findings indicated more memory-related activity for retrieval practice target words compared to retrieval practice cue words in several brain regions, including the prefrontal cortex, medial temporal lobe, and medial occipital lobe. These results suggest that similar memory performance improvements for retrieval practice cue and target words are supported by different neural processes. These findings indicate that retrieval practice effects are not solely reliant on memory retrieval but are significantly influenced by it during retrieval practice.

The Role of Memory Retrieval in Retrieval Practice.

By

Andrew T.J. Cawley-Bennett
M.A., Emory University, 2020

Advisor: Joseph R. Manns, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2023

# Table of Contents

CHAPTER I. GENERAL INTRODUCTION .... 1
Overview .... 2
Brief Historical Background and Additional Terminology .... 4
Background on Methods and Experimental Designs Formats .... 5
Stimuli .... 6
Delays .... 7
Between-subject vs within-subject designs .... 7
Studies Examining Testing Effects for Non-recalled Information During an Intervention .... 8
Retrieval practice fMRI investigations .... 13
Hypothesized Mechanisms for Retrieval Practice Effects .... 18
Current Aims .... 21
CHAPTER II. Retrieval is Unnecessary to Observe a Retrieval Practice Effect .... 23
Abstract .... 24
Retrieval is Unnecessary to Observe a Retrieval Practice Effect .... 26
Experiment 1 .... 31
Materials and Methods .... 31
Participants .... 31
Stimuli .... 31
Procedure .... 32
Analyses .... 34
Results .... 37
Discussion .... 40
Experiment 2 .... 43
Materials and Methods .... 43
Participants .... 43
Stimuli, Procedure, and Analyses .... 43
Results .... 45
Discussion .... 50
General Discussion .... 51
References .... 58
Figure 1 .... 64
Figure 2 .... 65
Figure 3 .... 66
Figure 4 .... 67
Figure 5 .... 68
Figure 6 .... 69
Figure 7 .... 70
Figure 8 .... 71
Figure 9 .... 72
Figure 10 .... 73
Figure 11 .... 74
Figure 12 .... 75
Supplemental Material .... 76
Supplementary Text .... 76

Results from Experiment 1 ........ 76
Results from Experiment 2 ........ 77
Figure S1 ........ 78
Figure S2 ........ 79

**CHAPTER III. Functional MRI Correlates of Recall-Based Retrieval Practice Effects..... 80**

**Abstract ........ 81**

**Functional MRI Correlates of Recall-Based Retrieval Practice Effects ........ 82**

**Materials and Methods ........ 88**
Participants ........ 89
Stimuli ........ 89
Procedure ........ 90
Behavioral Data Analyses ........ 93
Neuroimaging Acquisition ........ 96
Anatomical Data Preprocessing ........ 97
Functional Data Preprocessing ........ 98
fMRI Data Analyses ........ 100

**Results ........ 101**
Behavioral Results ........ 101
fMRI Results ........ 104
All Trials ........ 104
Correctly Answered Final Recognition Memory Test Trials Only ........ 109

**Discussion ........ 112**

**References ........ 121**
Figure 1 ........ 130
Figure 2 ........ 131
Figure 3 ........ 132
Figure 4 ........ 133
Figure 5 ........ 134
Figure 6 ........ 135
Figure 7 ........ 136
Figure 8 ........ 137
Figure 9 ........ 138
Figure 10 ........ 139
Figure 11 ........ 140

**CHAPTER IV. GENERAL DISCUSSION ........ 141**
Depth of Engagement Improving Memory ........ 144
Parallels Between the Retrieval Practice Effect and the Generation Effect ........ 145
Additional Factors Improving Memory for Retrieval Practice Target Words ........ 146
Connecting our Findings with Prior Retrieval Practice Hypotheses ........ 147
Revising Old Hypotheses Based on Current Findings ........ 149
Retrieval Practice Effects on a Molecular Basis ........ 151
Future Explorations of Memory Strength ........ 152
Conclusion ........ 153

**Chapter I AND IV. References ........ 154**

**Appendix A ........ 162**

# CHAPTER I. GENERAL INTRODUCTION

**Overview**

The focus of this dissertation is *retrieval practice*, an experimental intervention in which one is asked to retrieve a memory on an interim test with the intent to improve performance on a later final memory test (Roediger & Butler, 2011). For example, if one were provided the word pair “APPLE – WAGON” and moments later were asked to complete “APPLE - ????” from memory, one would be more likely to remember “WAGON” on a subsequent final test (retrieval practice condition: “APPLE - WAGON → APPLE - ???? → WAGON?”) than if the word pair were studied twice prior to the final test (restudy control condition: “APPLE - WAGON → APPLE - WAGON → WAGON?”). This memory benefit from retrieval practice relative to that achieved from the restudy condition is large. Moreover, this effect is robust, reliably reproduced, and commonly referred to as the *retrieval practice effect* (Karpicke, 2017).

A main goal of the present dissertation is to assess the extent to which the retrieval practice effect depends on recalling information during an intervention test. If participants were to achieve better performance on the final memory test for the cue word (e.g., “APPLE”) relative to those in the restudy control condition, the result would indicate that the retrieval practice effect does not depend on recalling the to-be-recalled information during the intervention. This possibility has not been fully tested, despite hundreds of studies on retrieval practice. Of course, tests result in increased engagement with the study information, which correspondingly can affect attention, arousal, or vigilance. The presence of increased participant engagement during testing introduces a potential confounding variable, meaning that the effects of retrieval practice may not be solely attributable to memory retrieval. In the given example, participants are required to recall the target word "WAGON" while the cue word "APPLE" is presented on the screen during the retrieval practice intervention. In the restudy control condition both the cue and target words are also presented on the screen. Therefore, if memory for cue words such as "APPLE" is better from retrieval practice relative to restudy control cue words, it suggests that the retrieval practice effect may be influenced, at least partially, by increased engagement with the stimuli rather than memory factors solely relying on recall alone.

The possibility that deep engagement with stimuli can benefit subsequent retention, even in the absence of memory retrieval, resonates with earlier theories on "levels of processing" (Craik & Lockhart, 1972; Craik & Tulving, 1975). According to these theories, the depth of memory formation can be attributed to factors associated with the extent of encoding and thus determines the subsequent retention of the stimuli. In the context of retrieval practice, it is possible that the practice of retrieving information not only facilitates memory retention, but in so doing also triggers deeper levels of processing for the stimuli (see Buchin & Mulligan, 2017; Harlen & Crick, 2003; Lozito & Mulligan, 2006; Mulligan & Picklesimer, 2016; Nielson & Powless, 2007 for other examples on encoding factors that can impact subsequent retention). By considering the interplay between depth of engagement and recall, we can gain a more nuanced understanding of the mechanisms underlying retrieval practice effects and how they may differentially impact retention of cue versus target words.

A secondary goal of this dissertation is to connect the findings from these studies to existing hypotheses that account for the mechanisms supporting retrieval practice effects. Many hypotheses have been proposed and the present dissertation seeks to contribute to this debate by considering the possibility that different processes contribute to retrieval practice effects for target words (e.g., "WAGON") and cue words (e.g., "APPLE"). To address this concern, we will assess final memory performance separately for cue words and target words to ask: 1) do participants demonstrate a retrieval practice effect for cue words, and, if so, 2) do different underlying factors contribute to the effect for cue words versus target words as determined by participants' performance patterns on the final memory test?

Across two behavioral experiments (Chapter 2) and one functional MRI experiment (Chapter 3), this dissertation addressed both goals. In all three experiments, participants studied word pairs, then received a retrieval practice and a restudy control intervention (with different word pairs appearing in each condition). One or two days later, participants then completed a final recognition memory test wherein either the cue or target word was presented from each word pair. The following introduction will first provide some background on findings from previous retrieval practice studies demonstrating its robustness. Then a brief review of a few studies that investigated testing effects for non-recalled information during the intervention will be provided. I will discuss the differences between the motivations of those studies relative to the motivations of the current dissertation or the methodological considerations that limit the strength of those findings. Subsequently, many of the influential hypotheses explaining the mechanisms supporting retrieval practice effects will then be reviewed. Finally, I will detail how all three experiments conducted in this dissertation answered both core questions.

## Brief Historical Background and Additional Terminology

Testing is typically used to measure retention of material taught in a classroom. Additionally, testing (a form of retrieval practice) can enhance encoding of information and increase retention of material (Brame & Biel, 2015; Halamish & Bjork, 2011; Roediger & Butler, 2011). While the benefits of testing have been studied since the early 20$^{th}$ century (Abbott, 1909), only in recent decades have studies incorporated better controlled methodological approaches when investigating retrieval practice (Roediger & Karpicke, 2006).

In modern studies evaluating the efficacy of retrieval practice, just after learning some set of information, researchers administer retrieval practice in an intervention and compare subsequent memory to an alternative re-exposure control condition, such as restudying the information. After a delay from the intervention, participants are then provided a final memory test to measure retention for all the information presented during initial study. Retrieval practice research has consistently demonstrated a strong pattern of improved memory associated with retrieval practice compared to the control condition (Rowland, 2014; Yang et al., 2021), commonly known as the retrieval practice effect or testing effect (Karpicke & Roediger, 2006; Malmberg et al., 2014).

Students often leverage the retrieval practice effect when learning a new language by studying with flashcards. In this method, individuals test their memory by attempting to recall the associated translational word, either using the English word or the foreign translation written on the opposite side as a cue. Using foreign translation word-pairs are an example of one type of stimuli used in retrieval practice studies. Additionally, flashcards represent one of a few formats of retrieval practice known as cued-recall retrieval practice. The following sections will briefly summarize some of the other various stimuli and formats used to investigate the efficacy of retrieval practice. Furthermore, the following will highlight the specific methods chosen in this dissertation to optimize the outcomes across the conducted experiments. By delving into the background of retrieval practice studies, readers will also gain a better understanding of the confounding factors that negatively affected previous similar studies investigating retrieval practice effects between cue and target words, thereby offering a more informed perspective on the flaws introduced in those studies' design.

## Background on Methods and Experimental Designs Formats

Retrieval practice research often utilizes one of three testing formats during the intervention or on the final memory test: 1) a recognition memory test, 2) a cued-recall test, or 3) a free-recall tests (see Whiffen & Karpicke, 2017 for additional but less common testing formats). In recognition tests, participants must select the correct response among two or more options, such as in a true or false test or a multiple-choice test. Cued-recall tests include partial information used as a memory cue while omitting other information, requiring individuals to retrieve the to-be-recalled target information. Free recall tests do not provide any specific memory cues and instead asks participants to produce from memory as much information from a studied set of stimuli as possible. Debate among which testing format leads to better memory retention remains unresolved. While some researchers argue that recognition memory tests can be as effective as cued-recall or free recall tests (e.g., Smith & Karpicke, 2014; Kang et al., 2007), evidence from a recent meta-analysis reported greater retrieval practice effect sizes on a final test from cued-recall or free-recall intervention formats (Rowland, 2014). To assess the extent to which retrieval practice effects rely on retrieving the to-be-recalled target word during the intervention for cue words, the experiments in the present dissertation will employ a cued-recall retrieval practice intervention. This format allows us to appropriately address many of the proposed processes underlying retrieval practice effects.

**Stimuli**

Retrieval practice studies often used various language-based stimuli. Such language-based material have included prose passages (e.g., Butler, 2010; Kang et al., 2007; Karpicke & Blunt, 2011; Roediger & Karpicke, 2006), factual knowledge (e.g., McDaniel et al., 2007), foreign language translations (e.g., Karpicke & Smith, 2012; Pyc & Rawson, 2009; 2010; 2012; Toppino & Cohen, 2009), and word pair associations (e.g., Carpenter et al., 2006; Carpenter, 2009; 2011; Hong et al., 2019; Lehman & Karpicke, 2016; Wing et al., 2013). Because factual knowledge and prose passages can contain potentially confounding factors affecting outcome performance, such as pre-existing knowledge effects, impaired reading comprehension, or taxing working memory, the experiments in this dissertation are better served by more simplified stimuli. For feasibility, the experiments in the current dissertation will use English noun word pairs.

**Delays**

Another variation commonly implemented in prior retrieval practice research reflects the variations in delay length between the intervention and the final memory test. These delay lengths vary from minutes (termed immediate tests; e.g., Carpenter & DeLosh, 2005; Hong et al., 2019; Lehman & Karpicke, 2016; Wheeler et al., 2003;), to days (e.g., Carpenter et al., 2006; Kang et al., 2007; Roediger & Karpicke, 2006), to a week or longer (e.g., Karpicke & Blunt, 2011; Karpicke & Roediger, 2007; Wheeler et al., 2003). Yet, retrieval practice effects are not always readily apparent on immediate final tests (e.g., Roediger & Karpicke, 2006; Toppino & Cohen, 2009; Wheeler et al., 2003). Moreover, larger effect sizes tend to be found with longer test delays relative to shorter test delays (e.g., Chan, 2010; Rowland, 2014). The stronger retrieval practice effects observed after longer test delays highlight the practical relevance of retrieval practice for long-term memory. Consequently, the experiments in this dissertation aimed to explore the real-world applications of retrieval practice effects on long-term memory, necessitating a minimum one-day delay period.

**Between-subject vs within-subject designs**

A key methodological consideration in prior retrieval practice research is the choice between using within-subject and between-subject experimental designs. Findings from one meta-analyses revealed greater retrieval practice effect sizes for between-subject experiments relative to within-subject experiments (Rowland, 2014; though see Yang et al., 2021). This may explain why a large number of studies might have used between-subject experimental designs. However, a limitation of using between-subject paradigms, as suggested by some researchers, is that participants assigned to the retrieval practice condition could be equipped with better learning or testing strategies during the intervention, potentially aiding them with stimulus-general advantages on the later final memory test (e.g., Cho et al., 2017; Lehman & Malmberg, 2013; Roediger & Karpicke, 2006). That is, between-subject designs leave open the possibility that retrieval practice effects could be more about literally practicing a particular memory strategy rather than stimulus-specific benefits obtained from reactivating representations of that stimulus. Accordingly, in this dissertation, a within-subject experimental paradigm was implemented to ensure that performance is not solely dependent on strategies benefitting a particular group of participants.

## Studies Examining Testing Effects for Non-recalled Information During an Intervention

The following section summarizes previous research that has examined the influence of interim memory testing of some stimuli on subsequent memory for other, associated stimuli. This prior literature can be partitioned into two paradigms that might initially seem similar but are rather different. One paradigm is similar to the specific retrieval practice procedure used in the present dissertation. It has focused on how interim memory tests can improve memory for the provided cue information that is present during the interim test—stimuli that appear on the interim test but do not require recall to benefit from the test. The second paradigm has investigated how interim memory tests can improve memory for stimuli that never appeared on those interim tests yet are related to the tested stimuli. Although, the focus of the present dissertation is the first paradigm, this summary will begin by reviewing the findings from the second paradigm, referred to as testing-induced facilitation effects (Chan et al., 2006). The motivation for summarizing testing-induced facilitation effects is to clarify how those effects differ fundamentally from the retrieval practice effects that are the focus of this dissertation.

The testing-induced facilitation effect studies examined how successful retrieval facilitated the memory for associated studied concepts with the tested stimuli. In these studies, researchers required participants to study either prose passages (Bauml & Tribl, 2022; Chan et al., 2006; Hinze & Wiley, 2011; LaPorte & Voss, 1975; Nungester & Duchastel, 1982), educational videos (Cranney et al., 2009), word pairs (e.g., Rowland & DeLosh, 2014), or noun word lists (Bauml & Tribl, 2022; Rowland & DeLosh, 2014). Participants then either received interim test questions, some form of a re-exposure control (e.g., restudy), or were not afforded with re-exposure to the material (e.g., baseline measure). After a varied delay length from the intervention, participants completed a final memory test including questions from the intervention as well as additional questions not provided during the intervention.

In one example study, Chan and colleagues (2006) provided reading passages about toucans and the big bang theory and then tested participants only about toucans during the interim testing phase. Later, on the final test, participants received the same test questions about toucans from the interim tests as well as additional new questions regarding both the toucan passage and the big bang theory passage, neither of which were presented during the intervention (Chan et al., 2006). The testing-induced facilitation effect was quantified as the extent to which participants performed better on the novel toucan questions as compared to the novel big bang theory questions, presumably reflecting the indirect benefit of being previously tested on the original toucan questions. Indeed, the study revealed a testing-induced facilitation effect with better memory found for the novel toucan questions relative to the novel big bang theory questions.

In another example study (Rowland & DeLosh, 2014), participants studied three separate word lists of semantically unrelated nouns. Immediately after studying each of those lists, participants were prompted to recall half of the words from those lists in an interim-testing intervention. To cue recall, the first letter of each word was presented and each list was carefully constructed so that no two words on a list began with the same first letter. On the final test, participants were asked to recall those same words from the interim test as well as the remaining words from the original word lists (see Experiment 2; Rowland & DeLosh, 2014). As a control, participants studied an additional three separate word lists, however no interim-testing was performed on these lists. On the final test, participants were also prompted to recall as many words from the control lists as they could. The testing-induced facilitation effect from this study revealed better recall for words that were not tested from the interim-test condition relative to words from the control condition.

Most of these studies demonstrated an interim-testing effect on the final test, and a testing-induced facilitation effect for information related to the tested material that was not initially tested. Some of the authors from these studies have suggested that these testing-induced facilitation effects are consistent with the concept of “spreading activation” (Collins & Loftus, 1974), the idea that successfully retrieving information will also activate strongly associated representations. A few studies, however, did not find a testing-induced facilitation effect (e.g., Hinze & Wiley, 2011; LaPorte & Voss, 1975; Nungester & Duchastel, 1982). The findings from these testing-induced facilitation effect studies hold important implications regarding the role of interim testing, but they do not directly address the primary goal of this dissertation. Rather than asking if interim testing effects depend on retrieval, these studies have presumed it does and have asked instead how successful retrieval can lead to improved downstream memory outcomes for other associated stimuli. In contrast, the primary goal of this dissertation is to assess the extent to which memory retrieval is necessary to observe a retrieval practice effect in the first place.

There have been two prior studies that used retrieval practice paradigms similar to the procedure used in the present dissertation (Carpenter, 2011; Carpenter et al., 2006). Despite the procedural similarity, neither experiment was designed to ask the same core question asked by the present dissertation—to what extent do retrieval practice effects actually depend on memory recall? Nevertheless, the results of these studies are worth describing, if only to clarify the knowledge gaps present in the existing literature.

In the first study (Carpenter et al., 2006), participants initially studied word pairs and then engaged in cued-recall retrieval practice for half the word pairs and restudied the remaining word pairs. Feedback was provided after each retrieval attempt. For the final memory test, conducted 18 to 48 hours later, participants were separated into two final test groups, with one group given a cued-recall test and the other a free-recall test. For both groups, retrieval practice effects were reported for both target words and also for cue words in both test formats, representing a potentially novel finding. However, several major limitations of the study made interpreting the results difficult. For instance, one limitation was that participants provided covert responses. Consequently, assessing the extent to which participants were successful in recalling the target words remained unclear, potentially diminishing the retrieval practice effects for those words. Another limitation was the provision of feedback after retrieval attempts. Feedback is considered a moderating factor that enhances the effects of retrieval practice after recall (Kang et al., 2007). Feedback can also impact metacognitive judgments (Butler et al., 2008), such as improving the level of confidence from one's response. For these reasons, feedback is considered an indirect effect of retrieval practice (Roediger et al., 2011). To study the direct effects of testing, a factor important in this dissertation, feedback interference should be avoided. One last limitation of the study was the absence of inferential statistics directly addressing the question of whether retrieval practice enhanced final test memory for cue words compared to the restudy control condition.

A second study (Carpenter, 2011) also found that cued recall retrieval practice improved memory for both cue and target words relative to a restudy control condition. After an immediate delay from the intervention, participants completed an old/new final recognition memory test. In old/new recognition memory tests, participants are to try and identify previously seen words from the word pairs as old and identify novel words that have never been seen as new. The immediate final test results revealed similar memory retention for cue and target words in both the retrieval practice and the restudy control conditions, though memory was found to be significantly greater for the retrieval practice condition (Experiment 1). However, one limitation with this study was the requirement for participants to remember a small number of words pairs (16) combined with a short delay length, on the order of minutes, between the intervention and final test. Using a small number of word pairs reduces the potential for decay or interference effects, especially when tested minutes after an intervention. It is often more practical to apply longer delay lengths to increase retrieval practice effect sizes (Rowland, 2014). A second limitation was the use of a between-subjects design, which can introduce confounds related to differences in participant engagement. Because only half the participants received a retrieval practice intervention, those participants may have been more engaged overall in the experiment. For example, those participants that received retrieval practice during the intervention could have rehearsed or ruminated on answers between or after intervention trials. Using a between-subject design can therefore potentially exacerbate the differences between the two conditions. Finally, participants provided a recognition response for both the cue and target words separately from each word pair. The problem with testing both words from a word pair is that it complicates the separability of memory performance for cue and target words. To address this concern, it is advisable for a recognition memory test to include either the cue or the target word exclusively, rather than both, to ensure a fair assessment of memory performance. Overall, the findings from these two studies (Carpenter, 2011; Carpenter et al., 2006) offer limited evidence pertaining to the central question of the present study regarding the influence of memory recall on retrieval practice effects.

## Retrieval practice fMRI investigations

While most retrieval practice research focuses on behavioral paradigms, a small set of research has included functional neuroimaging. The majority of these neuroimaging studies have primarily examined brain activity during the intervention phase, specifically comparing between a retrieval practice condition and a restudy control condition (e.g., Liu et al., 2014; Vanesst et al., 2012; Vestergren et al., 2014; Wiklund-Hornqvist et al., 2017; Wing et al., 2013). These studies have shown that retrieval practice during the intervention elicits distinct patterns of brain activity in posterior brain regions, including the lateral temporal and parietal cortical regions, when compared to restudying information (Liu et al., 2014; van den Broek et al., 2013; Vannest et al., 2012; Wing et al., 2013). Specifically, successfully recalled trials appear to enhance neural representations in these posterior regions, leading to the presumed stabilization of the memory traces for the retrieved information. Additional findings from these studies include increased engagement in the ventrolateral prefrontal cortex during retrieval practice relative to a restudy control during the intervention. This prefrontal region is thought to contribute to better controlled retrieval processes by suppressing competing information (van den Broek et al., 2013; Vannest et al., 2012; Wing et al., 2013). However, one study noted that higher activity in this prefrontal region may not always predict better learning (Vannest et al., 2012). Lastly, a few neuroimaging studies discovered greater activity in the anterior cingulate cortex and the anterior insula during the intervention for retrieval practice trials relative to restudy control trials (van den Broek et al., 2013; Vannest et al., 2012; Vestergren et al., 2014; Wing et al., 2013). The authors suggested these areas were associated with attentional processes, which may correlate with increased engagement during retrieval practice (for a comprehensive overview see van den Broek et al., 2016).

One limitation of scanning during the intervention is that it can exacerbate the task engagement differences between retrieval practice and restudy control conditions that are not specific to memory retrieval. For instance, restudying information is a less demanding cognitive process, whereas retrieval practice is an active cognitive process that typically requires overt responses. These distinctions in cognitive processes and overt responses would of course yield different neural activity. Even if overt responses were required in the restudy control condition, the results would still differ, such as with measures in response times. As a result, scanning during retrieval practice may hinder rather than facilitate the identification of retrieval-specific contributions to the retrieval practice effect on subsequent memory. To better understand how retrieval practice impacts memory relative to a restudy control condition, experiments should instead focus on the memory differences between conditions during the final test, when retrieval is required in both conditions. Rarely have neuroimaging studies investigated this approach (e.g., Eriksson et al., 2011; Keresztes et al., 2014; Wirebring et al., 2015), which are discussed next.

In one study (Eriksson et al., 2011), fMRI was used to examine the impact of retrieval practice from a cued-recall intervention on a final test. Native Swedish speakers learned Swedish-Swahili word pairs (e.g., "MASHUA-BOAT") and then underwent repeated study-test trials (e.g., "MASHUA-????"). The fMRI scan occurred the next day, but subsequent memory tests were also performed one week and five months later. On the following day, participants completed a cued-recall memory test. Participants viewed cue words (e.g., "MASHUA-????") and indicated their level of recall certainty for the to-be-recalled target word (i.e., recall, uncertain recall, or no recall) followed by a post-scan recall test as verification. To explore the neural correlates of successful memory formation from retrieval practice, the authors compared brain activity between intervention trials where participants less frequently recalled the correct target word versus trials where participants more frequently recalled the correct target word. On the one-day memory test, Eriksson and colleagues determined that more frequently recalled targets resulted in greater neural activity in the anterior cingulate cortex. More successful recalls also resulted in decreased activity in the mid-ventrolateral PFC and parietal cortex. One must consider when interpreting the results that a limitation of this study was its small sample size (n=12), which diminishes the statistical power of the results. An additional limitation was the inconsistent number of repeated test trials across participants. Four participants had only four repeated test trials per word pair while other participants had up to eight repeated test trials. This confound of repeated test trials could have resulted in biased distinctions in brain activity observed on the one-day memory test. Additionally, this study lacked a control condition, making it unclear if restudying the material would produce similar patterns of neural activity for memory.

A second fMRI study (Wirebring et al., 2015), conducted a similar study that also had native Swedish speakers learn Swedish-Swahili word pairs (e.g., “MASHUA-BOAT”) through multiple rounds of cued-recall retrieval practice (e.g., “MASHUA-????”). A final cued-recall memory test was administered one week later in an fMRI scanner. During the final test, participants were asked to mentally retrieve the missing target word when provided the cue word (e.g., “MASHUA-????”) and indicate their level of certainty for recalling the target word (i.e., knew it was correct, believed it was correct, could not recall). Participants then selected the second letter of the missing target word from four alternative options. The analysis focused on comparing brain activity between words that were remembered both during the initial intervention and during the final memory test versus words only remembered during the initial intervention. The researchers observed that there was higher activity in several brain regions during recall of words remembered at both instances compared to those only recalled in the initial intervention. These brain regions included the bilateral posterior parietal cortex, bilateral lateral temporal cortex, right medial temporal lobe (including the hippocampus), right middle occipital cortex, and bilateral frontal cortex. Like the previous study, this study lacked a suitable control condition, limiting the ability to distinguish memory performance between a retrieval practice and a restudy control condition.

In one last fMRI study (Keresztes et al., 2014), patterns of neural activity were also compared between a retrieval practice and a restudy control condition. However, the primary aim of this study was to compare successful memory performance from either condition on the final test to an *n*-back working memory task. In this working memory task, participants studied a sequence of stimuli and then were required to indicate whether the current stimulus was in a given prior $n^{th}$ position in the sequence. By comparing successful memory performance from either condition on the final test to the *n*-back working memory task the authors could examine if there were overlapping neural correlates associated with either of the two conditions and working memory. Native German speakers learned German-Swahili word pairs (e.g., "MASHUA-BOAT") followed by half the word pairs being provided in a cued-recall retrieval practice condition (e.g., "MASHUA-????") and the remaining half provided in a restudy control condition (e.g., "MASHUA-BOAT"). Corrective feedback was provided after each retrieval attempt. A final fMRI cued-recall test was conducted either immediately or one week later. During the final test, participants responded only when they knew the answer to the missing target word. Then the 2-back working memory task was administered in the fMRI. A post-scan test assessed recall verification. The behavioral results demonstrated a retrieval practice effect only in the one-week group. However, fMRI analyses at one week did not reveal any significant differences between conditions when correcting for multiple comparisons. Instead, a less stringent threshold was applied in the one-week group revealing greater brain activity differences for retrieval practice, primarily in the inferior frontal gyrus, medial frontal/anterior cingulate cortex, and occipital lobe. Overlapping patterns of neural activity was found in the retrieval practice condition and the working memory task at one week, whereas overlapping activity from both intervention conditions and the working memory task was found in the immediate test group. The study also contained limitations, including a small sample size per group (n=13) and confounding factors such as feedback during retrieval practice. An additional limitation was the difference in re-exposure times between conditions during the intervention (eight seconds for retrieval practice and five seconds for restudy control). These subtle differences in re-exposure times could result in subsequent memory disadvantages on the final test.

In sum, these fMRI studies investigating the effects of retrieval practice on a final memory test (Eriksson et al., 2011; Keresztes et al., 2014; Wirebring et al., 2015) provided limited insights into the memory differences in brain activity between a retrieval practice and a restudy control condition. Furthermore, none of these fMRI retrieval practice studies assessed the memory impact from recall during retrieval practice for cue words.

## Hypothesized Mechanisms for Retrieval Practice Effects

The retrieval practice literature contains multiple hypotheses for retrieval practice effects. For simplicity, we have grouped these into three hypothetical categories referred to as the effortful search, the elaborative association, and the episodic reactivation hypotheses. I will now individually review each of these hypotheses.

The effortful search hypothesis suggests that harder-to-retrieve information requires greater mental effort and such effort subsequently leads to better memory. Prior research on this hypothesis has often focused on observed behavioral effects rather than putative cognitive processes. For example, earlier studies examined how retrieving information strengthened the memory for that information supported by the extent of the effort involved during the retrieval process and further provided descriptions about how more difficult forms of retrieval should lead to greater memory improvements (Bjork & Bjork, 1992; Gardiner et al., 1973). Later research connected these ideas to retrieval practice and this memory benefit was later found to apply specifically to information that was successfully recalled during retrieval practice (Pyc & Rawson, 2009), an outcome consistent with the effortful search hypothesis. Other research found that the benefit from retrieval practice is consistently observed across longer test delays from a couple days to weeks after retrieval practice, whereas inconsistent results are more typically found on immediate tests (Halamish & Bjork, 2011). An interpretation of these results was that effortful, yet successful memory search delayed forgetting (Kornell et al., 2011; Pyc & Rawson, 2009). Most effort-based hypotheses lack specificity with respect to putative cognitive processes and do not offer a clear definition of mental effort, but the narrative descriptions highlight that the greater the search process, the better the memory benefit for only the successfully recalled target information.

The elaborative association hypothesis suggests that retrieval practice activates additional information relevant to the memory cue. This extra information subsequently leads to retrieval of the to-be-recalled target word, providing an improvement in later retrieval. A key distinction from the effortful search hypothesis is that the elaborative association hypothesis focuses on the experimenter-provided cues rather than the missing, to-be-recalled information. The elaborative association hypothesis builds on early studies that found that deeper "levels of processing" (Craik & Lockhart, 1972) lead to more semantic associations and thus better memory for to-be-recalled information. Semantic memory was later proposed to be made up of concepts, where the activation of a concept would activate other related concepts sharing semantic characteristics that were all contained within a much larger conceptual framework. This phenomenon is referred to as the spreading activation theory (Collins & Loftus, 1975). Later research connected these ideas to retrieval practice. For example, several studies found that memory cues during retrieval practice elicit an elaborate network of related semantic associations (Chan et al., 2006; Carpenter, 2009; 2011). These studies proposed that the generation of additional information provides more resources, or pathways, to subsequently retrieve the to-be-recalled information. In support of this idea, one retrieval practice study observed greater memory enhancement for word pairs that had weak semantic relatedness as compared to word pairs that had strong semantic relatedness (Carpenter, 2009). The author's interpretation was that the pre-existing link for strongly related words cut short the cue elaboration process. By this view, the elaboration process would indirectly benefit memory for the to-be-recalled target information but would improve memory to a greater degree for the association-triggering cue. Later research complicated the interpretation by suggesting that only one robust association generated during the elaborative process would be sufficient to support the retrieval practice effect (Carpenter & Yeung, 2017; Pyc & Rawson, 2012). However, the association indirectly linking the cue to the target information was proposed to be directional (i.e., cue → association → target). Thus, by this view, the benefit of retrieval practice focuses on the cue because encountering the target information alone on a final test would not offer the opportunity to capitalize on the enhanced directional associations.

The episodic reactivation hypothesis suggests that retrieval practice plays a role in reactivating the initial study episode, including the original stimuli and context. This hypothesis differs from the prior two in that successful reactivation of the entire study episode would benefit all information to a similar degree, including both the retrieval practice cue and the to-be-recalled target words. The episodic reactivation hypothesis encompasses ideas from autonoetic mental time travel (Tulving, 2005), in which humans are said to possess the ability to place oneself in the past to reinstate a memory in the present moment. However, as operationalized here, the hypothesis is agnostic with respect to the importance of autonoetic awareness. Indeed, the episodic reactivation hypothesis also includes ideas taken from mathematical models of episodic memory, particularly the Temporal Context Model (Howard & Kahana, 2002), which emphasizes reactivation of past item-context representations. Several studies have supported the idea that retrieval practice works by reinstating the original context, including the initial temporal order of the information (Karpicke et al., 2014; Rowland & DeLosh, 2014; Whiffen & Karpicke, 2017). The episodic reactivation hypothesis proposes that retrieval practice improves memory more so than restudying the material because reactivation of the original study episode during the intervention affords more opportunity to connect the original memory to the memory of the intervention—a synergy not available during restudy.

## Current Aims

Across three experiments, we assessed how much of the retrieval practice effect relies on recall and, simultaneously, examined the underlying contributing cognitive factors. The memory task across all three experiments consisted of three phases: a study phase, an intervention phase, and a final memory test phase. Briefly, participants initially studied word-pairs, then received a cued-recall retrieval practice or a restudy control intervention. One to two-days later, participants then received an old/new recognition memory test wherein only the cue or target word from each word-pair was tested. For example, if “APPLE - WAGON” was initially studied as a word-pair, then during the intervention participants either saw “APPLE - ????” or “APPLE - WAGON”. On the final test, participants were either presented with “APPPLE” or “WAGON” and were required to indicate if the word was old or new. Each old/new response on the final test was also provided with a confidence rating. By testing individual words on the final test, it was possible to separately assess the impact of retrieval practice on to-be-recalled target words and on cue words that were consistently displayed on the screen, like the cue words in the restudy control condition. An added benefit to implementing a word pair paradigm is the opportunity to test multiple retrieval practice hypotheses. The effortful retrieval, the elaborative association, and the episodic reactivation hypotheses will make distinct predictions regarding the memory benefits afforded to either retrieval practice target words, cue words, or both words respectively.

In Chapter 2, we assessed the extent to which recall during retrieval practice impacts subsequent memory. We included receiver operating characteristic (ROC) models to help disambiguate between the effects of memory strength and memory variability. We additionally conducted a second experiment in Chapter 2. The aim of the second experiment was to determine whether two distinct memory factors supported retrieval practice effects for cue words compared to target words, or if memory differences between cue and target words were attributable to a single set of enhancing factors that varied in terms of memory strength. We further attempted to connect the findings from Chapter 2 with the three categories of retrieval practice effect hypotheses. In Chapter 3, we extended the findings from Chapter 2 by applying functional MRI during the final test to determine whether the retrieval practice effects for cue and target words reflected similar or different neural processes. In Chapter 4, the general discussion reviews the findings from all three experiments and further connect the findings from the three experiments with the three categorized retrieval practice effect hypotheses.

# CHAPTER II. Retrieval is Unnecessary to Observe a Retrieval Practice Effect

Andrew T.J. Cawley-Bennett, J. Imani Bunn, and Joseph R. Manns

Department of Psychology, Emory University

## Abstract

Retrieval practice involves interim memory testing to improve subsequent retention on a final test. It has been studied for more than a century and found to be a reliable technique for enhancing memory retention relative to restudy. However, the precise role of retrieval itself remains unclear. Relative to a restudy control condition, memory testing during retrieval practice can uniquely involve memory factors involved with recall but can also differentially engage participants' attention or depth of processing of the test stimuli. The potential confound of increased participant engagement during testing means that retrieval practice effects may not be wholly attributable to memory retrieval. In the present word-pair memory study, a cued-recall retrieval practice condition (e.g., APPLE-WAGON; APPLE-?????) was compared to a restudy control condition (e.g., APPLE-WAGON; APPLE-WAGON). The final memory test was an old/new recognition memory test in which only one word from each word pair was tested, either the cue word (e.g., APPLE) or the target word (e.g., WAGON). The rationale was that, relative to control, memory for cue words would reflect general influences from memory testing, whereas memory for target words would additionally reflect the contribution of memory recall. Across two experiments, the results indicated that, relative to a restudy control condition, retrieval practice significantly improved final test memory for both target words and cue words, indicating that retrieval practice effects did not depend only on recalling words alone. However, additional analyses based on signal detection theory revealed underlying differences in retrieval practice effects for cue and target words. Specifically, cue words received a modest yet consistent benefit from memory testing, whereas target words received a larger yet more inconsistent benefit. The findings indicate that retrieval practice effects do not depend on, but are significantly influenced by, memory recall during a retrieval practice intervention.

*Keywords:* retrieval practice, practice testing, testing effect, recognition memory, signal detection theory, declarative memory

## Retrieval is Unnecessary to Observe a Retrieval Practice Effect

Practice tests have been shown to improve information retention relative to rereading the information a second time. For example, if participants were to initially learn the word pair "APPLE – WAGON", their memory would on average be improved more by testing themselves on the word pair (e.g., "APPLE - ????") as compared to rereading the word pair a second time (termed restudy). This form of interim testing is commonly referred to as retrieval practice (Rowland, 2014). For more than a century, researchers have explored the advantages of testing (Abott, 1909; Bauml & Tribl, 2022). However, only in recent decades have investigations embraced better controlled experimental designs, integrating a restudy control condition to equate stimulus exposure (Roediger & Karpicke, 2006). Many studies report stronger advantages of retrieval practice relative to a restudy control condition on a delayed final memory test, resulting in what is referred to as a retrieval practice effect (Rowland, 2014) or a testing effect (e.g., Malmberg et al., 2014). Retrieval practice effects have been observed across various stimuli (e.g., foreign language translations, prose passages, abstract symbols, and word lists; Rowland, 2014) and numerous delay intervals (e.g., minutes, days, and weeks; Roediger & Karpicke, 2006), demonstrating the generalizability of the effect and its robustness.

An unanswered question remains: does the retrieval practice effect depend on recalling the information alone? Retrieval practice may uniquely involve overt recall, but may also differentially engage participants' attention to or depth of processing of the test stimuli relative to when restudying/rereading materials. The potential confound of increased participant engagement during testing means that retrieval practice effects may not be wholly attributable to memory retrieval. In the example above, participants are asked to recall the target word "WAGON" from memory during the retrieval practice intervention, and therefore have only one exposure to the word. However, they have two exposures to the cue word “APPLE”, which appears on the screen during both learning and testing, as it does in the control condition. Accordingly, the presentation of a repeated cue word might trigger some memory for the original presentation of the cue word in both the experimental and control conditions. Therefore, if memory for cue words like “APPLE” benefit from retrieval practice, then the retrieval practice effect may be at least partly attributable to increased engagement with the stimuli rather than overt recall. Investigating this possibility further is the primary focus of the current study.

Two prior cued-recall word-pair studies have examined how retrieval practice impacts both target and cue words relative to a restudy control condition (Carpenter, 2011; Carpenter et al., 2006). Yet, methodological considerations limit the insights from the results from both studies. The first study (Carpenter et al., 2006) found that retrieval practice improved memory for both target words and cue words on a final test administered 18 to 48 hours later. One limitation of this study was that participants were not required to make overt responses during the cued-recall intervention and were provided the missing target word after 4 seconds on each trial. Accordingly, it was unclear how much effort participants devoted to recalling the target words or how successful they were at doing so, potentially attenuating retrieval practice effects for those words. Additionally, providing immediate feedback during retrieval practice has been critiqued for adding an indirect effect on memory beyond retrieval practice (Roediger et al., 2011) because it involves processes after retrieval (Kang et al., 2007), including metacognitive processes (Butler et al., 2008). Another limitation of the study was that final memory was assessed using procedures that made it difficult to distinguish memory for cue words from memory for target words. Specifically, for some participants, memory of cue words was assessed on the final test by providing what had been the target word and asking the participant to recall its associated cue word, thereby confounding cue word performance with memory for the target word and its association with the cue word. For other participants, memory for cue words was assessed with a free recall test in which the prompt was to recall only the cue words, again confounding cue word performance with memory for target words to the extent that well-remembered target words would aid recall of cue words and well-remembered target words would also be less likely to be erroneously included as intrusions. A final limitation of the study was that no inferential statistics were provided that directly addressed the specific question of whether retrieval practice improved final memory for cue words relative to the restudy control question.

A second study (Carpenter, 2011) also found that cued-recall retrieval practice improved memory for both target and cue words relative to a restudy control condition. In this study, no feedback was provided during the retrieval practice intervention, and final memory performance was assessed using separate old/new recognition memory tests for cue words and target words. Thus, the second study avoided some of the procedural limitations of the first. However, the second study included only a small number (16) of word pairs and used a between-subjects design in which only some of the participants engaged in retrieval practice. A between-subjects design can exacerbate confounds related to participant engagement between retrieval practice and restudy control conditions to the extent that participants in the retrieval practice condition would likely be more engaged overall in the experiment, for example, by rehearsing or ruminating on answers between or after intervention trials. Further, the delay in this second study between the intervention and final memory test was only 5 minutes. Some prior studies have found that shorter delays can result in better final memory for the restudy control condition relative to the retrieval practice condition (e.g., Karpicke & Roediger, 2006), and a general trend is that longer (e.g., 24-hour) delays tend to result in larger beneficial retrieval practice effects (Rowland, 2014). Finally, the second study assessed final recognition memory for cue and target words separately yet across all trials included both words from each pair on the final test, thereby complicating the separability of memory performance for cue and target words. In sum, these two studies (Carpenter, 2011; Carpenter et al., 2006) provided only limited evidence relevant to the current study's key question regarding the extent to which retrieval practice effects depend on recall.

Other studies have asked whether retrieval practice effects go beyond retrieving the missing target information, referred to as testing-induced facilitation effects (Chan et al., 2006). However, these studies have addressed questions that differ importantly from that of the current study. Rather than asking if retrieval practice effects depend on recall, these studies have presumed it does and have asked instead how successful recall can have beneficial downstream memory consequences for other related stimuli, such as content of prose passages (e.g., Chan et al., 2006; Chan, 2009; LaPorte & Voss, 1975; Nungester & Duchastel, 1982; Rowland & DeLosh, 2014), educational videos (e.g., Cranney et al., 2009), or noun word lists (e.g., Bauml & Tribl, 2022; Rowland & DeLosh, 2014; see also LaPorte & Voss, 1975 and Nungester & Duchastel, 1982 for counterexamples). For example, in one study participants either reread paragraphs about toucans or were given interim test questions about the birds before taking a final test with interim test questions and new questions (Chan et al., 2006). The interim testing condition led to better performance for these new questions relative to the restudy control condition. The authors' interpretation was consistent with the idea of "spreading activation" (Collins & Loftus, 1975), the notion that successfully retrieved information can activate representations of additional related information. In sum, these studies addressed an important topic but did not directly investigate the extent to which memory recall is necessary at all for retrieval practice effects.

The possibility that deep engagement with stimuli can benefit subsequent retention even in the absence of memory testing echoes earlier ideas that "levels of processing" primarily influence the amount of memory encoding and thereby subsequent retention (Craik & Lockhart, 1972; Craik & Tulving, 1975). Related to this view, retrieval practice might elicit deeper levels of processing of the stimuli in addition to retrieval itself (see Harlen & Crick, 2003; Lozito & Mulligan, 2006; Mulligan & Picklesimer, 2016; Nielson & Powless, 2007 for examples of deep encoding improving retention). If so, then retrieval practice effects for target words would be supported by the participant's depth of engagement and recall, whereas retrieval practice effects for cue words would only be supported by the participant's depth of engagement.

The current study consisted of two experiments that aimed to elucidate whether the retrieval practice effect was solely observed as a result from memory factors related to recall or whether the effect was also supported by the increased participant engagement associated with retrieval practice. We used a within-subject study design where participants were presented with noun word pairs and were then re-exposed to each word pair in either a retrieval practice condition or restudy control condition. After two days, we subsequently assessed memory for both cue and target words using an old/new recognition memory test. Analyses were bolstered by fitting receiver operating characteristic (ROC) curves to data to disambiguate effects on memory strength and memory variability. Results from both experiments indicated that retrieval practice significantly improved retention performance for both cue words and target words relative to a restudy control condition. However, results from ROC analyses indicated that the performance improvement for cue words was due to a modest yet consistent benefit to memory, whereas the performance improvement for target words was due to a larger yet more variable benefit to memory. The results are discussed in terms of implications for partitioning retrieval practice effects into factors related to levels of engagement and to memory retrieval.

## Experiment 1

### Materials and Methods

#### *Participants*

Sixty-seven undergraduate students were recruited by offering course credit for introductory psychology courses at a private university. An *a-priori* sample size of 52 participants was targeted for inclusion in the final data set to achieve a power of at least 0.8, Cohen's $d = 0.4$, and an alpha = .05. The estimated effect size between the retrieval practice and restudy control intervention conditions for cue and target words was based on pilot testing. Additional participants were recruited with the expectation that some data would not be included in the final analyses based on exclusion criterion (see Results for details). All procedures were approved by the Institutional Review Board at Emory University.

#### *Stimuli*

Pairs of five to seven letter nouns were compiled, constrained to moderate imageability (330-600), concreteness (400-600), and familiarity (300-550) based on ratings available in the MRC psycholinguistic database (https://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm; Coltheart, 1981). In Experiment 1, four separate randomly counterbalanced stimuli sets were created, and each participant received one set of stimuli. For each stimulus set, words were randomly paired together. Word pairs were then randomly assigned as "old" words (i.e., appearing in all phases of the study) or were assigned as "new" words (i.e., appearing only on the final memory test).

Additionally, all old word-pairs in each set were randomly assigned to either the retrieval practice condition or restudy control condition.

***Procedure***

Figure 1 provides a schematic of the procedure. All participants completed a study phase, an intervention phase, and a final recognition memory test phase. All phases were presented on a computer in a private research laboratory room using Qualtrics software (https://www.qualtrics.com; Provo, UT). Prior to beginning the experiment, participants provided written consent and were then given two examples of the study and intervention phases. Participants were informed that they would be tested on their memory for the word pairs presented to them.

In the study phase, participants first saw six word-pairs presented in a randomly-chosen order for six seconds each. Blocks of six trials were selected based on pilot testing to increase intervention performance in the retrieval practice condition. Participants were then immediately presented with the same six trials in a newly randomized order in the intervention phase, with each trial lasting 12 seconds. Half of the intervention trials were retrieval practice interventions, and half were restudy control trials. For retrieval practice trials, one word in a pair was provided as the cue and the other was replaced by an underlined blank. For restudy control trials, both words were presented. One word in the restudy control condition was underlined, mimicking the target from the retrieval practice trials, thus termed the control target words. The non-underlined words were termed control cue words. In both conditions, participants were asked to type in both words for each trial. The left versus right position of cue and target words was counterbalanced across trials for both retrieval practice and restudy control trials. No feedback was provided during the intervention based on prior literature highlighting that feedback may introduce influences on memory that are not directly attributable to retrieval (e.g., Kang et al., 2007). The six study and six corresponding intervention trials comprised one block of trials. Participants completed a total of 14 blocks. In total, participants saw 42 word-pairs in the retrieval practice condition and 42 word-pairs in the restudy control condition.

After two days (between 47–72 hours), participants returned to the laboratory and first completed a demographic questionnaire. Participants then completed an old/new recognition memory test, starting with seven test trial examples. Four of the trials included a word from the previously presented example word pairs, and the remaining trials included a new word. For each trial, a single word was presented, and participants were instructed to identify previously studied words as old and novel words as new, with a level of confidence. Thus, participants provided one of six response options: sure old, maybe old, guess old, guess new, maybe new, or sure new. In total, 168 test trials were presented (84 new words and 84 old words), with each trial lasting five seconds. Trials were partitioned into six 28-trial blocks with intervening self-paced rest periods. In each block, words were presented in a random order and were equally divided between old (14) and new (14) words. In each block, old words were equally divided between the retrieval practice (7) and restudy control (7) conditions. Across blocks, old words in each condition were equally likely to have previously served as either cue or target words.

At the end of each visit, participants additionally completed a free recall test with 10 novel words to assess participant engagement. Specifically, 10 new words taken from the same word pool as the test stimuli were presented simultaneously for 30 seconds. After answering three to five intervening survey questions, participants were asked to type in as many of the 10 words as they could remember within a 90 second period. Each free recall task used new words that did not appear elsewhere in the experiment.

***Analyses***

Analyses were conducted using MATLAB version 2022b (MathWorks; Natick, MA). The focus of the present study was the extent to which retrieval practice effects for cue words were similar to retrieval practice effects for target words. Accordingly, the three key analytic questions were 1) was there a significant retrieval practice effect for cue words, 2) was there a significant retrieval practice effect for target words, and 3) within the retrieval practice condition, was the memory for cue words and target words significantly different from one another? The analyses focused on three statistical comparisons: memory for retrieval practice cues versus restudy control cues, memory for retrieval practice targets versus restudy control targets, and memory for retrieval practice cues versus retrieval practice targets. These comparisons were evaluated with two-tailed paired t-tests, unless otherwise noted. Analysis of variance (ANOVA) was not used because neither the main effect of condition irrespective of word type nor the main effect of word type irrespective of condition would directly address the study's main questions. Moreover, a significant interaction term between condition and word type would be ambiguous as it could result, for example, from either a unique or disproportionate retrieval practice effect for targets relative to cues (i.e., disordinal or ordinal interaction terms).

Performance on the final recognition memory test was calculated as a d-prime score (denoted as d'; Tanner & Swets, 1954), a standard metric of memory that subtracts a normalized false alarm rate (proportion of new words judged as old) from a normalized hit rate (proportion of old words judged as old). Non-responses were considered incorrect guesses. For each participant, all new words contributed to the false alarm rate but separate hit rates were calculated for each word type from each condition. Therefore, separate d' scores were calculated for retrieval practice cues, restudy control cues, retrieval practice targets, and restudy control targets. Instances of perfect hit rates (1.0) or false alarm rates (0.0), which would yield infinite normalized values, were adjusted by subtracting (hits) or adding (false alarms) one-half trial from the numerator prior to normalization.

Additional analyses capitalized on fitting receiver operator characteristic (ROC) models to the data from final recognition memory test performance using the ROC Toolbox v.1.1.1 (Koen et al., 2017). Briefly, cumulative hit rates and false alarm rates were calculated across participants' confidence ratings, and curves were fitted to these data using maximum likelihood estimation. The number of trials per old word type (21 trials) was insufficient to fit ROC models to each participant's data, and thus ROC models were fit to group data by summing cumulative judgment values across participants. The ROC approach yielded estimates of memory performance that were theoretically similar to the d' scores described above. Signal Detection Theory conceptualizes memory as the discriminability of a memory signal distribution of old words from a noise distribution of new words (Pastore & Scheirer, 1974). We will refer to this discriminability as signal strength. An additional advantage of fitting ROC models is that the approach can also provide an estimate of how the shape (standard deviation) of the signal distribution differs from that of the noise distribution (Egan, 1958). Therefore, fitting ROC models to the data allows one to disambiguate whether increases in memory performance are due to relative increases in memory signal strength, decreases in signal standard deviation, or some combination of both. Accordingly, a commonly-used unequal variance signal detection model (UVSD; Mickes et al., 2007; Wixted, 2007) was fit to the ROC data to yield parameter estimates of both signal strength and the signal standard deviation (as a ratio of the standard deviation of the noise distribution).

A caveat to this ROC approach is that consensus does not exist regarding which model is best for recognition memory data. Accordingly, we also fit the data with a dual-process signal detection (DPSD) model that assumes two independent memory processes contribute to recognition memory judgments (Yonelinas, 1994). One process, termed familiarity, is conceptually similar to the estimate of graded memory signal strength described by signal detection theory. The other process, termed recollection, is depicted as an all-or-none threshold process that involves episodic-like mental time travel (Yonelinas, 1994). To be clear, the motivation of the ROC analyses here was not to arbitrate which model was best but instead to use ROC models in general to ask if the memory supported by retrieval practice differed between cue words and target words. Moreover, both the UVSD and DPSD ROC models allow for the possibility that multiple memory processes contribute to recognition memory judgments (Wixted, 2007). The theoretical distinction between these models is whether it assumes that multiple memory processes contribute independently to memory judgments (DPSD) or aggregate into a combined signal for the purposes of making memory judgments (UVSD).

For both UVSD and DPSD models, we used a random bootstrap procedure to estimate variability across participants with respect to ROC parameter estimates (UVSD: signal strength and signal standard deviation; DPSD: recollection and familiarity). Specifically, in each of the 1,000 bootstrap iterations, a random sample of n participants was drawn from the original n participants by random sampling with replacement. For each iteration, the ROC model was fit to the resampled data in the same manner as it was for the original data. The standard deviations of the ROC parameter estimates across these 1,000 iterations served as the standard errors of the mean for the parameters.

A random shuffling procedure was used to estimate statistical significance for the ROC parameter differences between retrieval practice and restudy control conditions for both cue and target words. In this approach, only experimental and control conditions were randomly shuffled using all original participants. No participant bootstrap resampling took place during these shuffles. Specifically, in each of the 1,000 iterations, the retrieval practice and restudy control labels were swapped for a randomly-selected subset of participants. In each iteration, ROC models were then fit to the shuffled data in the same manner as for the original data, and for each ROC parameter estimate an experimental-control difference was calculated. Original parameter estimates that fell in the outer 5$^{th}$ percentile of these distributions was considered statistically significant at an alpha level of .05.

**Results**

We excluded data from individuals whose performance was two standard deviations below the mean on either the final recognition memory test, the cued-recall intervention trials, or the two free recall word lists that were administered solely to assess task engagement (n=4). Data generated from non-primary English speakers (n = 13) and data from trials where a computer error occurred (n = 10) were also excluded from our dataset. Of the 40 remaining participants (Age range = 18-24; n = 27 self-identified as female, n = 4 as a nonbinary gender), the mean proportion of correct responses on cued-recall intervention trials was 0.748 (SD = 0.173). Subsequent analyses of final test performance were based on all old words, irrespective of intervention performance, to avoid retrievability biases that occur when one analyzes final test data for only the stimuli that participants recall during the intervention.

We first assessed how memory performance on the final recognition memory test differed as a function of the word type (cue or target word) and intervention condition (retrieval practice or restudy control). Figure 2 shows performance on the final recognition memory test, using d' based on the normalized hit rate minus the normalized false alarm rate (i.e., not based on ROC models or a specific theory of recognition memory). Three paired t-tests were used to compare memory of retrieval practice cues versus restudy control cues, retrieval practice targets versus restudy control targets, and retrieval practice cues versus retrieval practice targets (see Analyses section for rationale of approach). Memory for both target and cue words learned in the retrieval practice condition was significantly better than memory for words learned in the restudy control condition (target words: $t(39) = 5.41$, $p < .001$; cue words: $t(39) = 3.44$, $p < .01$). There was not, however, a significant difference in memory found between retrieval practice cue words and retrieval practice target words ($t(39) = 0.88$, $p > .05$, $d = 0.14$). Thus, retrieval practice improved final recognition memory test performance relative to restudy control for both cue and target words, though the size of the retrieval practice effect (Cohen's *d*) for target words ($d = 0.86$, indicating a large effect) was numerically greater than that of cue words ($d = 0.54$, indicating a medium effect).

We next assessed how retrieval practice performance during the intervention may have influenced final recognition memory test performance, separately for cue and target words. Specifically, for each participant, cued-recall intervention performance (proportion correct) was calculated separately for trials for which the target word appeared on the final test and trials for which the cue word appeared on the final test. Those intervention scores were then separately correlated with their corresponding final recognition memory test scores (d') for target words and cue words. Figure 3 shows separate scatterplots and highlights significant correlations between retrieval practice intervention performance and final test performance for both target words ($r(39) = 0.70$, $p < .001$) and cue words ($r(39) = 0.423$, $p < .01$). An analysis of the two correlations that accounted for the dependence (within-subject) of the data indicated that the correlation was significantly greater (two-tailed, $p = .018$) for target words compared to cue words (http://quantpsy.org; Lee & Preacher, 2013).

Final recognition memory test performance was also assessed by fitting the data to a classic ROC model of recognition memory, the unequal variance signal detection model (UVSD; Egan, 1958). The model fits two parameters: the difference between distributions of memory strength for old words (signal) and new words (noise), referred to here as signal strength, as well as the signal standard deviation (as a ratio of the standard deviation to the noise distribution). An advantage of ROC models over the simpler estimate of d' (normalized hit rate minus normalized false alarm rate) is that the ROC model parameter estimates offer empirical information about the shape of the memory distribution in addition to information about the strength of the memory (Brady et al., 2023). Figure S1 in the Supplemental Material additionally shows results from fitting the data based on another common ROC model of recognition memory, the dual process signal detection model (DPSD; Yonelinas, 1994). The basic findings from the UVSD and DPSD ROC models were similar: both cue words and target words benefitted from retrieval practice relative to restudy control, but the specific impact on ROC parameter estimates differed between cue words and target words. The main analysis will focus on the results from the UVSD model.

Figure 4 shows the two parameters (signal strength and signal standard deviation) fit by the UVSD model, separately for restudy control cue words, retrieval practice cue words, restudy control target words, and retrieval practice target words. A random shuffling approach was used to determine whether observed differences were greater than one would expect by chance (see Method Section). Relative to the restudy control condition, retrieval practice improved signal strength scores for both cue ($p < .05$) and target ($p < .001$) words. However, a direct comparison between retrieval practice cue and target words indicated that the signal strength scores were significantly ($p < .01$) greater for target words. The estimates of signal standard deviations indicated that, relative to restudy control, retrieval practice numerically increased the variability of the signal for target words yet numerically decreased the variability of the signal for the cue words. Although those comparisons to the control condition did not reach statistical significance (both *p*'s > .05), a direct comparison between retrieval practice cue and target words indicated that the signal standard deviations were significantly ($p < .01$) smaller for cue words. These findings suggest that relative to restudy control, retrieval practice benefitted memory for cue words with a smaller but more consistent boost to memory and target words with a larger but more variable boost to memory.

Using another approach to determine how the influence of retrieval practice might differ between cue words and target words, we compared only correct final test trials for old items (hits) across the conditions. Figure 5 depicts this analysis wherein we analyzed the proportion of hits that were given with high confidence ("sure old") relative to all hits ("sure old", "maybe old", and "guess old"). Proportion of hits given high confidence ratings were found to be similar for cues in the retrieval practice and restudy control condition ($t(39) = 0.24$, $p = .81$, $d = 0.04$). In contrast, the proportion of hits given high confidence ratings for target words was significantly higher in the retrieval practice condition relative to the restudy control condition ($t(39) = 4.62$, $p < .001$, $d = 0.73$). In addition, there was a greater proportion of high confidence ratings for retrieval practice target words relative to retrieval practice cue words ($t(39) = 4.53$, $p < .001$, $d = 0.72$). These results suggest that participants, on average, more frequently provided correct high confidence ratings for retrieval practice target words relative to cue words.

**Discussion**

As demonstrated in prior literature (Rowland, 2014), we found that retrieval practice improved memory of target words relative to memory of restudy control target words, a phenomenon known as the retrieval practice effect. Additionally, we found that retrieval practice also improved memory for cue words. This finding indicates that at least some of the memory benefits from retrieval practice may not depend exclusively on recall during the intervention.

Additional findings from our study suggest that the retrieval practice effect for cue words may be caused by a process distinct from the retrieval practice effect for target words. This interpretation is supported by the following findings. Firstly, the retrieval practice effect size from memory accuracy (d') was numerically greater for target words relative to cue words. Secondly, memory for retrieval practice target words was significantly more correlated with intervention performance than correlations for retrieval practice cue words. Thirdly, we found significantly greater signal strength for retrieval practice target words compared to cue words, along with significantly greater memory variability for target words compared to cue words (signal standard deviation). The UVSD model used to derive these findings accounts for two parameters as opposed to the single parameter with d', thereby providing a more accurate and precise estimation of the signal to noise distribution (Brady et al., 2023). Further results from Experiment 1 also revealed a greater proportion of high confidence responses for retrieval practice target words relative to cue words. Therefore, we hypothesize that there are at least two effects occurring here: a target-only memory retrieval effect and an effect of attention or engagement that applied to both retrieval practice cue and target words.

An alternative interpretation of the data is that a single set of memory-enhancing factors operated on both cue and target words during retrieval practice and that these factors differed from the restudy control condition only in terms of magnitude. That is, relative to restudy control, retrieval practice might have benefitted cue words and target words in the same manner but just to a greater extent for target words. By this interpretation, all of the differences in the patterns of results between retrieval practice cue and target words could be traced back to the observation that the d' retrieval practice effect size was numerically (though not statistically significantly) larger for target words (Cohen's $d = 0.86$) than for cue words (Cohen's $d = 0.54$). Unless a memory-enhancing process adds the same amount of memory strength to each item, that process will tend to increase memory variability in addition to mean memory strength (Wixted, 2007). Thus, the increase in signal standard deviation for retrieval practice target words could have resulted as a byproduct of increased memory signal strength for target words to the extent that, broadly speaking, increased signal could also lead to increased variability. Moreover, the increased proportion of correctly endorsed target words (target word hits) with high confidence in the retrieval practice condition (Figure 5) could be a byproduct of the increased standard deviation for memory of those words. Figure 6 depicts an illustration of this point. Specifically, a wider distribution includes heavier tails, which means that a greater proportion of the distribution would reside in the upper limit of memory strength. To the extent that these stronger memories would be endorsed with high confidence, the wider memory distribution would include a greater proportion of high confidence hits. Similarly, the greater correlation between intervention and final test performance for retrieval practice target words versus cue words could also be attributed to increased variability in final test memory strength. More variability would hypothetically open more parametric space for a stronger correlation to emerge. In short, it was unclear as to whether all of the results from Experiment 1 could have been attributable to the somewhat greater retrieval practice effect size for targets as compared to cues.

Experiment 2 sought to address this possibility by including an additional restudy control condition in which participants restudied word pairs two-times rather than once. The rationale for this design was that two restudy opportunities might improve final memory test performance for cues and targets at a level similar to a single retrieval practice opportunity for cues and targets. We sought to determine whether an increase in overall memory performance via a single retrieval practice or two restudy opportunities would lead to similar or dissimilar changes in other outcomes, such as signal standard deviations, in a manner that still permitted separate evaluation of final test performance for cue words and target words. We pursued this rationale in Experiment 2 rather than trying to equate performance between retrieval practice cues and retrieval practice targets so as to minimize changes to the retrieval practice intervention procedure used in Experiment 1.

# Experiment 2

## Materials and Methods

### *Participants*

Seventy-three undergraduate students were recruited. The *a-priori* sample size for Experiment 2 was based on similar estimations for the sample size in Experiment 1, yet we included additional participants in Experiment 2 because of the unanticipatedly large number of excluded participants from Experiment 1. All procedures were approved by Institutional Review Board at Emory University.

### *Stimuli, Procedure, and Analyses*

Figure 7 provides a schematic of the procedures for Experiment 2. The same procedures and stimuli used in Experiment 1 were applied to Experiment 2, with four exceptions. First, the general administration of the tasks in Experiment 2 were carried out using PsychoPy (http://psychopy.org; Peirce et al., 2019). For each participant, PsychoPy randomly assigned old word pairs evenly to each intervention condition and randomly assigned new words for the final recognition memory test. With PsychoPy we also collected response times from each trial on the final recognition memory test.

Second, an additional restudy control condition was included. In Experiment 2 there were three intervention conditions: a single restudy control condition, a restudy two-times control condition, and a retrieval practice condition. Each block during the initial study consisted of six study trials (two trials per condition), and the intervention phase now contained eight trials. These eight intervention trials consisted of two trials for the single restudy control condition, two trials for the retrieval practice condition, and two trials for the restudy two-times control condition that were repeated (i.e., four total trials for restudy two-times control).

Third, the number of word-pairs used in the intervention conditions reduced from 42 to 36 per condition. Despite the decreases in word pairs per intervention condition, there was an overall increase in the total number of old word trials on the final test. This increase went from 84 to 108 words. With the increase in old words on the final test, we also increased the number of new word trials to 108. Trials were partitioned into six 36-trial blocks with intervening self-paced rest periods in between. Each block was equally divided between old (18) and new (18) words. In each block, old words were equally divided between retrieval practice (6), restudy control (6), and restudy two-times control (6) conditions.

Fourth, the delay interval between the intervention and the final recognition memory test decreased from a two-day delay to a one-day delay. The adjustments to the number of word pairs and test delay were based on pilot testing wherein preliminary results suggested that as the total overall number of word pairs increased, final test performance after a two-day delay decreased.

Despite the decrease in performance after a two-day delay, we found more similar memory retention after a one-day delay to the d' results from Experiment 1.

Further, all analyses carried out in Experiment 1 were also applied to data in Experiment 2. We continued to make direct contrasts between conditions for each word type using two-tailed paired samples t-tests. For Experiment 2, additional comparisons included retrieval practice cues versus restudy two-times control cues and retrieval practice targets versus restudy two-times control targets. A repeated-measures analysis of variance (ANOVA) is not suitable as it does not directly address the present study's core questions. For instance, a main effect for word type or intervention condition, collapsed across all three conditions, would not directly address the study's primary questions. Additionally, an interaction between intervention conditions and word types, although relevant, could still be ambiguous as it could result, for example, from either a unique or disproportionate retrieval practice effect for targets relative to cues. Therefore, to better address the key questions of interest, we continued with analyses using two-tailed paired samples t-tests. One additional analysis was introduced in this experiment that tested response times between conditions on the final recognition memory test. Analyzing response times provides valuable insights into potential variations in cognitive processes associated with decision-making.

**Results**

We excluded data from individuals whose performance was two standard deviations below the mean on either the final recognition memory test, the cued-recall intervention trials, or the two free recall word lists that were administered solely to assess task engagement (n=4). Data generated from non-primary English speakers (n = 15) were also excluded from our dataset. Of the 54 remaining participants (age range = 18-24; n = 37 self-identified as female, n = 3 as a nonbinary gender), mean proportion of correct responses on cued-recall intervention trials was 0.68 (*SD* = 0.20). Subsequent analyses of final test performance were based on all old words, irrespective of intervention performance.

We first addressed how memory performance on the final recognition memory test differed as a function of word type and intervention condition. Figure 8 shows performance on the final recognition memory test using d' (normalized hit rate minus the normalized false alarm rate). Paired t-tests revealed a statistically significant retrieval practice effect for memory of both cue and target words relative to the single restudy control condition (cues: $t(53) = 2.58$, $p < .05$, $d = 0.35$; targets: $t(53) = 3.86$ , $p < .001$, $d = 0.52$). Direct contrasts between retrieval practice cue and target words revealed that memory accuracy was slightly better for retrieval practice target words, though not significantly better ($t(53) = 1.91$, $p = .06$, $d = 0.26$). Thus, the results from Experiment 1 were largely replicated for the conditions that were repeated again in Experiment 2 insofar as retrieval practice improved performance relative to a single restudy control condition for both target and cue words. Moreover, the effect size was numerically larger for memory of retrieval practice target words relative to retrieval practice cue words.

The results for the restudy two-times control condition indicated that the extra study opportunity improved final test performance (Figure 8). However, for both cue and target words, final test performance in the restudy two-times control condition was still numerically lower than that of cue and target words in the retrieval practice condition. The effect sizes for those differences were small and did not approach statistical significance (cues: $t(53) = 1.17$, $p = .25$, $d = 0.16$; targets: $t(53) = 1.63$, $p = .11$, $d = 0.22$). Thus, providing an additional restudy opportunity to engage with word pairs provided a gain in memory accuracy for cue and target words but did not quite match the retrieval practice condition performance.

To further explore memory performance differences on the final test based on word type and intervention condition, we analyzed response times. Measuring response times allows us to assess variations in decision-making cognitive processes. Figure 9 depicts the direct contrasts in response times for both words between intervention conditions. The direct contrasts revealed significantly faster response times for retrieval practice target words relative to target words from both restudy control conditions (restudy one-time: $t(53) = 3.32$, $p < .01$, $d = 0.45$; restudy two-times: $t(53) = 3.63$ , $p < .001$, $d = 0.49$). A similar benefit in response times was found for retrieval practice cue words relative to restudy control cue words ($t(53) = 2.08$, $p < .05$, $d = 0.28$). However, response times for cue words between the retrieval practice and restudy two-times control condition were not significantly different from one another ($t(53) = 0.17$ , $p = .87$, $d = 0.02$). A direct comparison between participant response times for retrieval practice cue versus target words also revealed significantly faster response times for target words ($t(53) = 2.76$, $p < .01$, $d = 0.37$). In sum, these results demonstrated faster decision-making processes for both cue and target words in the retrieval practice condition compared to the single restudy control condition. Additionally, there was a greater advantage in decision-making processes specifically for retrieval practice target words, in which response times were also faster relative to restudy two-times control target words and retrieval practice cue words.

We next assessed how retrieval practice performance during the intervention influenced final recognition memory test performance, separately for cue and target words. These correlations were carried out similar to the procedure described in Experiment 1. Figure 10 demonstrates separate scatter plots and correlations between retrieval practice intervention performance and final test performance (d') for both target words ($r(53) = 0.733$, $p < .001$) and cue words ($r(53) = 0.464$, $p < .001$). An analysis of the two correlations that accounted for the dependence (within-subject) of the data indicated that the correlation was significantly greater (two-tailed $p = .019$) for target words compared to cue words. These results replicate the parallel analyses from Experiment 1.

Final recognition memory test performance across conditions was also assessed by fitting the data to the two parameters from the UVSD model. Figure 11 depicts the signal strength and signal standard deviation fit by the UVSD model and assessed separately for restudy control cue words, restudy two-times control cue words, retrieval practice cue words, restudy control target words, restudy two-times control target words, and retrieval practice target words. Figure S2 in the Supplemental Material shows the results from fitting the data to the parameters from the DPSD model. Results from the random shuffling approach, applied to ask if there were observed differences between the conditions indicated that, relative to the single restudy control condition, retrieval practice improved signal strength for both target ($p < .001$) and cue ($p < .01$) words. Retrieval practice also improved signal strength relative to the restudy two-times control condition for both target ($p < .001$) and cue ($p = .001$) words. Signal strength was also found to be significantly greater for retrieval practice target words relative to retrieval practice cue words ($p < .001$).

Direct comparisons of the signal standard deviation from the UVSD model for target words between the retrieval practice and both restudy control conditions revealed greater variability for retrieval practice target words (single restudy control: $p < .001$; restudy two-times control: $p < .001$). While the signal standard deviation for cue words between the retrieval practice condition and the restudy control condition was similar ($p = .82$), variability was greater for retrieval practice cue words relative to restudy two-times control cue words ($p < .05$). The direct comparison between retrieval practice cue and target words also indicated that there was more variability for retrieval practice target words ($p < .001$). Although, the additional study opportunity in the restudy two-times control condition did not boost overall d' memory performance to the same level as retrieval practice (Figure 8), it did bring overall d' performance closer to retrieval practice levels for both cue and target words. Despite that increase in overall d' performance for restudy two-times cue and target words, there was no evidence in the ROC analyses that this increase in performance resulted in a greater signal variability. Indeed, the signal standard deviation was slightly numerically less for restudy two-times cue and target words as compared to cues and targets in the single restudy control condition. This result signifies that an increase in memory strength from an additional restudy opportunity does not consequently lead to more response variability.

We additionally explored how the influence of retrieval practice might differ between cue words and target words by examining only correct final test trials for old items (hits) among the conditions. Figure 12 depicts the results of the proportion of hits given high confidence analysis. The proportion of high confident responses for retrieval practice target words were found to be significantly higher relative to target words from both restudy control conditions (single restudy: $t(53) = 6.89$, $p < .001$, $d = 0.94$; restudy two-times: $t(53) = 6.25$, $p < .001$, $d = 0.85$). The proportion of high confident responses for retrieval practice cue words were also significantly higher relative to cue words from both restudy control conditions (single restudy: $t(53) = 2.54$, $p < .05$, $d = 0.35$; restudy two-times: $t(53) = 2.57$, $p < .05$, $d = 0.35$). Nonetheless, retrieval practice target words contained a greater proportion of high confidence ratings relative to retrieval practice cue words ($t(53) = 4.07$, $p < .001$, $d = 0.55$). In sum, these results suggest that correct high confident responses were more pronounced for both retrieval practice words relative to words from both restudy control conditions. Moreover, there were particularly more frequent high confidence ratings for retrieval practice target words relative to retrieval practice cue words, replicating the findings from Experiment 1.

**Discussion**

A retrieval practice effect was observed for both cue and target words compared to words from the single restudy control condition, replicating findings from Experiment 1. Similar to the UVSD results from Experiment 1, relative to the restudy control condition, retrieval practice cues received a modest yet consistent memory improvement whereas retrieval practice target words received a stronger and more variable memory improvement. However, the results of the memory accuracy analysis (d') did not reveal a retrieval practice effect for cue and target words compared to words from the restudy two-times control condition. In contrast to results from the memory accuracy analysis, the parameters from the UVSD model indicated greater memory strength and variability for both retrieval practice cue and target words compared to words from the restudy two-times control condition. Furthermore, there was a greater proportion of hits given high confidence responses for retrieval practice cue and target words relative to words from both restudy control conditions.

It seems that providing a second restudy opportunity in Experiment 2 does improve final memory test performance at least relative to the single restudy control condition. Despite this small memory improvement, memory strength and memory variability were still lower for restudy two-times control cue and target words relative to both words from the retrieval practice condition. These findings therefore do not provide support for the notion that an improvement in memory performance necessarily translates into corresponding increases in other outcome variables, specifically in terms of memory variability. Nevertheless, the results obtained in Experiment 2 confirm that a single retrieval practice condition consistently leads to a significantly greater memory improvement compared to restudying the information twice. Moreover, similar to findings from Experiment 1, we continued to find a greater memory improvement for retrieval practice target words relative to retrieval practice cue words. What remains unclear is if the potential differences in memory performance found between retrieval practice cue and target words result from differences in memory strength or if these differences are the result of two distinct memory processes.

## General Discussion

The current study across two experiments revealed a retrieval practice effect for target and cue words relative to a restudy control condition. The finding of a retrieval practice effect for cue words confirmed that retrieval practice effects do not necessarily depend on recall, the type of memory retrieval of missing information. The memory accuracy analyses (d') from both experiments revealed a significant memory improvement for both retrieval practice cue and target words relative restudy control words, with no significant differences between retrieval practice cue and target words. The results from the unequal variance signal detection (UVSD) model, however, across both experiments revealed underlying differences in retrieval practice effects between cue and target words. Specifically, relative to the restudy control condition, retrieval practice cue words received a modest yet consistent memory improvement, whereas retrieval practice target words received a larger and more variable memory improvement. The results from additional analyses, including the correlation analysis, response times analysis (Experiment 2), and proportion of hits given high confidence analysis, further support the notion that retrieval practice enhances memory for target words to a greater extent than for cue words. In sum, these findings suggest that retrieval practice effects for cue words are influenced by factors associated with deeper levels of engagement during testing, while retrieval practice effects for target words are influenced by both deeper levels of engagement and memory recall.

The results from the current study provides findings with substantial implications compared to prior studies investigating retrieval practice effects for cue and target words (Carpenter et al., 2006; Carpenter, 2011). While the previous studies provided valuable insights, they had some methodological limitations that affected the strength of their conclusions. To address some of those limitations, the present study used a more robust analysis to evaluate recognition memory performance. For instance, whereas the Carpenter (2011) study reported responses as a function of hit rates and false alarms separately, the present study used a more standard measure of memory using d' to assess recognition memory test results. We additionally included an additional method using ROC analyses from recognition memory signal detection models. An additional limitation that was addressed in the current study was the separate assessment of memory for retrieval practice cue and target words, whereas Carpenter (2011) tested both cue and target words from each word pair which could have biased their results. Thus, the present study included additional analyses for both retrieval practice cue and target words, which were not conducted in prior studies, to assess retrieval practice effects for stimuli requiring recall versus stimuli that did not require recall. These analyses allowed us to examine these retrieval practice effects in a more comprehensive and distinct manner, highlighting the contribution of the current study.

A retrieval practice effect was found for cue words, yet both retrieval practice and restudy control cue words were provided on the screen during the intervention. Moreover, the task requirements for both conditions included typing in the provided cue word. The key distinction between the conditions was that retrieval practice cue words were associated with a to-be-recalled target word. Since the cue word from both conditions was present during initial study and the intervention, it would be expected that automatic recognition memory processes on the final test would be similar for cue words from both conditions. Despite this assumption, the parameters from the UVSD model depicted a greater yet consistent memory improvement to retrieval practice cue words relative to restudy control cue words. The findings from both experiments indicate that this memory improvement for retrieval practice cue words is not dependent on recall but rather relies on encoding factors related to deeper engagement during the retrieval process, such as attention, arousal, or vigilance. Furthermore, it is evident that restudying word pairs twice did not result in similar improvements to memory strength or memory variability of cue words as observed in the retrieval practice condition. Nevertheless, this memory improvement for retrieval practice cue words was less than the memory improvement afforded to retrieval practice target words. Memory for retrieval practice target words improved both from deeper levels of engagement with the stimuli and memory recall. This greater memory benefit was found in both experiments. Results from the UVSD model indicated a larger and more variable memory improvement, as well as a higher proportion of high confidence and faster decision-making responses, relative to that for retrieval practice cue words.

These findings can be better contextualized and understood with consideration to hypotheses that explain the mechanisms behind retrieval practice effects. These hypotheses provide valuable implications regarding the occurrence of retrieval practice effects on the cue word, the target word, or both words. One collection of ideas proposed that harder-to-retrieve information that requires greater mental effort supports the retrieval practice effect (e.g., Bjork, 1975; Bjork & Bjork, 1992; Halamish & Bjork, 2011; Kang et al., 2007; Kornell et al., 2011). We have categorized these ideas as the effortful search hypothesis. One research study expanded on this idea suggesting that more difficult retrieval attempts are beneficial only if they are successful (Pyc & Rawson, 2009). The disproportionate memory benefit provided to retrieval practice target words from the current study supports this idea. However, according to this hypothesis, retrieval practice should benefit memory for only target words and not cue words since no mental effort was expended to retrieve cue words. With respect to memory strength and memory variability, an interpretation of this hypothesis would assume an improvement to memory strength and variability only for target words relative to restudy control target words. There should be no improvements to memory strength or variability for retrieval practice cue words relative to restudy control cue words. The observation of a retrieval practice effect for cue words from the current study contradicts this simple interpretation. For this reason, this hypothesis as currently provided, does not account for the finding of a retrieval practice effect for both cue and target words.

Another set of ideas revolves around retrieval practice cue words. According to this view, the provided cue word triggers semantically-related associations and memory recall for the to-be-retrieved target word. This activation of additional associations enriches the memory for word pairs and contributes to retrieval practice effects (e.g., Carpenter, 2009; 2011; Carpenter & Yeung, 2017; Pyc & Rawson, 2012). We will categorize this set of ideas as the elaborative association hypothesis. Later research added to this hypothesis suggesting that the association linking cue to the target information is unidirectional (Carpenter & Yeung, 2017), leaving to interpretation that memory would be better for the elaborated-on cue words relative to target words. Based on this interpretation, we would anticipate a greater memory advantage specifically for retrieval practice cue words. With respect to memory strength and memory variability, we could interpret from this hypothesis that a greater improvement in memory strength and variability for retrieval practice cue words should be observed relative to restudy control cue words. In contrast, improvements in memory strength and variability for retrieval practice target words would be modest relative to restudy control target words. The findings from the current study, however, do not align with this hypothesis. Yet, the concept that retrieval promotes elaborative associations and establishes more connections, leading to a memory benefit, aligns with previous notions that deeper processing of information also enhances memory (Craik & Lockhart, 1972; Craik & Tulving, 1975). A key distinction between the elaborative association hypothesis and previous ideas regarding deeper levels of processing is that depth of processing does not necessarily rely upon retrieval to receive a memory benefit.

An alternative set of ideas proposed that retrieval practice uses provided cues to reactivate the original study episode, including the items and prior context (e.g., Karpicke et al., 2014; Lehman & Malmberg, 2013; Rowland & DeLosh, 2014; Tulving, 2005; Whiffen & Karpicke, 2017). Furthermore, successful retrieval practice would integrate the study and retrieval trial episodes, leading to a subsequent memory benefit. We will categorize this set of ideas as the episodic reactivation hypothesis. This hypothesis builds on the foundational principles of the Temporal Context Model (Howard & Kahana, 2002) which emphasizes the reactivation of prior item-context representations, such that retrieval reinstates the original context and leads to better memory. According to the episodic reactivation hypothesis account, the retrieval practice effect should benefit equally both the cue and target word, as reactivating and integrating the study and retrieval episodes should benefit both words. In relation to memory strength and memory variability, an interpretation of this hypothesis would assume that a similar memory improvement for both retrieval practice cue and target words should be observed relative to both restudy control words. This account raises two potential concerns. First, is reactivation of the initial study episode restricted to only retrieval, or is it possible for reactivation to also occur during a restudy control trial? Perhaps it does occur in the restudy control condition, but not to the same extent as retrieval practice. Second, even if retrieval practice leads to a greater advantage in reactivation, the findings in the present study, beyond the memory accuracy analysis, did not demonstrate similar retrieval practice effects for cue and target words. Instead, many of the results demonstrated some unique memory distinctions. Thus, the findings suggest the memory processes for retrieval practice cue words relative to target words are different.

The findings in the present study cannot be fully explained by a single categorical hypothesis alone, as currently proposed, but it is possible that these hypotheses are not mutually exclusive (Carpenter & Yeung, 2017). For instance, the retrieval practice effect for cue words could be attributed to the episodic reactivation hypothesis. Alternatively, additional encoding factors like deeper engagement, attention, and vigilance during retrieval practice may be contributing to the effect. We believe this latter interpretation is more likely to be what is occurring. These encoding factors may be activated during memory search, resembling concepts from the elaborative association hypothesis, even though recall is not dependent for the effect on retrieval practice cue words to occur. These ideas align more closely with the concept of deeper levels of processing (Craik & Lockhart, 1972). Even though retrieval practice also increased memory for cue words, the disproportionate retrieval practice effect for target words aligns with the effortful search hypothesis, wherein increased mental effort during retrieval provides a greater memory improvement. This greater effort for target words from recall may also contribute to increases in memory strength and greater memory variability. Further research, for instance including neuroimaging methods, is needed to test these ideas more broadly.

A potential limitation of Experiment 1 was that we did not meet the intended sample size. Despite aiming *a-priori* for a sample of 52 participants, we experienced a higher participant exclusion rate than anticipated, which may have decreased overall statistical power. Even with this complication, the results were largely replicated in Experiment 2 using the initially planned sample size, providing evidence for the replicability of the findings from Experiment 1.

Additionally, it is important to note that the findings from retrieval practice studies using word pairs as stimuli may not always apply to real-world memory scenarios or different types of memory tasks. Memory performance can vary across different contexts and tasks, suggesting caution in generalizing the results beyond the specific experimental conditions employed in the present study. It is also possible that performance from the restudy two-times control condition may have benefitted from a longer interval between each restudy opportunity during the intervention than what was currently provided. Accordingly, future research may benefit from investigating this idea with a longer delay between restudy opportunities, and then assessing memory performance relative to a retrieval practice condition. Despite these limitations, the findings of the present study indicate that recall during retrieval practice is not strictly required to observe a retrieval practice effect. Nonetheless, memory recall during a retrieval practice intervention can significantly influence the effect.

**Acknowledgements**

We would like to thank Daniela Minondo, Solana Rivera, and Bennett Levine for their contributions to the study design and data collection. No external funding sources were obtained for these studies. None of the authors have any financial disclosures to provide.

**References**

Abott, E. E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements*, *11*(1), 159-177.

Bäuml, K. H. T., & Trißl, L. (2022). Selective memory retrieval can revive forgotten memories. *Proceedings of the National Academy of Sciences*, *119*(8), 1-6.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In *Information Processing and Cognition* (Ed.), 123-144. Wiley Publishing.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *From Learning Processes to Cognitive Processes: Essays in honor of William K. Estes*, *2*, 35-67.

Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2023). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin & Review*, *30*(2), 421-449.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918-928.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563-1569.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1547-1552.

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin and Review*, *13*(5), 826-830.

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language*, *92*, 128-141.

Chan, J. C. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, *61*(2), 153-170.

Chan, J. C., McDermott, K. B., & Roediger III, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*(4), 553-571.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-428.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497-505.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268-294.

Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology*, *21*(6), 919-940.

Egan, J. P. (1958). Recognition memory and the operating characteristic. *USAF Operational Applications Laboratory Technical Note*.

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 801-812.

Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, *10*(2), 169-207.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269-299.

Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4-5), 528-558.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In *Psychology of Learning and Motivation*, *61*, 237-284. Academic Press.

Koen, J. D., Barrett, F. S., Harlow, I. M., & Yonelinas, A. P. (2017). The ROC Toolbox: A toolbox for analyzing receiver-operating characteristics derived from confidence ratings. *Behavior Research Methods*, *49*, 1399-1406.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85-97.

LaPorte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. *Journal of Educational Psychology*, *67*(2), 259-266.

Lee, I. A., & Preacher, K. J. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.

Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, *120*(1), 155-190.

Lozito, J. P., & Mulligan, N. W. (2006). Exploring the role of attention during memory retrieval: Effects of semantic encoding and divided attention. *Memory & Cognition*, *34*, 986-998.

Malmberg, K. J., Lehman, M., Annis, J., Criss, A. H., & Shiffrin, R. M. (2014). Consequences of testing memory. *Psychology of Learning and Motivation*, *61*, 285-313**.**

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review*, *14*, 858-865.

Mulligan, N. W., & Picklesimer, M. (2016). Attention and the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(6), 938-950.

Nielson, K. A., & Powless, M. (2007). Positive and negative sources of emotional arousal enhance long-term word-list retention when induced as long as 30 min after learning. *Neurobiology of Learning and Memory*, *88*(1), 40-47.

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, *74*(1), 18-22.

Pastore, R. E., & Scheirer, C. J. (1974). Signal detection theory: Considerations for general application. *Psychological Bulletin*, *81*(12), 945-958.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195-203.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437-447.

Pyc, M. A., & Rawson, K. A. (2012). Why is test–restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 737-746.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255.

Roediger III, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. *Psychology of Learning and Motivation*, *55*, 1-36.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432-1464.

Rowland, C. A., & DeLosh, E. L. (2014). Benefits of testing for nontested information: Retrieval-induced facilitation of episodically bound material. *Psychonomic Bulletin & Review*, *21*, 1516-1523.

Tanner, W., & Swets, J. (1954). The human use of information--I: Signal detection for the case of the signal known exactly. *Transactions of the IRE Professional Group on Information Theory*, *4*(4), 213-221.

Tulving, E. (2005). Episodic Memory and Autonoesis: Uniquely Human? In H. S. Terrace & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness*, 3–56. Oxford University Press.

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.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152-176.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341-1354.

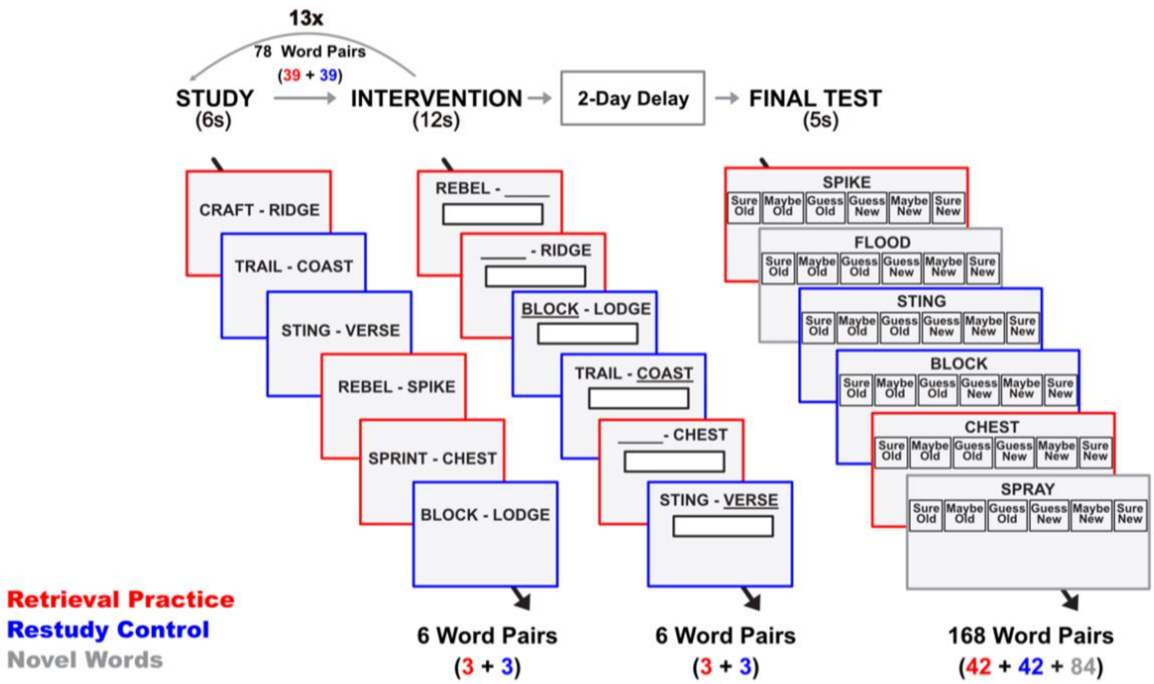


**Figure 1**

*Schematic of Procedures in Experiment 1*

*Note.* Each participant underwent an initial study phase, intervention phase, and a final test phase administered after a 2-day delay. Word pairs were presented in blocks of 6 in the initial study followed immediately by the intervention phase. This cycle repeated a total of 14 times.

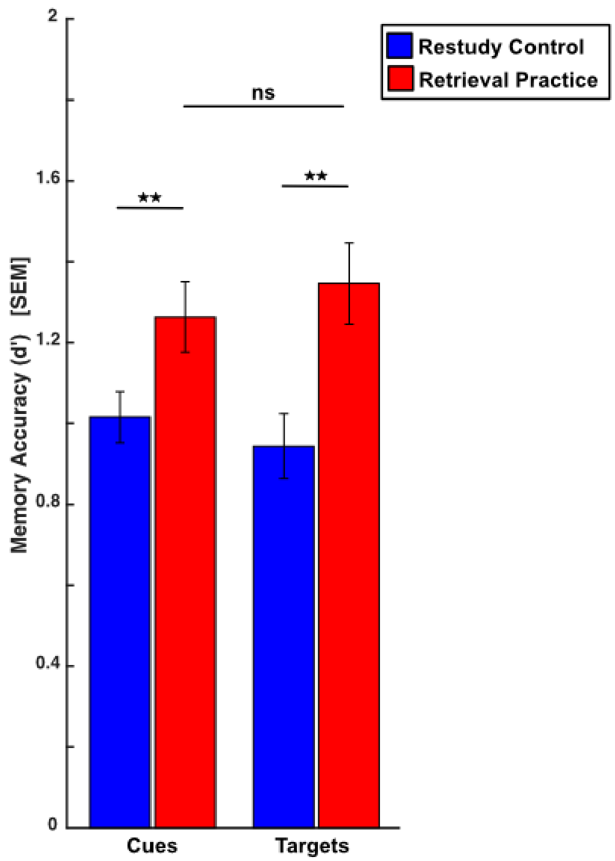


**Figure 2**

*Memory Accuracy Results from Experiment 1*

*Note.* Error bars based on SEM; **indicates $p < .01$; ns = not significant.

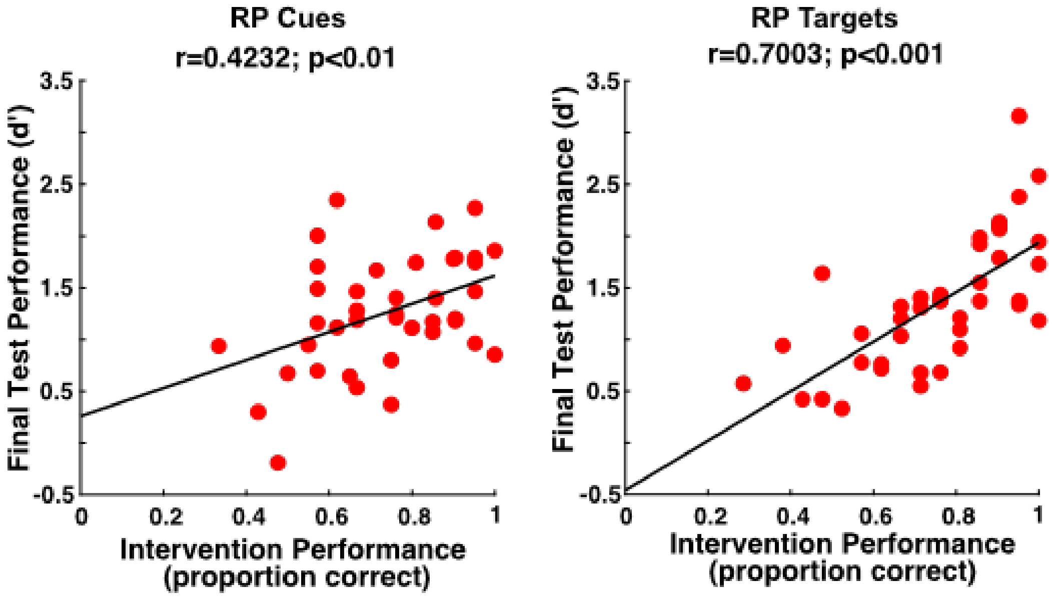


**Figure 3**

*Scatterplot and Correlations from Experiment 1*

*Note.* Correlations between performance during the retrieval practice intervention and on the final test, separately for retrieval practice cues (left) and retrieval practice targets (right) in Experiment 1.

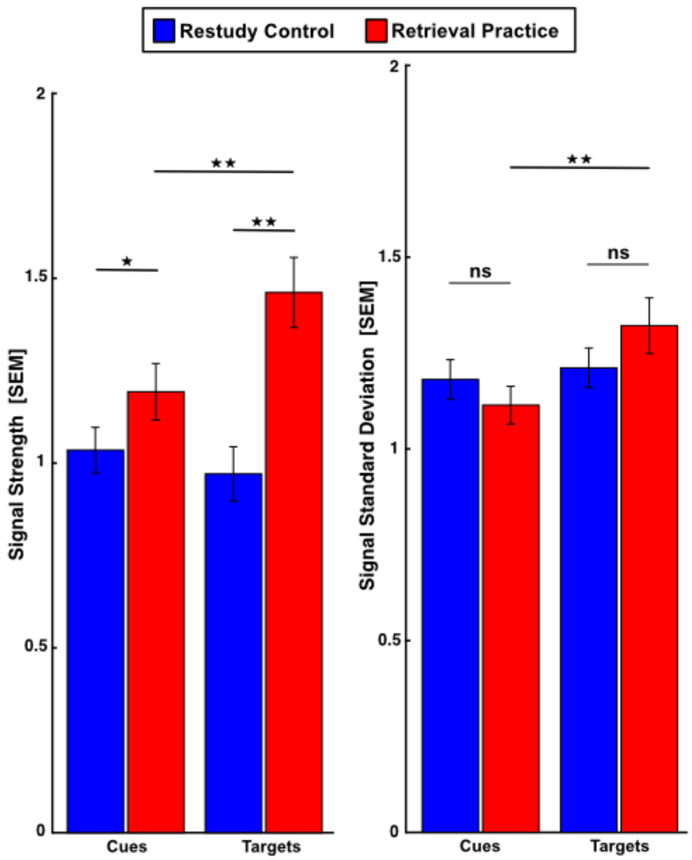


**Figure 4**

*Unequal Variance Signal Detection Model Results from Experiment 1*

*Note.* The two parameters from the unequal variance signal detection theory model depict signal strength (left) and the signal standard deviation (right) in Experiment 1. Error bars depict SEM based on random permutation, whereas p-values are based on random shuffling procedures. *indicates $p < .05$; **indicates $p < .01$; ns = not significant.

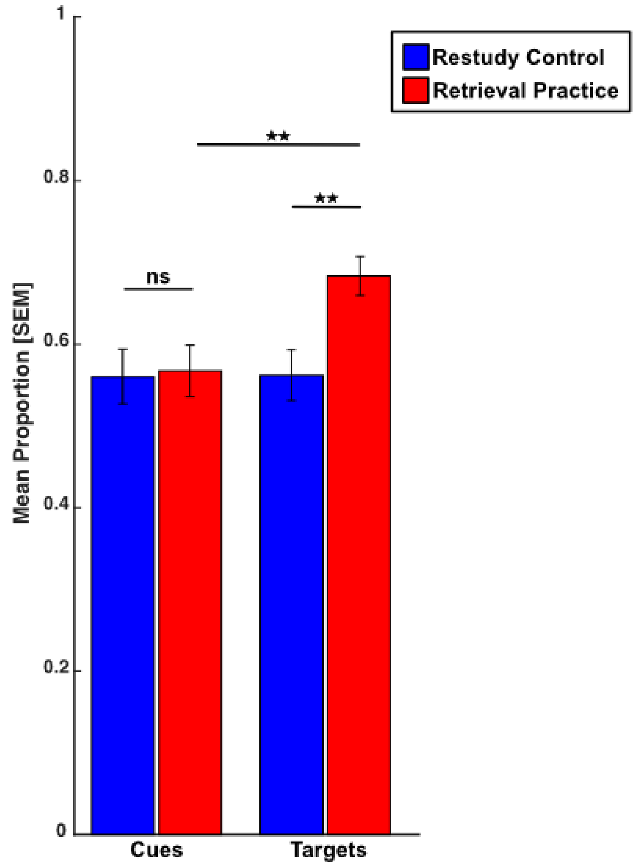


**Figure 5**

*Proportion of Hits Given High Confidence Analysis from Experiment 1*

*Note.* Error bars based on SEM; **indicates $p < .01$; ns = not significant.

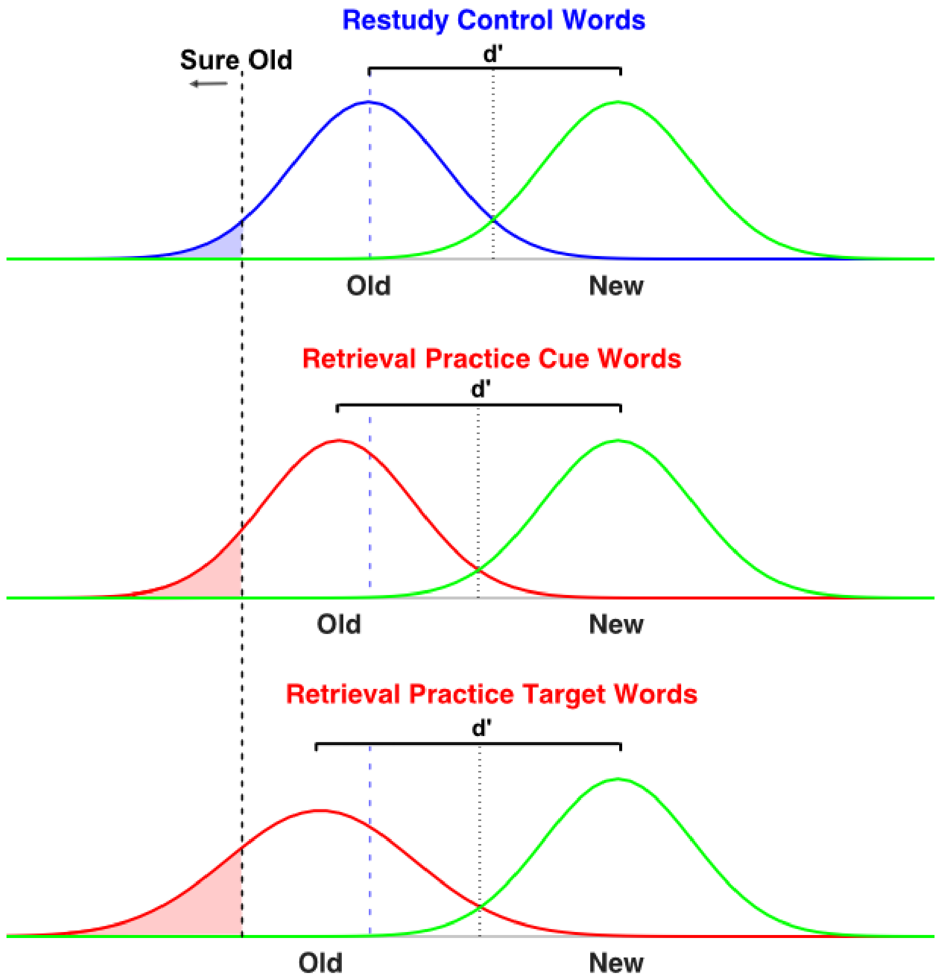


**Figure 6**

*Example of Signal Detection Theory Results from Experiment 1*

Note. Signal strength for retrieval practice cue words (middle) is greater relative to words from the restudy control condition (top), though the signal standard deviation is the same. However, signal strength and the signal standard deviation for retrieval practice target words (bottom) is greater relative to both retrieval practice cue words and words from the restudy control condition. The middle-dashed line depicts old/new discrimination whereas the far left-dashed line depicts the proportion of high confident responses. The blue dotted line depicts signal strength for restudy control words as a comparison relative to retrieval practice words.

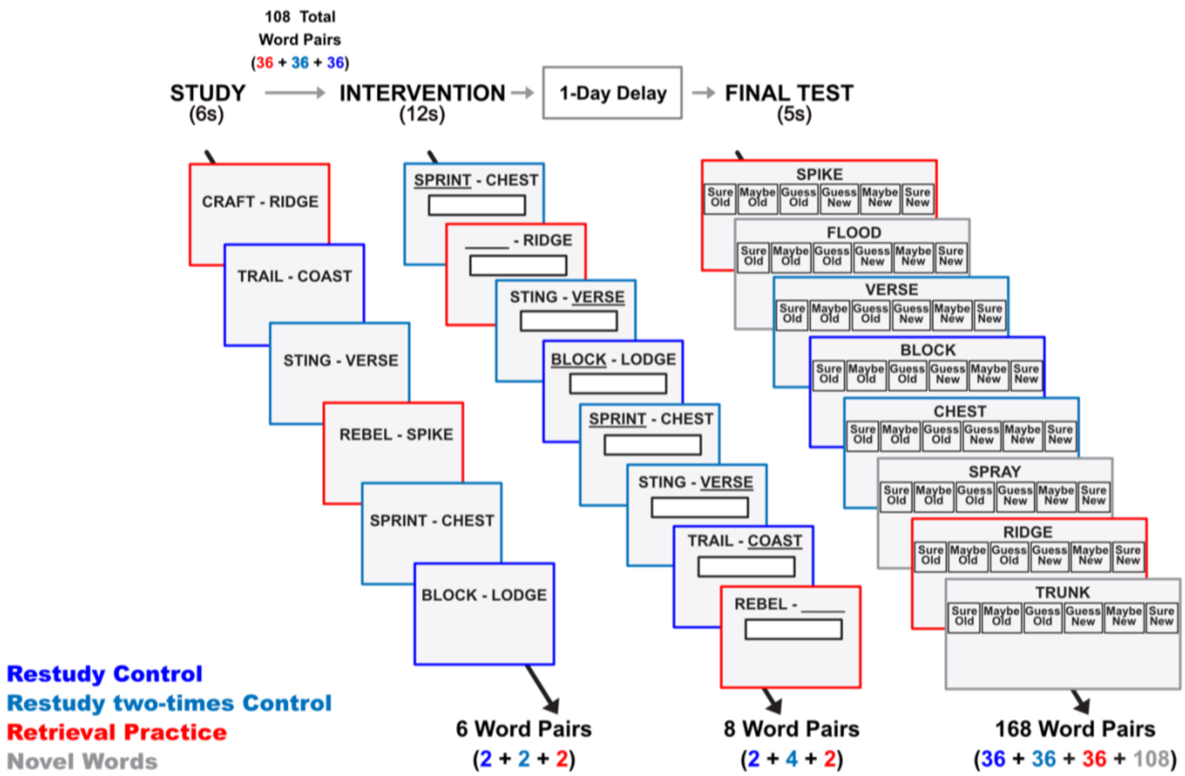


**Figure 7**

*Schematic of Procedures in Experiment 2*

*Note.* The procedure consisted of three phases (initial study, intervention, final memory test), with the initial study of six word-pairs. Unlike in Experiment 1, the immediate intervention phase, however, included 2 additional word-pairs in the restudy two-times control condition. The initial study followed immediately by the intervention phase was repeated for a total of 18 times.

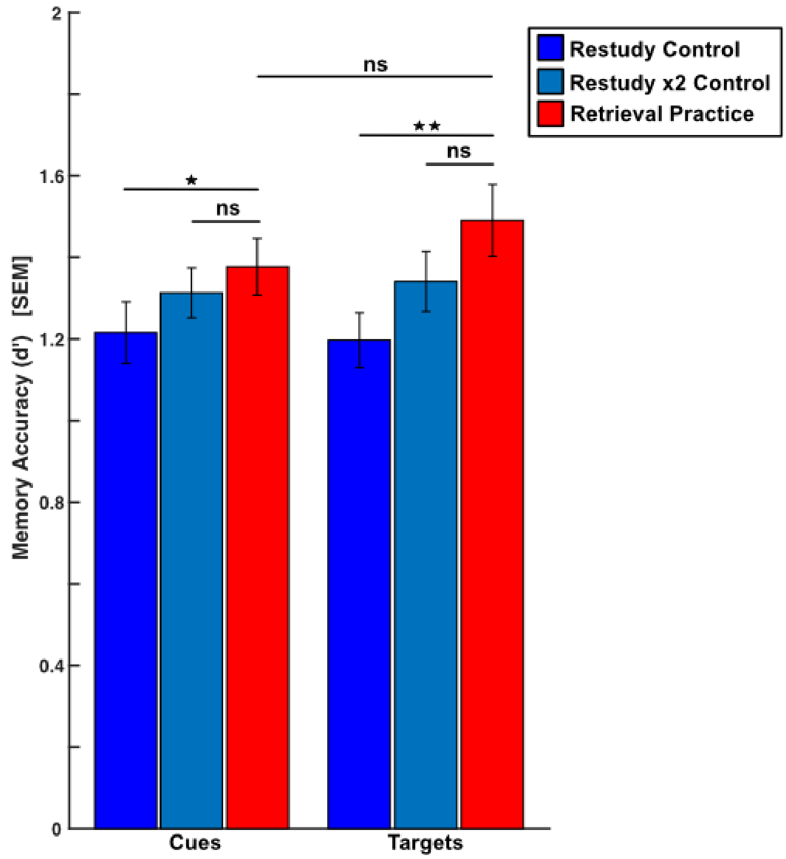


**Figure 8**

*Memory Accuracy Results from Experiment 2*

*Note.* Error bars based on SEM; *indicates $p < .05$; **indicates $p < .01$; ns = not significant.

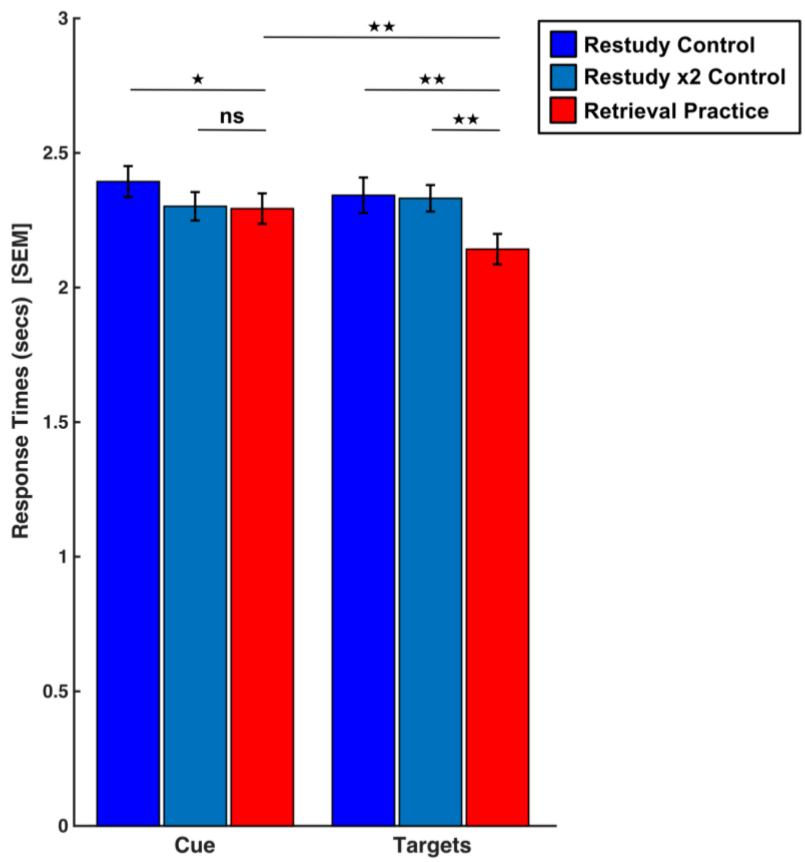


**Figure 9**

*Response time analysis from Experiment 2*

*Note.* Error bars based on SEM; *indicates $p < .05$; **indicates $p < .01$; ns = not significant.

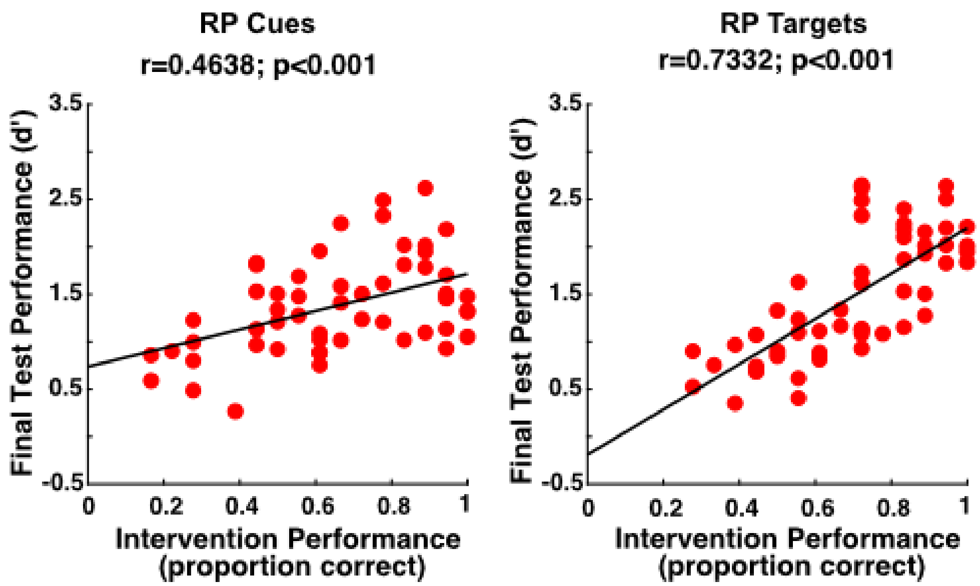


**Figure 10**

*Scatterplot and Correlations from Experiment 2*

*Note.* Correlation results from intervention performance relative to final test performance separately for retrieval practice cue and target words in Experiment 2.

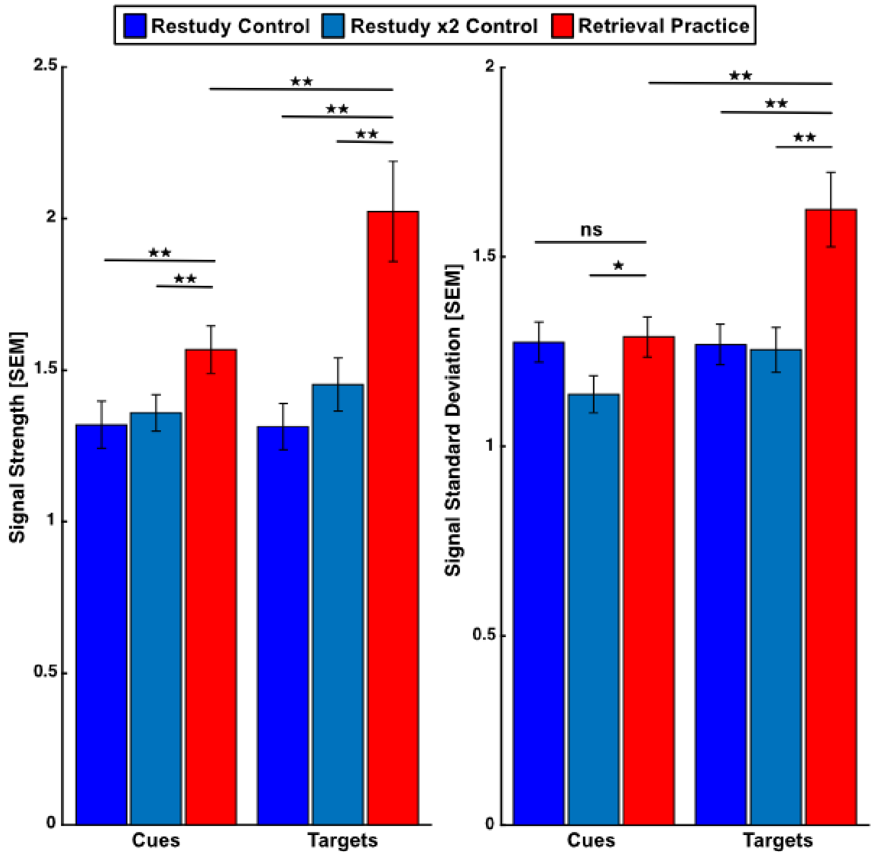


**Figure 11**

*Unequal Variance Signal Detection Model Results from Experiment 2*

*Note.* The two parameters from the UVSD model depicting signal strength and signal standard deviation from Experiment 2. Error bars depict SEM based on random permutation, whereas p-values are based on random shuffling procedures. *indicates $p < .05$; **indicates $p < .01$; ns = not significant.

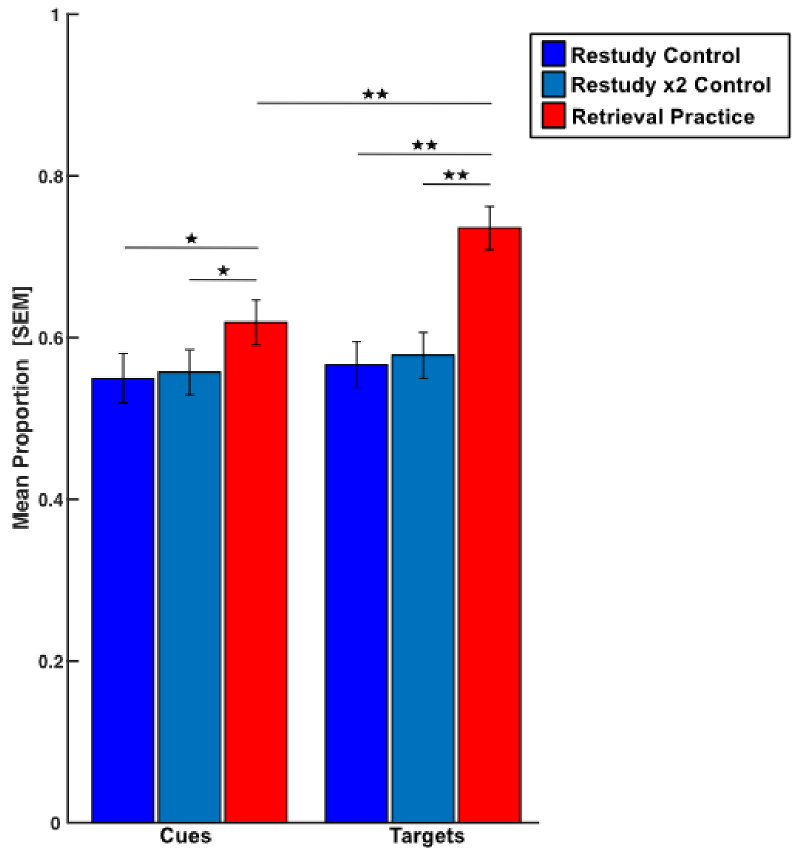


**Figure 12**

*Proportion of Hits Given High Confidence Analysis from Experiment 2*

*Note.* Error bars based on SEM; *indicates $p < .05$; **indicates $p < .01$.

# Supplemental Material

## Supplementary Text

### *Results from Experiment 1*

A common ROC model of recognition memory is the dual process signal detection model (DPSD; Yonelinas, 1994). The two parameter estimates from the DPSD model include familiarity, similar to signal strength from the UVSD model, and recollection, an all-or-none threshold process. Familiarity is considered a fast and automatic memory process whereas recollection is considered a slower memory process. A random shuffling approach was used to ask if observed differences were greater than one would expect by chance, similar to results from the UVSD model (see Method Section). Figure S1 depicts the results of this analysis. Relative to the restudy control condition, retrieval practice improved familiarity for cue words ($p < .01$), though not for target words ($p = .09$). A direct comparison between retrieval practice cue and target words found no differences in familiarity scores between both word types ($p = .17$). In contrast to findings for familiarity, the results from the recollection parameter indicated that recollection for retrieval practice target words was better relative to restudy control target words ($p < .01$), though the same memory benefit did not apply for cue words between the retrieval practice and restudy control conditions ($p = .42$). A direct contrast for recollection between retrieval practice cue and target words revealed a greater recollection for target words compared to cue words ($p < .001$). These results suggest that relative to restudy control words, recollection was strongest for retrieval practice target words whereas familiarity was greater for retrieval practice cue words.

***Results from Experiment 2***

Figure S2 depicts the results of the analysis from the DPSD model. The results from the DPSD model, with a random shuffling approach, revealed improved familiarity from restudying the word lists two times for target words relative to the familiarity of retrieval practice target words, ($p < .05$). Furthermore, familiarity for retrieval practice cue words was also better relative to familiarity for retrieval practice target words ($p < .001$). There was however, no other significant distinctions in familiarity between retrieval practice and the single restudy control condition for cue ($p = .11$) or target ($p = .38$) words, nor between retrieval practice and the restudy two-times control condition for cue words ($p = .47$). However, recollection was found to be significantly greater for retrieval practice target words relative to target words from both restudy control conditions (single restudy: $p < .001$; restudy two-times: $p < .001$) and retrieval practice cue words ($p < .001$). Moreover, recollection for retrieval practice cue words was better relative to restudy two-times control cue words ($p < .05$), although they were not any different when compared to cue words from the single restudy control condition ($p = .13$). In sum, these results suggest that an additional restudy opportunity improved familiarity for cue and target words, yet this improvement was substantially better relative to retrieval practice target words. Familiarity was also better for retrieval practice cue words relative to retrieval practice target words, although the opposite was true in regards to recollection with a much stronger sense of recollection for target words. Moreover, recollection memory for retrieval practice target words was vastly superior relative to words from both restudy control conditions.

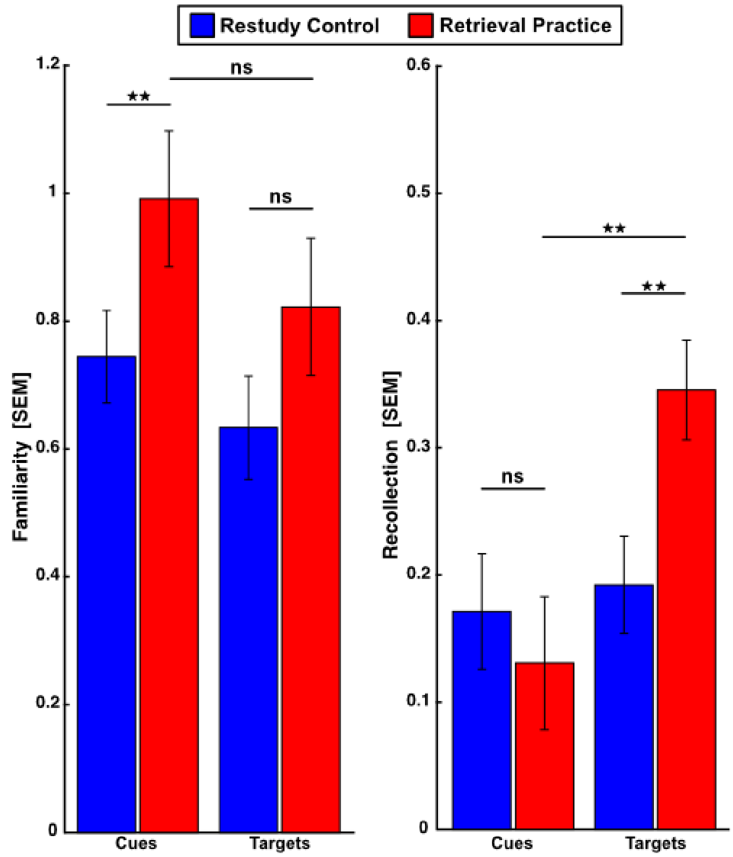


**Figure S1**

*Dual Process Signal Detection Model Results from Experiment 1*

*Note.* The two parameters from the dual process signal detection theory model depict familiarity (left) and recollection (right) in Experiment 1. Error bars depict SEM based on random permutation, whereas p-values are based on random shuffling procedures. **indicates $p < .01$; ns = not significant.

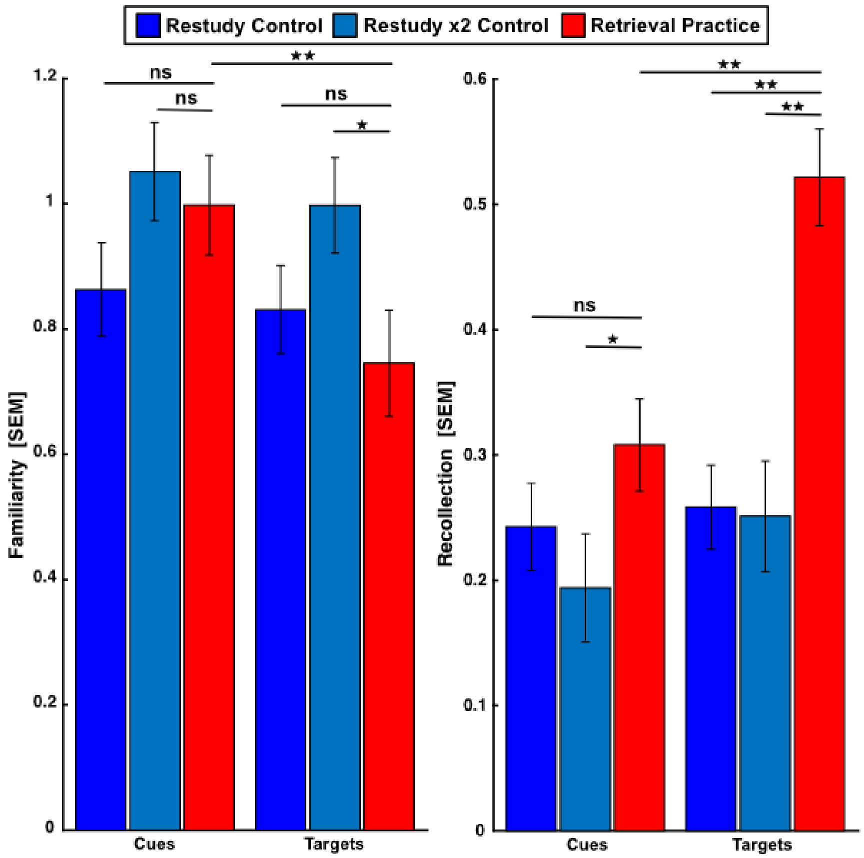


**Figure S2**

*Dual Process Signal Detection Model Results from Experiment 2*

*Note.* The two parameters from the dual process signal detection theory model depict familiarity (left) and recollection (right) in Experiment 2. Error bars depict SEM based on random permutation, whereas p-values are based on random shuffling procedures. *indicates $p < .05$; **indicates $p < .01$; ns = not significant.

# CHAPTER III. Functional MRI Correlates of Recall-Based Retrieval Practice Effects

Andrew T.J. Cawley-Bennett and Joseph R. Manns

Department of Psychology, Emory University

**Abstract**

The retrieval practice effect (also known as the testing effect) refers to the benefit to subsequent memory from intervening practice tests compared to restudying information. The effect has been studied extensively, yet the extent to which retrieval itself contributes to the effect was questioned in a recent study that used a cued-recall retrieval practice intervention (Cawley-Bennett et al., Chapter 2). The study found that, compared to a restudy control condition, the retrieval practice intervention improved subsequent memory for the to-be-recalled target words but also for the cue words that appeared on the screen during retrieval practice. The present fMRI study aimed to determine if the memory improvement for cue words and target words were supported by similar neural processes. Participants learned word pairs in either a cued-recall retrieval practice or a restudy control condition and completed a final recognition memory test in a scanner one day later for cue words and target words. Similar to the previous study, retrieval practice improved final test memory for both cue words and target words relative to the restudy control condition. In contrast, the fMRI results revealed more neural correlates of memory for retrieval practice target words compared to retrieval practice cue words in several brain regions, including the prefrontal cortex, medial temporal lobe, and medial occipital lobe. These findings suggest that the similar improvement in memory performance for cue words and targets words was supported by different neural processes.

**Functional MRI Correlates of Recall-Based Retrieval Practice Effects**

We tend to be better at remembering information that we actively tested ourselves on compared to information that we simply read twice. For example, if one were presented with the word pair “CRAYON-BARGE” and moments later were asked to complete “CRAYON-?????” from memory, one would later remember the to-be-recalled target word “BARGE” better on average than if one had read the word pair twice (termed restudied). This benefit to memory is termed the retrieval practice effect, or testing effect. The benefits of testing have been studied for over a century (e.g., Abott, 1909), though only in recent decades have studies incorporated better controlled experimental paradigms that include a restudy control condition to match exposure to the stimuli (e.g., Roediger & Karpicke, 2006). The memory benefits of retrieval practice relative to a restudy control condition have proven to be reliable and robust (Rowland, 2014).

Despite this recent progress in implementing a proper control condition, an important question that remains unanswered is to what degree factors other than memory recall during the intervention contribute to the retrieval practice effect, specifically when comparing it to a restudy control condition. In particular, a retrieval practice trial could elicit more participant engagement with the task as compared to a restudy control trial. This increased engagement could hypothetically include sharpened attention, increased vigilance, or deepened levels of processing, for example. Thus, a retrieval practice intervention could benefit subsequent memory relative to a restudy control due to memory recall only, these other factors related to engagement, or to both.

Cawley-Bennett and colleagues (Chapter 2) recently addressed this question in a study that used a cued-recall retrieval practice intervention, similar to the example above. Participants learned word pairs, with half of the word-pair stimuli presented in a retrieval practice condition and half of the word-pair stimuli presented in a restudy control condition. One to two days later, participants completed an old/new recognition memory test for the stimuli from both conditions. A novel aspect of the study was that only one word from each word pair appeared on the final test. Testing one word at a time allowed for a separate assessment of the effect of retrieval practice on the to-be-recalled target words and cue words. We suspected that the confound of deeper engagement might improve memory during retrieval practice for both cue and target words relative to the words from the control condition, even though the cue word from both conditions was always provided on the screen. Yet, retrieval practice target words might receive an additional memory improvement from both recall and increased engagement. The Cawley-Bennett et al. (Chapter 2) study found that, compared to restudy control condition, the retrieval practice intervention improved subsequent memory for cue words in addition to target words. This finding indicated that the retrieval practice condition differed from the restudy control condition by factors possibly other than memory recall. Despite the observation that retrieval practice improved subsequent memory accuracy for both targets and cues, additional findings from the prior study noted other differences on the final test between retrieval practice cue words and retrieval practice target words. These differences included a numerically larger effect size for target words relative to cue words and an increased likelihood of participants expressing high confidence in their correct memory judgments for target words as compared to cue words. Further analysis from a receiver operating characteristic (ROC) model revealed a modest yet consistent memory boost provided for retrieval practice cue words, whereas retrieval practice target words received a greater and more variable memory boost relative to words from the restudy control condition. Thus, the retrieval practice effect for target words appeared to be influenced by recall during the cued-recall intervention in addition to factors related to task engagement that also likely applied to cue words.

The present study sought to extend the findings of Cawley-Bennett et al. (Chapter 2) by using fMRI during the final memory test of a retrieval practice study to determine the extent to which the memory improvement for cue words and target words reflected similar or different neural processes. The overall rationale for this study was that patterns of brain activity might distinguish differing underlying cognitive processes even if accuracy from memory performance was similar between retrieval practice cues and retrieval practice targets. Moreover, a large body of functional imaging studies on the neural correlates of recognition memory broadly (Eichenbaum et al., 2007; Squire et al., 2007) and a small set of functional imaging studies on the neural correlates of retrieval practice specifically (van den Broek et al., 2016) are available to help frame the current study. The following section focuses on fMRI studies of retrieval practice.

Many neuroimaging studies investigating the effects of retrieval practice on memory have focused on neural activity during the retrieval practice intervention and have contrasted activity between retrieval practice and restudy control trials (e.g., Liu et al., 2014; Nelson et al., 2013; van den Broek et al., 2013; Vannest et al., 2012; Vestergren & Nyberg, 2014; Wing et al., 2013). These studies have typically found distinct neural engagement patterns during retrieval practice compared to during restudy, often involving posterior brain regions and the ventrolateral prefrontal cortex. These regions are presumed to contribute to strengthening neural representations and controlled retrieval processes, respectively. The anterior cingulate cortex and anterior insula have also been implicated for their involvement in attentional factors during retrieval practice (see van den Broek et al., 2016 for a review).

One limitation of scanning during the intervention is that it can exacerbate the task engagement differences between retrieval practice and restudy control conditions that are not specific to recall. For instance, if participants were to make overt memory responses during retrieval practice trials yet silently reread information during restudy control trials, overt behavior would differ and spuriously influence neural activity. Even if overt responses were made in both conditions, subtler differences, such as differences in response times, could similarly complicate the contrasts between retrieval practice and restudy control conditions. Thus, scanning during the intervention could make it harder rather than easier to identify retrieval-specific contributions to the retrieval practice effect on subsequent memory. An alternative approach is to scan neural activity during the final test, when all stimuli is presented in the same fashion and memory can be assessed identically. Relatively few studies have adopted this approach (e.g., Eriksson et al., 2011; Keresztes et al., 2014; Wirebring et al., 2015), and their findings are discussed next.

One fMRI study (Eriksson et al., 2011) investigated the memory effects from repeated cued-recall retrieval practice. The fMRI scan occurred one day after repeated study-test intervention trials of word pairs, and subsequent memory tests were also provided one week and five months after the fMRI scan. In the scanner, participants indicated their recall certainty for the target word when provided with the cue word, followed by a post-scan recall test for verification. The authors correlated the frequency of correctly retrieved trials during repeated retrieval practice the day prior with subsequent patterns of neural activity observed on the memory test one day later. The authors determined that the more times words were successfully recalled during encoding, the greater the amount of activity was observed in the anterior cingulate cortex one day later. In contrast, more successful recalls resulted in decreased activity in the mid-ventrolateral prefrontal cortex (PFC) and parietal cortex. The authors suggested that repeated testing reduced cognitive control in these specific brain regions, as linking the cue and target words became easier. Nonetheless, this study had limitations that limit the strength of the fMRI results. For instance, this study included a small sample size (n=12), which diminish the statistical power of the fMRI results. An additional limitation of the study included an inconsistent number of repeated test trials across participants. A few participants received only four repeated test trials per word pair whereas others received up to eight repeated trials. This confound of repeated test trials could have resulted in biased distinctions in brain activity observed during the one-day memory test. An additional limitation was the lack of a control condition to compare memory with the retrieval practice condition. Thus, it was unclear if good memory resulting from restudying the word pairs would have yielded similar neural activity patterns.

In another study (Wirebring et al., 2015), participants also learned word pairs through repeated cued-recall retrieval practice. The final test was conducted one week later using fMRI. In the scanner, participants mentally recalled the missing target word given the cue word and indicated their confidence in the correctness of their recall. They were then presented with four single letters and asked to select the second letter of the target word. The analysis focused on comparing patterns of brain activity between words that were both remembered during the intervention and also on the final test versus words only remembered during the intervention. The results revealed higher patterns of neural activity for those words remembered at both instances in several brain regions including the bilateral posterior parietal cortex (primarily the left inferior parietal cortex including the angular and supramarginal gyri), lateral temporal cortex, right medial temporal lobe (including the hippocampus), right occipital cortex, and bilateral frontal cortex (including the ventrolateral PFC, dorsolateral PFC, inferior and superior frontal gyri). Nevertheless, this study also lacked a suitable control condition. The lack of a control condition obscures the ability to assess whether good memory from restudy would elicit comparable neural activity patterns.

Lastly, Keresztes et al. (2014) compared patterns of brain activity for word pairs learned across a cued-recall retrieval practice and a restudy control condition. Corrective feedback, which is considered a memory modifier (Kang et al., 2007), was provided after each retrieval attempt. The fMRI final test was conducted immediately or one week later, with participants clicking a button only when they knew the missing target word when provided with the cue word. A post-scan test was then conducted for verification. An additional fMRI scan was conducted that included a *n*-back working memory task. The study's primary aim was to examine shared neural correlates between memory recall on the final test with the working memory task. Behavioral results revealed a retrieval practice effect only in the one-week group whereas the fMRI data did not reveal any significant differences in neural correlates of memory-related activity between conditions in either the immediate or one-week group when correcting for multiple comparisons. However, using a less stringent statistical threshold revealed differences between conditions for the one-week group with greater patterns of neural activity for the retrieval practice condition in the inferior frontal gyrus, medial frontal/anterior cingulate cortex, and occipital lobe. Additional analyses revealed overlapping patterns of neural activity between the retrieval practice condition and working memory tasks in the one-week group, while patterns of neural activity from both intervention conditions and the working memory task overlapped in the immediate group. Nonetheless, the findings from this study were also limited by the small sample sizes per group (n=13), which diminish the statistical power of the fMRI results. An additional limitation included the time differences in exposure between conditions in the intervention (i.e., eight seconds for retrieval practice trials and five seconds for restudy control trials). Differences in exposure times could have resulted in biased behavioral performance on the final test. In sum, these fMRI studies investigating retrieval practice offer limited insights into the neural correlates related to memory on a final test. Moreover, none of these studies depicted the neural correlates of memory-related activity for retrieval practice cue words, which is a goal of the present study.

The present study builds upon findings from Cawley-Bennett and colleagues (Chapter 2). One goal of this study was to replicate the behavioral findings of the previous study. While the research by Cawley-Bennett and colleagues (Chapter 2) demonstrated that memory factors related to recall was not necessary to observe retrieval practice effects for cue words, it remains unclear whether this behavioral similarity between retrieval practice cue and target words is supported by the same neural mechanisms. As such, the second goal of the present study was to determine if the retrieval practice effects for cue and target words reflected similar neural activity. Therefore, the present study examined neural activity related to cue and target words separately by comparing neural activity when remembering words learned in a retrieval practice condition or a restudy control condition during a final recognition memory test conducted inside an fMRI scanner. The behavioral results of this study replicated findings from Cawley-Bennett and colleagues (Chapter 2); we observed a retrieval practice effect for both cue and target words. However, the fMRI results revealed more widespread neural correlates of memory for retrieval practice target words relative to retrieval practice cue words, demonstrating that the memory processes for both words are different.

## Materials and Methods

**Participants**

Thirty participants were recruited from Emory University, with local Institutional Review Board approval. In order to be included, participants had to be native English speakers or use English as their primary language, at least 18 years of age to participate, and right-handed or right hand dominant. Right-handed participants were only included because brain regions involving language are considered to be lateralized to some degree and neural activity during episodic memory related tasks varies according to handedness in the medial temporal lobe (MTL) (Cuzzocreo et al., 2009). All participants were initially screened for prior history of head trauma (e.g., traumatic brain injury, concussion resulting in loss of consciousness for 30 seconds or longer) or cognitive impairments (e.g. Tourette syndrome, strokes, epilepsy, autism spectrum disorder). The *a-priori* sample size was determined based on previous fMRI retrieval practice investigations that included sample sizes ranging from 22 participants (Wiklund-Hörnqvist et al., 2017) to 35 participants (Hulbert & Norman, 2015). An initial sample size of 30 participants was determined to suffice, wherein potential exclusion of a few participants based on excessive head movement, systematic errors during scanning (such as computer synchronization errors), or attrition loss would be accounted for. Participants were required to complete two separate sessions and were compensated $50 after they completed both sessions. All participants were screened for eligibility and provided written consent prior to participation.

**Stimuli**

The stimuli were unrelated noun word-pairs taken from the MRC psycholinguistic database (Coltheart, 1981), with each word made up of five to seven letters. Additional factors, such as imageability, concreteness, and familiarity were also constrained to range from 330–600, 400–600, and 300–550, respectively (from the database’s scale of 100–700; see Appendix A for the list of words). Words were randomly paired together for each participant. The new words used as lures in the recognition memory final test in Session 2 were selected at random from the remaining list of unused words. An additional 20 words were taken from the word pool and used to create two additional follow-up immediate free recall memory tests. These words for the immediate free recall tests were never provided in any other portion of the study and each immediate free recall test was administered after the study related tasks. These immediate free recall tests were used as to assess participant engagement and in addition with final memory test and intervention performance, used as a basis for exclusion analysis.

**Procedure**

Administration of the task in both sessions was conducted in a private research laboratory on a computer using PsychoPy v2021.2.3 (http://psychopy.org; Peirce et al., 2019). Figure 1 provides a schematic of the procedures. Participants completed an initial study phase, an intervention phase, and a final memory test phase. Participants were initially instructed in Session 1 that they would learn and need to memorize many word pairs as their memory for the word-pairs would be later tested. Participants were then provided with instructions and two examples. One example demonstrated the procedural task for the retrieval practice condition and the other demonstrated the restudy control condition.

Once participants completed the instructions and examples, they then proceeded with the study phase. In the study phase, six randomly selected word-pairs were presented in a randomly chosen order for six seconds. Participants then viewed the same six word-pairs a second time in a new randomly chosen order for 12 seconds in the intervention phase. Half of the trials in the intervention were assigned to the restudy control condition and the remaining half were assigned to the retrieval practice condition. For retrieval practice trials, one word was provided as a cue and the other was replaced by an underlined blank line, termed the target. For restudy control trials, word-pairs were presented similarly to the initial study, with both words provided on the screen. To mimic a target word in the restudy control condition, one word was underlined and thus termed the control target word. The non-underlined word was termed the control cue word. In both conditions, participants were required to type in both words. The left versus right position of cue and target words was counterbalanced across trials for both retrieval practice and restudy control trials. No feedback was provided during the intervention. The six word-pairs presented during initial study and then immediately in the intervention made up one block of trials. Participants completed a total of 18 blocks. In total, participants studied 54 word-pairs in the retrieval practice condition and 54 word-pairs in the restudy control condition.

After twenty-four hours from the first session, participants returned to the MRI center to complete Session 2. In Session 2, participants were given old/new recognition memory tests in the scanner. Half of the trials in the recognition memory test were made up of old words, either the cue or target word from each word pair, and half of the trials were made up of new words. Prior to beginning the task, participants were first provided with instructions and nine examples of the recognition memory test and 10 baseline control trials. The exemplar recognition memory trials consisted of four words taken from the initial example words used in Session 1 (i.e., both words from the two word-pairs) intermixed with five new words. For each recognition memory trial, a single word was presented and participants provided one of six response options: Sure Old, Maybe Old, Guess Old, Guess New, Maybe New, or Sure New. Responses were provided in a single horizontal row. Counterbalancing of response options was applied with half of the participants being presented with responses as described from left to right, whereas the other half of participants were presented with response options in reverse going from left to right starting with Sure New. The presentation of Old and New recognition memory responses was counterbalanced across participants to prevent bias in brain activity in the primary motor and sensory cortices corresponding to the selection of responses. Participants selected one of six buttons from two separate MRI compatible button boxes (Current Designs, 8-button bimanual boxes; Winona, MN). Each recognition memory trial lasted five seconds.

Intermixed with the recognition memory trials were baseline control trials. The 10 baseline control examples were presented just after the recognition memory examples. These baseline control trials replaced passive resting state trials to prevent mind wandering (Stark & Squire, 2001). The baseline trials were designed to look similar to the recognition trials, but did not recruit long-term memory retrieval. In these baseline trials, a randomly selected single digit from one to six was presented between 1000 and 3750 milliseconds after the start of the trial. The presentation of the digit was shown briefly for either 100, 150, or 200 milliseconds. The onset presentation time and duration of the number was also randomly determined. Participants were to respond by pressing one of six buttons from two separate MRI compatible button boxes that corresponded to the number that participants viewed on the screen (i.e., 1 through 6 from left to right). Each baseline trial also lasted five seconds.

Once participants read the instructions and completed the example sets, participants then began the recognition memory tests. The recognition memory test was partitioned into four runs with each run beginning with a six second countdown to allow for MRI equilibration effects. Each run consisted of 27 new lures and 27 old words. The old words were evenly split (as best as possible) between the retrieval practice (13-14 trials) and restudy control conditions (13-14 trials). Only an individual cue or target word was presented from each old word pair. In total, 216 test trials were presented (108 old words and 108 new words). Intermixed with the recognition memory trials in each run were 27 baseline control trials. A blank inter-stimulus interval (ISI) of 500 milliseconds was provided between all trials. Based on prior simulations, we randomly intermixed half of the recognition and baseline trial-run set in the first half of the run and the remaining set in the second half. We applied two constraints to the presentation of trials: 1) each run should begin with a recognition trial after the six-second countdown and 2) no more than two baseline control trials were consecutively presented. After all of the recognition and baseline trials were completed in each run, two additional baseline control trials followed to allow for the tapering of the hemodynamic response from the last recognition memory trial.

Upon completing the tasks in the MRI, participants were removed from the scanner and were then provided with a demographic questionnaire, using a Qualtrics survey (Provo, UT). Once participants completed each session, participants completed an immediate free recall task to assess task engagement. In this task, participants studied a new list of 10 words, from the same word pool as the stimuli, presented simultaneously for 30 seconds on Qualtrics. Participants were instructed to remember the words from the list. Participants were then provided with three questions as distractors, followed by an immediate free recall test. During the test, participants were instructed to type in all the words they could recall from the initial word list within 90 secs. Once complete with all study requirements in Session 2, participants received monetary compensation for their participation.

**Behavioral Data Analyses**

Analyses were conducted using MATLAB version 2022b (MathWorks; Natick, MA). While the focus of the present study was to determine if retrieval practice effects for cue words reflected similar patterns of neural activity as retrieval practice effects for target words, we needed to assess behaviorally that there was a retrieval practice effect for cue and target words.

The three key analytic questions were accordingly: 1) was there a significant retrieval practice effect for cue words, 2) was there a significant retrieval practice effect for target words, and 3) is memory resulting from retrieval practice significantly different between cue words and target words? Thus, the behavioral analyses focused on three statistical comparisons: 1) retrieval practice cues versus restudy control cues, 2) retrieval practice targets versus restudy control targets, and 3) retrieval practice cues versus retrieval practice targets. These comparisons were evaluated with two-tailed paired t-tests, unless noted otherwise. Analysis of variance (ANOVA) was not used because neither the main effect of condition irrespective of word type nor the main effect of word type irrespective of condition would directly address the study's main questions. Moreover, a significant interaction term between condition and word type would be ambiguous as it could result, for example, from either a unique or disproportionate retrieval practice effect for targets relative to cues.

Performance on the final recognition memory test was calculated as a d-prime score (denoted as d'; Tanner & Swets, 1954), a standard metric of memory that subtracts a normalized false alarm rate (proportion of new words judged as old) from a normalized hit rate (proportion of old words judged as old). Non-responses were considered incorrect responses. For each participant, all new words contributed to the false alarm rate but separate hit rates, and thus separate d' scores, were calculated for retrieval practice cues, restudy control cues, retrieval practice targets, and restudy control targets. Instances of perfect hit rates (1.0) or false alarm rates (0), which would yield infinite normalized values, were adjusted by subtracting (hits) or adding (false alarms) one-half trial from the numerator prior to normalization.

A second set of analyses capitalizes on fitting receiver operator characteristic (ROC) models to the data from final recognition memory test performance, using the ROC Toolbox v.1.1.1 (Koen et al., 2017). Cumulative hit rates and false alarm rates were calculated across participants' confidence ratings, and curves were fit to these data using maximum likelihood estimation. The number of trials per old word type (27 trials) was insufficient to fit ROC models to each participant's data, and thus ROC models were fit to group data by summing cumulative judgment values across participants. All participant trials containing missing responses were substituted with a low confidence (guess) response. Determining the guess response as an old or new judgment was arbitrarily, yet evenly, selected. The ROC approach yielded estimates of memory performance that were similar to the d' scores described above in which memory is conceptualized in signal detection theory as the discriminability of a memory signal distribution of old words from a memory noise distribution of new words (Pastore & Scheirer, 1974). We will refer to this discriminability as signal strength. An additional advantage of fitting ROC models is that the approach can also provide an estimate of how the shape (standard deviation) of the signal distribution differs from that of the noise distribution (Egan, 1958). Thus, fitting ROC models to the data allows one to disambiguate whether increases in memory performance are due to relative increases in memory signal strength, decreases in signal standard deviation, or some combination of both. Accordingly, a commonly-used unequal variance signal detection model (UVSD; Mickes et al., 2007; Wixted, 2007) was fit to the ROC data to yield parameter estimates of both signal strength and the signal standard deviation (as a ratio of the standard deviation of the noise distribution).

After capturing both estimates for the UVSD model, a random bootstrap procedure was used to estimate variability across participants with respect to the ROC parameters estimates. Specifically, in each of the 1,000 bootstrap iterations, a random sample of n participants was drawn from the original n participants by random sampling with replacement. For each iteration, the UVSD ROC model was fit to the resampled data in the same manner as the original data. The standard errors of the mean for the parameters were calculated based on the standard deviations of the ROC parameter estimates across these 1,000 iterations.

An additional random shuffling procedure was used to estimate statistical significance for the original ROC parameter differences between retrieval practice and restudy control conditions for both cue and target words. In this approach, only experimental and control conditions were randomly shuffled using all original participants. No participant bootstrap resampling took place. Specifically, in each of the 1,000 iterations, the retrieval practice and restudy control labels were swapped for a randomly-selected subset of participants. In each iteration, ROC models were then fit to the shuffled data in the same manner as for the original data, and for each ROC parameter estimate an experimental-control difference was calculated. Original parameter estimates that fell in the outer $5^{th}$ percentile of these distributions was considered statistically significant at an alpha level of .05.

**Neuroimaging Acquisition**

Functional imaging was conducted using a 3T Siemens Trio scanner at the Facility for Education and Research in Neuroscience (FERN) at Emory University. All functional and anatomical images were acquired using a 32-channel head coil. A localizer scan was conducted prior to running a high resolution anatomical T1-weighted (MPRAGE) sagittal scan (1900 ms TR; 2.26 ms TE; 9° flip angle; 1 x 1 x 1 mm voxel resolution) for each participant to aid in spatial normalization to standard MNI-atlas space. Blood oxygen level dependent (BOLD) functional images were then acquired using a gradient-echo, echoplanar T2*-weighted pulse sequence (1200 ms TR; 33.2 ms TE; 45° flip angle; 2.5 x 2.5 x 2.5 mm voxel resolution; 3 multi-band acceleration factor).

All fMRI data were preprocessed using fMRIPrep pipeline 21.0.1 (Esteban et al., 2018; Esteban et al., 2019), which is based on Nipype 1.6.1 (Gorgolewski et al., 2011). fMRIPrep generates text regarding the preprocess pipeline to be implemented in the analyses section. Accordingly, the following unaltered text is provided per the request of the authors.

### *Anatomical Data Preprocessing*

A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1-weigted image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008, RRID:SCR_004757), and used as a T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 6.0.5.1:57b01774, RRID:SCR_002823, Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, RRID:SCR_001847, Fischl et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al., 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* [Fonov et al., 2009, RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

### *Functional Data Preprocessing*

For each of the 4 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1:57b01774, Jenkinson et al., 2002). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al., 2014) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al., 2002). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. (2007) in that instead of eroding the masks by two pixels on BOLD space, the aCompCor masks are subtracted a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer). After fMRIPrep pre-processing, further fMRI preprocessing was conducted using SPM12 software (http://fil.ion.ucl.ac.uk/spm/). Images after fMRIPrep were spatially smoothed to a six-millimeter gaussian kernel size using full width at half maximum (FWHM).

**fMRI Data Analyses**

Functional data analyses were conducted using SPM12, except where noted otherwise. First level models were created on a participant basis that included 29 nuisance regressors per run in a general linear model (GLM). These regressors included six head motion parameters (three translational and three rotational parameters), six temporal derivatives of motion parameters, six quadratic terms of motion parameters, six quadratic terms of temporal derivatives, and the five top-anatomical CompCor (aCompCor) based on the largest singular values. A high pass filter of 128 seconds was additionally applied to the GLM in each run to account for slow signal drifts. All trials were combined together (i.e., incorrect and correct trials), irrespective of confidence. Combining trials was done in an effort to increase statistical power and limit differences in patterns of neural activity related to memory strength between conditions for fMRI analyses. Response times were collected and utilized as trial duration in the study. This approach was adopted because the duration of the BOLD response for each trial is variable, persisting until a response is provided. Failing to account for reaction time can potentially disturb the accurate representation of BOLD-related activity (Grinband et al., 2008). Each individual participant's normalized-to-MNI-template anatomical brain, provided by fMRI prep (i.e., MNI152NLin2009cAsym_desc-brain_mask), was applied as an explicit brain mask. Model estimation was then calculated after establishing each participant's GLM. Following, one sample t-test contrasts were constructed on a participant basis to examine patterns of neural activity differences between the following conditions: a) retrieval practice cue words relative to new words, b) retrieval practice target words relative to new words, c) restudy control cue words relative to new words, d) restudy control target words relative to new words, e) retrieval practice cue words relative to restudy control cue words, f) retrieval practice target words relative to restudy control target words, and g) retrieval practice target words relative to retrieval practice cue words. Subsequent group level contrasts, using each participant's first level contrasts, were then constructed followed with an additional model estimation. Group contrast analyses were subsequently error corrected using statistical non-parametric mapping software (SnPM13.1.09; http://nisox.org/Software/SnPM13/; Nichols & Holmes, 2002) wherein a cluster-based approach with a corrected p-value of .05 and 10,000 bootstrap permutation iterations were applied. All group level contrasts that underwent error correction utilized an initial cluster-based threshold p-value of .001 for evaluation.

# Results

## Behavioral Results

Data were excluded for three participants based on: 1) excessive motion across more than one functional run as indicated by more than 25 framewise displacements (FD) detected per run (n=1) and 2) scores that were at least two standard deviations below the mean for either overall performance on the final recognition memory test, performance on the cued-recall intervention trials, or averaged performance across the two immediate free recall tests (n=2). Of the 27 remaining participants (age range = 19-29; n = 16 self-identified as female, n = 1 as a nonbinary gender), the mean proportion of correct responses for cued-recall intervention trials was 0.84 (SD = 0.16). Final test performance was evaluated to ensure discovery of retrieval practice effects for cue and target words as well as to replicate prior findings from Cawley-Bennett and colleagues (Chapter 2).

We first assessed how memory performance on the final recognition memory test differed as a function of word type (cue or target) and intervention condition (retrieval practice or restudy control). Figure 2 shows performance on the final recognition memory test, using d' (i.e., normalized hit rate minus the normalized false alarm rate; not based on ROC models). Three paired t-tests were used to assess comparisons between retrieval practice cues versus restudy control cues, retrieval practice targets versus restudy control targets, and retrieval practice cues versus retrieval practice targets. Direct contrasts between intervention conditions revealed significantly greater memory for both retrieval practice target ($t(26) = 2.67$, $p < .05$, $d = 0.51$) and cue words ($t(26) = 3.22$, $p < .01$, $d = 0.62$). There was, however, no differences in memory performance between retrieval practice cue words and retrieval practice target words ($t(26) = 0.004$, $p > .05$, $d = 0.0007$). Thus, we found a retrieval practice effect for cue and target words on the final recognition memory test. Furthermore, the direct comparison between retrieval practice cue and target words revealed that memory performance was nearly identical between both retrieval practice word types. These results parallel the findings from Chapter 2 demonstrating that retrieval practice improved subsequent memory for cue words in addition to target words.

To further explore memory performance differences on the final test based on word type and intervention condition, we analyzed response times. Measuring response times allows us to assess variations in decision-making cognitive processes. Figure 3 depicts the direct contrasts in response times for both words between intervention conditions. The results revealed significantly faster response times for retrieval practice target words relative to restudy control target words ($t(26) = 4.02$, $p < .001$, $d = 0.77$). There was, however, no significant differences in response times between retrieval practice cue words and restudy control cue words ($t(26) = 1.30$, $p > .05$, $d < 0.25$). A direct contrast between retrieval practice cue and target words also revealed significantly faster response times for target words ($t(26) = 4.38$, $p < .001$, $d = 0.84$). Thus, these results suggest that decision-making processes were significantly faster for retrieval practice target words relative to both words from the control condition and relative to retrieval practice cue words. To account for these observed differences, we also included response times as a covariate in all the fMRI analyses.

Final recognition memory test performance was also assessed by fitting the data to parameters from the unequal variance signal detection model (UVSD; Mickes et al., 2007). The model fits two parameters: the difference between distributions of memory strength for old words (signal) and new words (noise), referred to here as signal strength, and the signal standard deviation (as a ratio of the standard deviation from the noise distribution). An advantage of ROC models over the simpler estimate of d' is that the ROC model parameter estimates offer more empirical information about the shape of the memory distribution in addition to information about the strength of the memory (Brady et al., 2023). Figure 4 depicts the two parameters fit by the UVSD model separately for restudy control cue words, retrieval practice cue words, restudy control target words, and retrieval practice target words.

A random shuffling approach was used to ask if observed differences were greater than one would expect by chance. Relative to the restudy control condition, retrieval practice improved signal strength scores for both cue ($p < .01$) and target ($p < .001$) words. The signal standard deviation for retrieval practice target words relative to restudy control target words was significantly more variable ($p < .05$), though the standard deviation of the memory strength distribution for cue words between conditions was not ($p > .05$). When comparing the signal strength and the signal standard deviation between retrieval practice cue and target words, there was greater signal strength and more signal variability for target words ($p$'s $< .001$). These results, suggest that relative to restudy control, retrieval practice benefitted cue words with a smaller but more consistent memory boost and benefitted target words with a larger but more variable memory boost. These results further replicate findings from Chapter 2.

As another approach to asking how the influence of retrieval practice might differ between cue words and target words, we examined only correct final test trials for old items (hits) among the conditions. Figure 5 depicts the results from the proportion of high confident hits ("sure old") relative to all hits ("sure old", "maybe old", and "guess old") analysis. Correct high confidence responses were found to be significantly more frequent for retrieval practice cue ($t(26) = 2.48$, $p < .05$, $d = 0.48$) and target ($t(26) = 5.64$, $p < .001$, $d = 1.09$) words relative to restudy control words. Furthermore, there was a significantly greater proportion of high confidence ratings for retrieval practice target words relative to retrieval practice cue words ($t(26) = 3.20$, $p < .01$, $d = 0.62$). These results suggest that correct high confidence ratings were particularly more frequent for retrieval practice words relative to restudy control words, though the memory benefit was better for targets relative to cues. These results also further replicated previous findings from Chapter 2.

**fMRI Results**

***All Trials***

We first contrasted fMRI activity for new words and each of the four old word conditions to assess if patterns of neural activity from the contrasts were similar or dissimilar. The following fMRI results are based on direct contrasts between conditions when all trials were combined (i.e., incorrect and correct trials) and with all confidence responses included, to maximize statistical power and equate memory strength among all of the old word conditions. Figure 6 depicts all of the contrasts between old words and new words based on significant activity ($p <$ .05) after correcting for multiple comparisons using SnPM (Nichols & Holmes, 2002). Common patterns of neural activity that were greater from each of the old word conditions when compared to new words (e.g., retrieval practice cue words > new words) was found in the left inferior frontal cortex (IFC), the left lateral orbital frontal cortex (OFC), the left posterior cingulate cortex (PCC), the left precuneus primarily along the parieto-occipital sulcus, the left striatum (caudate nucleus and putamen), and the left angular gyrus. The posterior cingulate and precuneus are commonly associated with self-referential processes and subjective experiences of memory, related to memory introspection (Kim, 2010; Sestieri et al., 2010). The lateral inferior parietal cortex, particularly the angular and supramarginal gyri, have been implicated in attentional control, memory monitoring, and retrieval of contextual information (Cabeza et al., 2008; Vilberg & Rugg, 2009).

Moreover, similar patterns of neural activity for retrieval practice cue words, retrieval practice target words, and restudy control cue words were found in the left dorsal lateral prefrontal cortex (dlPFC), the left superior medial frontal gyrus (MFG; also known as the supplementary motor area), the left inferior temporal gyrus (ITG), and the right caudate nucleus. Similar patterns of neural activity in the left lateral occipital lobe, the right OFC, and right ventral lateral prefrontal cortex (vlPFC) was specifically found for restudy control cue words and retrieval practice target words when compared with new words. The dlPFC is linked to processes in retrieval monitoring, response inhibition, manipulation of retrieved information, or evaluation of memory representations (Ranganath & Blumenfeld, 2005; Simons & Spiers, 2003). The vlPFC is associated with retrieval of relevant information during memory tasks, particularly for contextually or semantically associated familiar items (Dobbins & Wagner, 2005; Wagner et al., 2001). Additionally, the inferior temporal gyrus is associated with the identification and discrimination of previously encountered items (Murray et al., 2007; Ranganath & D'Esposito, 2001).

Additional activity in the left ventral medial PFC (vmPFC), which includes the medial OFC, was found specifically for both retrieval practice cue and target words relative to new words. The vmPFC is associated in evaluating memory confidence during retrieval or assesses memory congruency (e.g., Brod & Shing, 2018; Hebscher & Gilboa, 2016; Moscovitch & Winocur, 2002). More pronounced and widespread patterns of neural activity were observed primarily for retrieval practice target words throughout many of the regions previously listed as well as additional activity found in the left medial temporal lobe (MTL), an area crucial in memory retrieval (Eichenbaum et al., 2007; Manns, 2017; Squire et al., 2007). Additional neural activity associated with retrieval practice target words was found in the left insula, the left fusiform gyrus (FFG), the middle and superior temporal gyri (MTG and STG; respectively), and bilateral activation in the medial occipital lobe such as the calcarine and lingual gyri. In sum, the contrasts between old word conditions and new words revealed that patterns of neural activity for old words was primarily left lateralized, consistent with some prior recognition memory studies involving verbal information (e.g., Yonelinas et al., 2005). Additionally, the old-new contrasts revealed more widespread patterns of neural activity for retrieval practice target words relative to retrieval practice cue words and both words from the restudy control condition.

We continued to contrast patterns of activity between old and new words by reversing the contrast and examining activity found to be significantly greater for new words relative to each of the old conditions (e.g., new words > retrieval practice cue words). Greater patterns of brain activity for new words was found predominantly in the right hemisphere and restricted primarily to the right inferior parietal cortex (including the supramarginal gyrus and angular gyrus; except when contrasted with restudy control cue words) and the right superior PFC (except when contrasted with restudy control target words). Uniquely significant neural activity for new words when contrasted with restudy control targets revealed additional patterns of brain activity in the right superior parietal lobe. Similarly, significant patterns of neural activity for new words was found in the right middle frontal gyrus when contrasted with restudy control cue and target words. These results suggest that patterns of neural activity for new words was predominantly lateralized to the right hemisphere. Additionally, sporadic patterns of neural activity were found when contrasted with different old word conditions. Yet, consistent neural activity appeared in the inferior parietal cortex and superior PFC.

We next performed a conjunction analysis to help visualize the similarities and dissimilarities in statistically significant BOLD activity for each of the above four contrasts of old words (retrieval practice cue words, retrieval practice target words, and restudy control cue words, and restudy control target words) relative to new words. These conjunction analyses were not meant to replace more formal direct contrasts between old word conditions (which are reported below) but instead were meant to help highlight trends in the data presented in Figure 6. Accordingly, Figure 7 (top) depicts patterns of brain activity unique to each contrast condition by masking out activity observed in more than one contrast. The contrast of retrieval practice target words versus new words yielded the most unique patterns of neural activity. Figure 7 (bottom) also depicts patterns of brain activity found to be common across all four contrasts by masking out activity that was not statistically significant for each of the four contrasts. Relatively few patterns of neural activity were statistically significant across all four contrasts. Thus, the results of contrasts between old words in each condition and new words indicated that old-new memory signals were relatively distinct for retrieval practice target words.

We next contrasted fMRI activity directly between only old word conditions. Similar to the comparisons for the memory performance data, we contrasted patterns of brain activity between retrieval practice cue words and restudy control cue words, between retrieval practice target words and restudy control target words, and between retrieval practice cue words and retrieval practice target words. Figure 8 depicts these additional contrasts based on significant neural activity ($p < .05$) found after multiple comparison correction using SnPM. Results from the direct contrast between retrieval practice and restudy control target words revealed distinct patterns of neural activity for retrieval practice targets in the left dlPFC, left vlPFC, left IFC, bilateral OFC, bilateral vmPFC, bilateral superior medial frontal, bilateral insula, left caudate nucleus, right putamen, left ITG, left MTG, bilateral STG, bilateral PCC, bilateral precuneus, bilateral medial parietal cortex, left MTL, as well as bilateral activation in the medial occipital lobe. No distinct patterns of neural activity were observed for restudy control target words. Similar results from contrasting between memory for retrieval practice targets and retrieval practice cues also revealed patterns of neural activity for retrieval practice targets in the left vlPFC, left dlPFC, left IFC, bilateral insula, left precuneus, left PCC, left MTL, bilateral hippocampus, bilateral FFG, bilateral ITG, left MTG, bilateral STG, areas of the right striatum such as the pallidum and putamen, as well as bilateral activation in the medial occipital lobe. Again, there were no distinct patterns of neural activity observed when for retrieval practice cue words. Furthermore, there were not any significant patterns of brain activity after multiple comparison correction from the direct contrast between retrieval practice cue words and restudy control cue words with all trials combined. In sum, these direct old word contrasts revealed more patterns of neural activity for retrieval practice target words relative to restudy control target words and retrieval practice cue words, similar to the analyses between old word conditions relative to new words.

***Correctly Answered Final Recognition Memory Test Trials Only***

The previous fMRI analyses focused on all old word trials, regardless of correctness, to increase statistical power and equate memory strength among all of the old word conditions. To obtain contrasts that depicted neural correlates specifically related to successfully remembering old words, we now conducted contrasts using correct trials only. Thus, all the contrasts that were previously conducted are now repeated with the following contrasts including correct only responses. Figure 9 depicts the results from the direct contrasts between old word conditions (i.e., retrieval practice cue and target words, restudy control cue and target words) relative to new words based on significant neural activity ($p < .05$) found after correcting for multiple comparisons using SnPM. Common neural correlates of memory found to be greater for each of the old word conditions when compared to new words (e.g., retrieval practice target words > new words) was found in the left IFC, left dlPFC, left OFC, left vlPFC, the left superior medial PFC, the left PCC, the left precuneus, the left angular gyrus, bilateral caudate nucleus, and the left lateral occipital lobe. Furthermore, similar patterns of neural activity for retrieval practice cue words, retrieval practice target words, and restudy control cue words were found in the left MTG. There was also greater MTL activity for restudy control cue words and retrieval practice target words. Analogous to contrasts with all trials, there were more neural correlates of memory-related activity for retrieval practice target words in all of the previously listed regions and with additional neural activity in the left ITG and STG, as well as bilateral activation in the medial occipital lobe in the calcarine and lingual gyri. Thus, these findings indicate that the contrasts between old word conditions and new words resulted in more neural correlates of memory for retrieval practice target words compared to retrieval practice cue words. Similarly, these contrasts also demonstrated more patterns of neural activity for retrieval practice target words relative to both words from the restudy control condition. The overall result of more neural correlates of memory-related activity observed for retrieval practice target words relative to each of the other old word conditions replicates the earlier old-new contrasts with all trials.

We continued to contrast patterns of neural activity between old and new words by reversing the contrast and examined patterns of activity found to be significantly greater for new words relative to each of the old conditions (e.g., new words > retrieval practice cue words). Patterns of neural activity for new words were found in the right inferior parietal cortex and the right superior PFC, as well as additional activity in the right dlPFC. Patterns of brain activity were also observed in the right precuneus for the new words condition relative to the restudy control target word condition and the retrieval practice cue word condition. Unique patterns of neural activity for new words relative the retrieval practice cue words were also found in the right MTG, the right ITG, and the right IFC. Once again, these results revealed that patterns of neural activity for correctly identifying new words were lateralized to the right hemisphere and were primarily observed in the inferior parietal cortex and superior PFC, similar to previous old-new contrasts with all trials. However, unlike with the previous analyses, we found additional consistent activity in the right dorsolateral PFC when correct only trials were considered.

We next performed the conjunction analysis to help visualize the differences and similarities in significant neural correlates of memory for each of the four old word contrasts relative to new words with correct only trials. Figure 10 (top) depicts these patterns of neural activity unique to each old word condition by masking out activity observed in more than a single contrast. Figure 10 (bottom) depicts common neural correlates related to memory across all four contrasts. Similar to the previous conjunction analyses with all trials, the results of the current conjunction analyses between each old word conditions and new words indicates that there were relatively few old-new memory signals, but that there were more distinct patterns of neural activity for retrieval practice target words.

We next directly contrasted fMRI activity between only old word conditions including retrieval practice cue words versus restudy control cue words, retrieval practice target words versus restudy control target words, and retrieval practice cue words versus retrieval practice target words using correct only trials. Figure 11 shows the results for the significant differences ($p < .05$) found after multiple comparison correction using SnPM. The results from the direct contrast between retrieval practice and restudy control cue words revealed greater patterns of neural activity only for restudy control cues located in bilateral occipital lobe, the right fusiform gyrus, the right ITG, right MTG, and the right insula. No patterns of brain activity were found for retrieval practice cue words. Results from the direct contrast between retrieval practice cue and target words revealed more neural correlates of memory-related activity for retrieval practice target words including bilateral activity in the OFC, the left dlPFC, the left IFC, the left vlPFC, the left superior PFC, bilateral activity in the vmPFC, bilateral activity in the medial occipital lobe, bilateral activation of the hippocampi but right MTL, bilateral activation of the FFG, and the right anterior-medial temporal pole. Additional patterns of neural activity was found in bilateral STG, left MTG, left ITG, bilateral activation of the insula, right thalamus, and bilateral activation of the striatum including the caudate nucleus and putamen, and the right pallidum. No neural correlates of memory were found for retrieval practice cue words. There was, however, no significant differences in patterns of neural activity between retrieval practice target words and restudy control target words. In sum, the results from the direct contrasts between old word conditions with correct only trials revealed more neural correlates of memory-related activity for restudy control cue words and, once again, more memory-related neural activity for retrieval practice target words.

Overall, direct contrasts between only the old word conditions were inconsistent. That is, when the direct contrasts between conditions for target words involved all trials, we found significant differences in patterns of neural activity between conditions. However, when the same contrasts involved correct only trials, we did not find any differences in neural activity. When the direct contrasts between conditions for cue words involved all trials, we did not find any differences in neural activity between conditions. However, when the same contrasts involved correct only trials, we did find differences in patterns of neural activity between conditions. Nonetheless, the contrasts between retrieval practice cue words and retrieval practice target words when including all trials or correct only trials both depicted more neural correlates of memory-related activity for retrieval practice target words.

## Discussion

The behavioral results from the current study revealed a retrieval practice effect for cue and target words. The memory accuracy (d') results for retrieval practice cue words resembled memory for retrieval practice target words. However, the memory improvement for retrieval practice target words varied more relative to retrieval practice cue words as revealed in the additional analyses. Specifically, the results from the UVSD model revealed that relative to restudy control words, retrieval practice cue words received a modest yet consistent memory improvement whereas retrieval practice target words received a larger and more variable memory improvement. These findings replicate those reported by Cawley-Bennett and colleagues (Chapter 2).

The memory benefits from the retrieval practice condition compared to the restudy control condition may be attributed to deeper engagement with the stimuli, yet we aimed to account for these engagement-related differences by contrasting retrieval practice target words with retrieval practice cue words. While similar memory accuracy was found for retrieval practice cue and target words, fMRI results revealed that retrieval practice led to dissimilar neural correlates of memory-related activity for target words relative to cue words. That dissimilarity was highlighted by analyses that compared old versus new memory signals across each condition and word type as well as by analyses that directly contrasted retrieval practice target words and retrieval practice cue words. Moreover, that dissimilarity was present regardless of whether the analyses included all trials or trials that were answered correctly on the final test. The following sections will discuss these fMRI findings in more detail with respect to contrasts specifically involving retrieval practice cue words and retrieval practice target words.

An initial set of analyses focused on the neural correlates of memory for each of the four types of old words (i.e., retrieval practice targets, restudy control targets, retrieval practice cues, and restudy control cues) by contrasting each separately with activity for new words. When these analyses were restricted to correct only trials, memory-related neural activity common to all four-word types were found in a few brain regions (Figure 10, bottom). These brain regions included the left inferior frontal cortex (IFC), the left dorsolateral prefrontal cortex (dlPFC), the left ventrolateral prefrontal cortex (vlPFC), left superior medial prefrontal cortex, left posterior cingulate cortex (PCC), the left precuneus, and the left angular gyrus. The left middle temporal gyrus also showed activity when remembering all of the old words except for restudy control target words. All of these brain areas were found in previous fMRI studies as typical neural sources involved in recognition memory (Henson et al., 2000; Henson et al., 2005; Horn et al., 2016; Wagner et al., 1998). However, when considering correct only trials in these old-new analyses, there were more patterns of memory related activity specifically for retrieval practice target words relative to the other old-new contrasts. Additional patterns of neural activity associated with retrieval practice target words included the left medial temporal lobe (MTL), areas in the left lateral temporal lobe (ITG, MTG, STG), and bilateral medial occipital lobe. Many of these aforementioned brain regions were also identified in one previous fMRI study on retrieval practice (Wirebring et al., 2015).

When similar old-new analyses were conducted with all trials, there were far fewer common patterns of neural activity found (Figure 7, bottom). Despite the reduction in shared patterns of neural activity in the old-new analyses using all trials, the left ventromedial prefrontal cortex (vmPFC) exhibited increased activity exclusively for retrieval practice cue and target words. However, activity in the vmPFC was not observed in any of the old-new analyses using correct only trials. Prior research has typically associated the vmPFC with confidence related decision-making processes (Hebscher & Gilboa, 2016; Moscovitch & Winocur, 2002). Given this finding, it is unclear why the vmPFC did not additionally show increased activity in these old-new analyses for restudy control words and in the old-new analyses including correct only trials. Nevertheless, the old-new contrasts with all trials also revealed more patterns of neural activity for retrieval practice target words relative to the other old-new contrasts. This finding complements the previous finding from the old-new contrasts with correct only trials.

An additional set of analyses contrasted patterns of neural activity between only old word conditions (i.e., retrieval practice cues versus restudy control cues, retrieval practice targets versus restudy control targets, retrieval practice cues versus retrieval practice targets). When restricting these contrasts between retrieval practice cue words and retrieval practice target words using either all or correct only trials, patterns of neural activity was solely identified for retrieval practice target words. No patterns of neural activity were observed for retrieval practice cue words. The patterns of neural activity for retrieval practice target words, whether including all or correct only trials, included the left dlPFC, the left vlPFC, bilateral insula, bilateral MTL, bilateral hippocampus, bilateral FFG, the left temporal lobe, and bilateral medial occipital lobe. These findings further support the notion that neural processes for retrieval practice target words involve more neural correlates of memory relative to retrieval practice cue words. Moreover, these fMRI results indicated that distinct memory-related processes were operating separately for retrieval practice cue words and retrieval practice target words. Similar to these fMRI results, the results from the additional behavioral analyses indicated that memory processes were dissimilar between retrieval practice cue words and retrieval practice target words.

Upon initial behavioral analysis using a memory accuracy metric (d'), it appeared that both retrieval practice cue and target words demonstrated a reliance on similar memory processes. However, further analyses from the UVSD model revealed a distinction in memory processes, indicating a more variable memory improvement for retrieval practice target words. Additional results from the response time analysis revealed faster decision-making processes for target words relative to cue words. Lastly, results from the high confidence hits analysis revealed a greater proportion of responses for retrieval practice target words relative to retrieval practice cue words. From both the behavioral and fMRI findings, we suggest that recall of target words during a retrieval practice intervention leads to both larger and more variable memory performance and instantiates more neural correlates related to memory processing relative to retrieval practice cue words.

However, our behavioral and fMRI findings did not always agree with one another. For instance, our behavioral data revealed a retrieval practice effect for cue words. Both memory accuracy and signal strength from the UVSD model depicted a stronger memory benefit for retrieval practice cue words relative to cue words from the restudy control condition. Furthermore, the proportion of hits given high confidence analysis also depicted a greater proportion of responses for retrieval practice cues relative to restudy control cues. Yet, we did not find any neural correlates of memory for retrieval practice cue words when compared to restudy control cue words. Rather, we only found neural correlates of memory for restudy control cue words. It is possible that the statistically significantly reduced neural activity in multiple brain regions for retrieval practice cue words relative to restudy control cues reflected more efficient processing and less mind wandering. Nevertheless, we can still connect the behavioral and fMRI findings for retrieval practice cue and target words with hypotheses explaining retrieval practice effects.

Cawley-Bennett and colleagues (Chapter 2) previously discussed three categories of hypotheses proposed to explain retrieval practice effects and attempted to connect their data with these hypotheses. The present study will expand on this idea by attempting to connect the current findings with those three categories of hypotheses. The first hypothesis was referred to as the effortful search hypothesis. This hypothesis proposed that the retrieval practice effect is driven by greater mental effort required to successfully recall information from memory (Bjork & Bjork, 1992; Kornell et al., 2011). According to this hypothesis, however, the retrieval practice effect would be observed only for the target words, as retrieval practice for cue words does not require any mental exertion to retrieve. Rather, the cue word is already present on the screen during the intervention, similar to both words from the restudy control condition. The fMRI results from the present study align with this hypothesis by revealing more neural correlates of memory-related activity for retrieval practice target words relative to retrieval practice cue words. However, the behavioral memory accuracy results do not support this hypothesis since we found a retrieval practice effect for both cue and target words from d'.

The second hypothesis was referred to as the elaborative association hypothesis. This hypothesis proposed that the experimenter-provided cue word triggers related semantic associations during memory search to help establish a link to the missing target information (Carpenter, 2009; 2011; Pyc & Rawson, 2012). This elaborated-on cue leads to the unidirectional retrieval of the target word (Carpenter & Yeung, 2017). An interpretation of this hypothesis would suggest that the benefit of retrieval practice would support better memory for the cue word because encountering the target word alone on a final memory test would not offer the opportunity to capitalize on the enhanced directional association. Therefore, the retrieval practice effect would greatly improve memory for cue words relative to target words. Nevertheless, neither the behavioral nor fMRI results from the present study provide supporting evidence for this hypothesis.

The third hypothesis was referred to as the episodic reactivation hypothesis. According to this hypothesis, retrieval practice attempts to reinstate the memory of the original study episode. This reinstatement attempts to retrieve prior items and other associated contextual information. If the original memory is reinstated, the memory representation is integrated with the retrieval practice event along with the restriction of the memory search process, thereby leading to subsequent memory improvements for both words (Karpicke et al., 2014; Lehman & Malmberg, 2013; Rowland & DeLosh, 2014). Based on this hypothesis, retrieval practice should equally benefit memory for both cue and target words since both words would be similarly reinstated. The behavioral results from the present study did find a retrieval practice effect for both words with similar memory accuracy results, providing supporting evidence for this hypothesis. However, findings from the UVSD model indicated that there was a stronger and more variable memory boost provided for retrieval practice target words relative to retrieval practice cue words. The fMRI results also revealed different neural correlates between retrieval practice cue and target words or that at least brain activity for retrieval practice target words could have been attributable to stronger memory representation than a different memory process. Irrespective of the interpretation of the fMRI results, the findings from the UVSD model and the fMRI findings make clear that there is not an equal memory benefit for both retrieval practice cue and target words as proposed by the episodic reactivation hypothesis.

In sum, while various hypotheses have been suggested to explain the mechanisms supporting retrieval practice effects, the current behavioral and fMRI findings do not strongly support any single hypothesis as initially proposed. Nevertheless, there is some evidence to support the effortful search hypothesis. Our fMRI findings revealed more neural correlates of memory for retrieval practice target words relative to retrieval practice cue words, which supports this idea. Additional evidence provided by the behavioral findings, such as faster decision-making responses, greater memory strength and more variability from the UVSD model, and a greater proportion of hits given high confidence responses all for retrieval practice target words relative to cue words also support this hypothesis.

One limitation of the present study was the inability to specifically investigate neural correlates of memory-related activity associated with only high confidence responses. This fMRI analysis would have paralleled the behavioral analysis that examined the proportion of high confidence responses across each old word condition. The limited availability of high confidence responses from multiple participants in the current study hindered our ability to perform the intended analysis using the fMRI data with adequate statistical power. Future research could address this limitation by reducing the number of trials, thereby enhancing memory for a smaller set of word pairs per condition and potentially enabling the analysis of high confidence responses.

A limiting factor in fMRI analyses, like those conducted in the present study, is the inability to examine corresponding activity in particular brain regions and reverse infer psychological processes (Poldrack, 2006). When attempting to draw definitive conclusions about the involvement of specific brain regions in cognitive processes, researchers should exercise caution and avoid oversimplifying findings post hoc based solely on observed patterns of neural activity. For this reason, the present study remains agnostic in defining differences in cognitive processes between retrieval practice cue and target words based on the observed patterns of neural activity. Future research may benefit from conducting a connectivity analysis, wherein investigators could better determine how brain regions communicate and interact with one another, therefore providing a more descriptive analysis.

Nonetheless, the primary motivation for using fMRI in the present study was to examine the extent to which the memory processes for retrieval practice cue and target words were similar or different. The behavioral findings from the present study revealed retrieval practice effects for cue and target words. However, the memory benefit was greater and more variable for retrieval practice target words relative to retrieval practice cue words, replicating findings from Cawley-Bennett and colleagues (Chapter 2). The fMRI results demonstrated different neural processes between retrieval practice cue and target words with more neural correlates of memory observed for target words.

**Acknowledgements**

We would like to thank Kate Revill, Solana Rivera, Kristina Dahlgren, and J. Imani Bunn for their contributions to the study design and/or data collection. Funding sources were obtained for this study from the Facility for Education and Research in Neuroscience (FERN) at Emory University. None of the authors have any other financial disclosures to provide.

# References

Abott, E. E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements*, *11*(1), 159-177.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26-41.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage*, *37*(1), 90-101.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. From *Learning processes to cognitive processes: Essays in honor of William K. Estes*, *2*, 35-67.

Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2023). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin & Review*, *30*(2), 421-449.

Brod, G., & Shing, Y. L. (2018). Specifying the role of the ventromedial prefrontal cortex in memory formation. *Neuropsychologia*, *111*, 8-15.

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: An attentional account. *Nature Reviews Neuroscience*, *9*(8), 613-625.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563-1569.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1547-1552.

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language*, *92*, 128-141.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497-505.

Cuzzocreo, J. L., Yassa, M. A., Verduzco, G., Honeycutt, N. A., Scott, D. J., & Bassett, S. S. (2009). Effect of handedness on fMRI activation in the medial temporal lobe during an auditory verbal memory task. *Human Brain Mapping*, *30*(4), 1271-1278.

Dobbins, I. G., & Wagner, A. D. (2005). Domain-general and domain-sensitive prefrontal mechanisms for recollecting events and detecting novelty. *Cerebral Cortex, 15*(11), 1768-1778.

Egan, J. P. (1958). Recognition memory and the operating characteristic. *USAF Operational Applications Laboratory Technical Note*.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci.*, *30*, 123-152.

Eriksson, J., Kalpouzos, G., & Nyberg, L. (2011). Rewiring the brain with repeated retrieval: A parametric fMRI study of the testing effect. *Neuroscience Letters*, *505*(1), 36-40.

Esteban, O., Blair., R., Markiewicz, C.J., Berleant, S.L., Moodie, C., Ma, F., Ilkay, A., et al. (2018). fMRIPrep 22.0.0rc3. *Software*.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., ... & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111-116.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, *9*(2), 195-207.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, (47), S102.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, 13.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*(1), 63-72.

Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., & Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *Neuroimage*, *43*(3), 509-520.

Hebscher, M., & Gilboa, A. (2016). A boost of confidence: The role of the ventromedial prefrontal cortex in memory, decision-making, and schemas. *Neuropsychologia*, *90*, 46-58.

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

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825-841.

Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4-5), 528-558.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In *Psychology of Learning and Motivation*, *61*, 237-284. Academic Press.

Keresztes, A., Kaiser, D., Kovács, G., & Racsmány, M. (2014). Testing promotes long-term learning via stabilizing activation patterns in a large network of brain areas. *Cerebral Cortex*, *24*(11), 3025-3035.

Kim, H. (2010). Dissociating the roles of the default-mode, dorsal, and ventral networks in episodic memory retrieval. NeuroImage, 50(4), 1648-1657.

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., ... & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLoS Computational Biology, 13*(2), e1005350.

Koen, J. D., Barrett, F. S., Harlow, I. M., & Yonelinas, A. P. (2017). The ROC Toolbox: A toolbox for analyzing receiver-operating characteristics derived from confidence ratings. *Behavior Research Methods*, *49*, 1399-1406.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85-97.

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, *1*(1), 76-85.

Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, *120*(1), 155-190.

Liu, X. L., Liang, P., Li, K., & Reder, L. M. (2014). Uncovering the neural mechanisms underlying learning from tests. *PLoS One, 9*(3), 1-7.

Manns, J.R. (2017). Neurobiology of recognition memory. In: J.L. Byrne (ed). *Learning and memory: A comprehensive reference*, 177-187. San Diego: Elsevier.

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review*, *14*, 858-865.

Moscovitch, M., & Winocur, G. (2002). The frontal cortex and working with memory. *Principles of Frontal Lobe Function*, 188-209.

Murray, L. J., Ranganath, C., & Knight, R. T. (2007). The effects of spatial attention on memory recognition: Behavioral and electrophysiological evidence. *Brain Research, 1184*, 255-269.

Nelson, S. M., Arnold, K. M., Gilmore, A. W., & McDermott, K. B. (2013). Neural signatures of test-potentiated learning in parietal cortex. *Journal of Neuroscience*, *33*(29), 11754-11762.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1-25.

Pastore, R. E., & Scheirer, C. J. (1974). Signal detection theory: Considerations for general application. *Psychological Bulletin*, *81*(12), 945-958.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195-203.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59-63.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage*, *84*, 320-341.

Pyc, M. A., & Rawson, K. A. (2012). Why is test–restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 737-746.

Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences, 9*(8), 374-380.

Ranganath, C., & D'Esposito, M. (2001). Medial temporal lobe activity associated with active maintenance of novel information. *Neuron, 31*(5), 865-873.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432-1464.

Rowland, C. A., & DeLosh, E. L. (2014). Benefits of testing for nontested information: Retrieval-induced facilitation of episodically bound material. *Psychonomic Bulletin & Review*, *21*, 1516-1523.

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., ... & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage*, *64*, 240-256.

Sestieri, C., Shulman, G. L., & Corbetta, M. (2010). Attention to memory and the environment: Functional specialization and dynamic competition in human posterior parietal cortex. *Journal of Neuroscience*, *30*(25), 8445-8456.

Simons, J. S., & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience*, *4*(8), 637-648.

Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience*, *8*(11), 872-883.

Stark, C. E., & Squire, L. R. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences*, *98*(22), 12760-12766.

Tanner, W., & Swets, J. (1954). The human use of information--I: Signal detection for the case of the signal known exactly. *Transactions of the IRE Professional Group on Information Theory*, *4*(4), 213-221.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310-1320.

Van den Broek, G. S., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *Neuroimage*, *78*, 94-102.

van den Broek, G., Takashima, A., Wiklund-Hörnqvist, C., Wirebring, L. K., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the "testing effect": A review. *Trends in Neuroscience and Education*, *5*(2), 52-66.

Vannest, J., Eaton, K. P., Henkel, D., Siegel, M., Tsevat, R. K., Allendorfer, J. B., ... & Szaflarski, J. P. (2012). Cortical correlates of self-generation in verbal paired associate learning. *Brain Research*, *1437*, 104-114.

Vestergren, P., & Nyberg, L. (2014). Testing alters brain activity during subsequent restudy: Evidence for test-potentiated encoding. *Trends in Neuroscience and Education*, *3*(2), 69-80.

Vilberg, K. L., & Rugg, M. D. (2009). Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. *Neuropsychologia*, *47*(13), 2691-2699.

Wagner, A. D., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Prefrontal cortex and recognition memory. Functional-MRI evidence for context-dependent retrieval processes. *Brain: A Journal of Neurology*, *121*(10), 1985-2002.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31*(2), 329-338.

Wiklund-Hörnqvist, C., Andersson, M., Jonsson, B., & Nyberg, L. (2017). Neural activations associated with feedback and retrieval success. *NPJ Science of Learning*, *2*(12), 1-7.

Wing, E. A., Marsh, E. J., & Cabeza, R. (2013). Neural correlates of retrieval-based memory enhancement: An fMRI study of the testing effect. *Neuropsychologia*, *51*(12), 2360-2370.

Wirebring, L. K., Wiklund-Hörnqvist, C., Eriksson, J., Andersson, M., Jonsson, B., & Nyberg, L. (2015). Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. *Journal of Neuroscience*, *35*(26), 9595-9602.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152-176.

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, *25*(11), 3002-3008.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45-57.

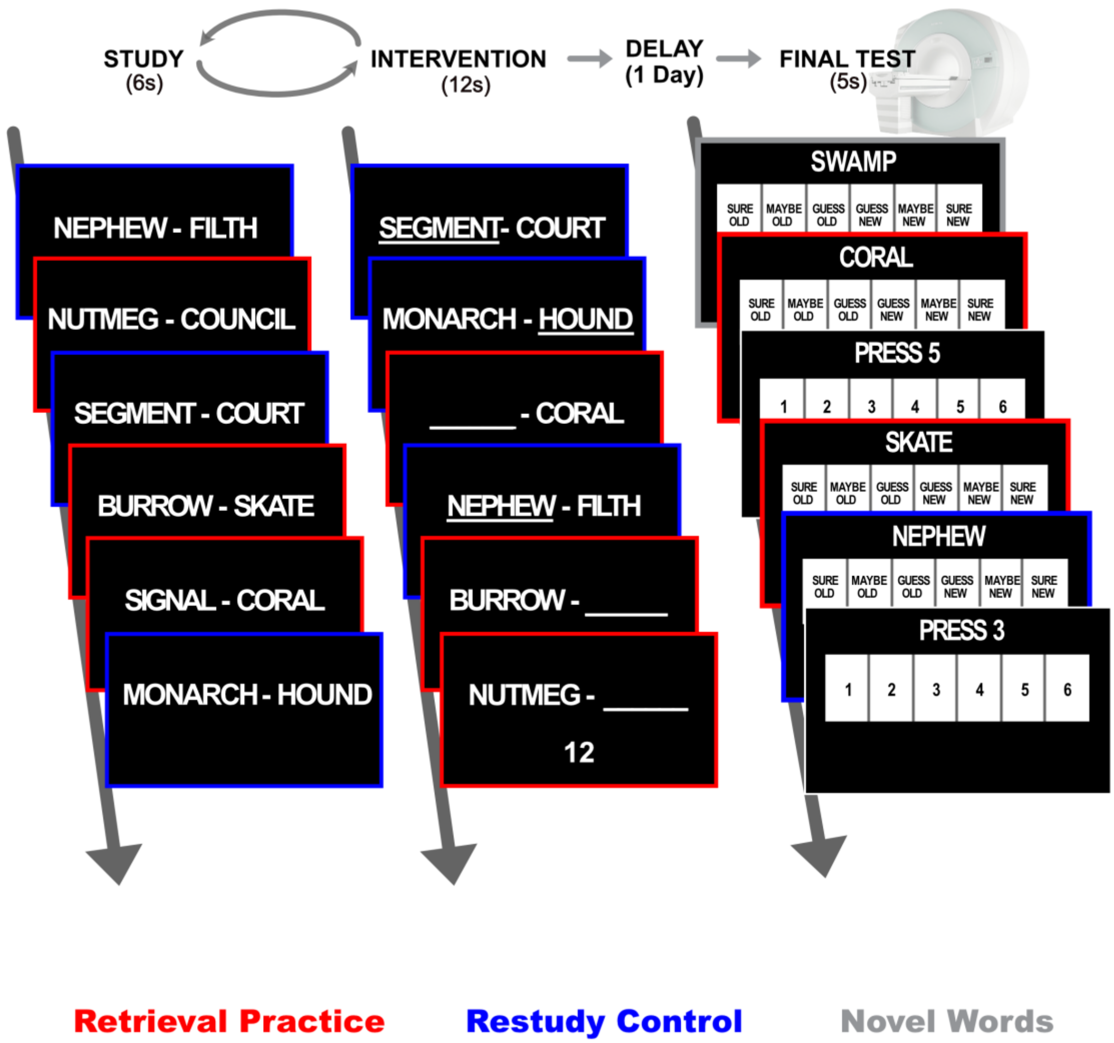


**Figure 1**

*Schematic of Task Procedures*

*Note.* Participants underwent three phases in the study: 1) a study phase, 2) an intervention phase, and 3) a final test phase. The six trial-blocks for initial study followed immediately by the intervention phase were repeated in total 18 times. The final test took place one-day later in the fMRI.

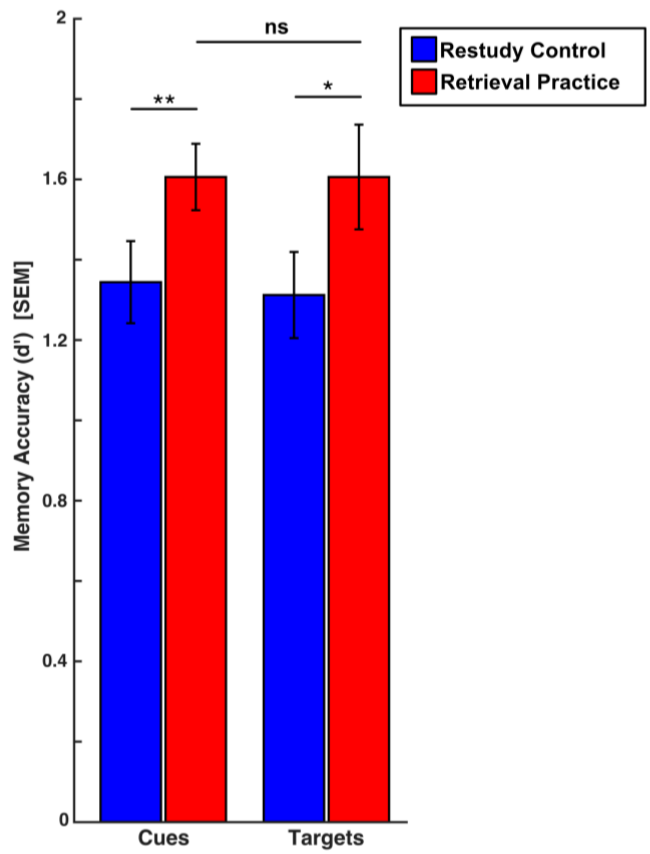


**Figure 2**

*Memory Accuracy Results*

*Note.* Results for cue and target words from both the retrieval practice (red) and restudy control (blue) conditions. Error bars depict SEM; *indicates p < .05, **indicates p < .01; ns = not significant.

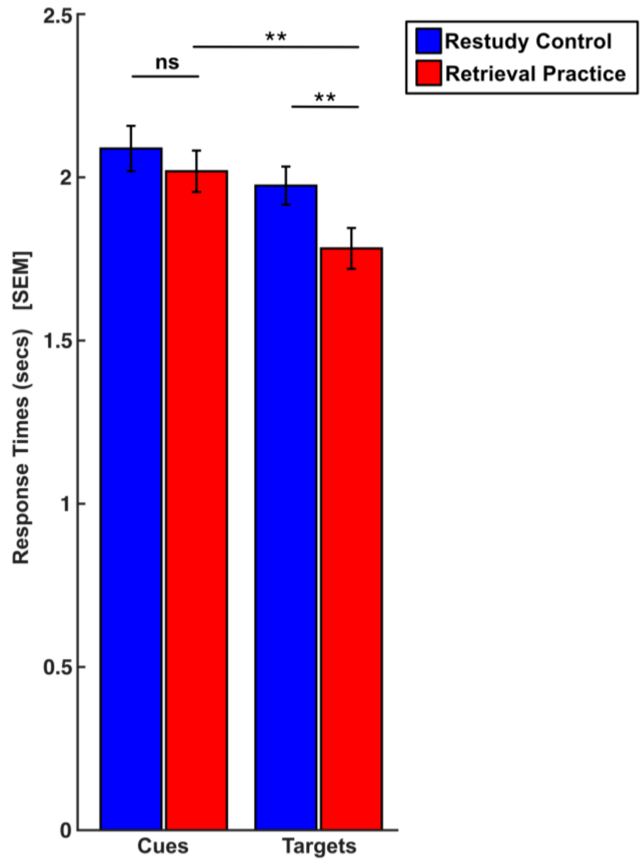


**Figure 3**

*Response Time Analysis*

*Note.* Error bars based on SEM; **indicates p < .01; ns = not significant.

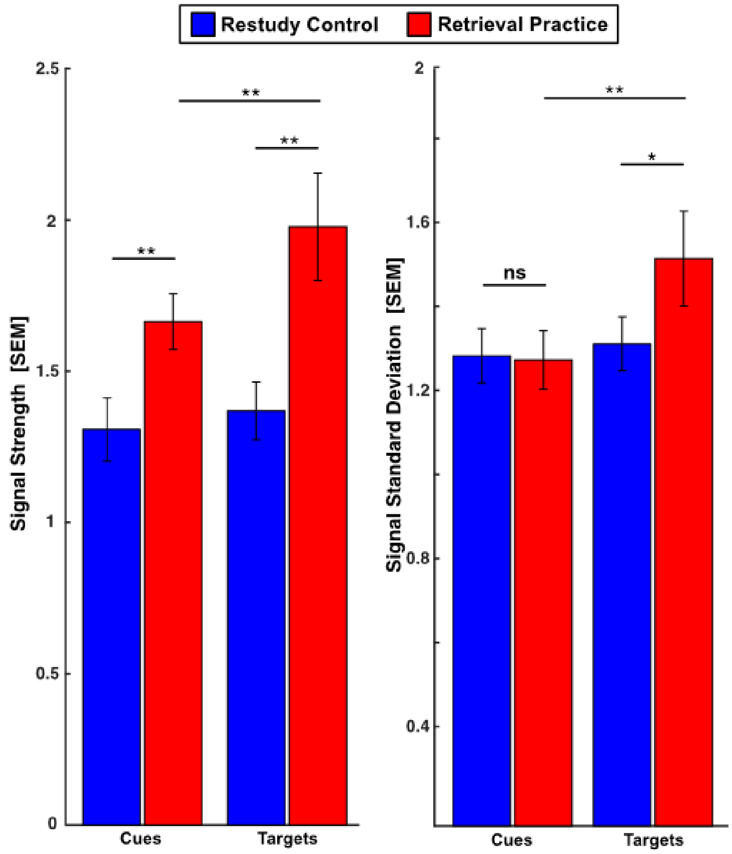


**Figure 4**

*Unequal Variance Signal Detection Model Results*

*Note.* Results from the two parameters of the unequal variance signal detection model: signal strength and the signal standard deviation. Error bars are based on SEM from 1,000 bootstrap iterations. P-values are based on 1,000 random shuffles by swapping n-sample conditions for cues and targets, then estimating the two parameters. *indicates $p < .05$, **indicates $p < .01$; ns = not significant.

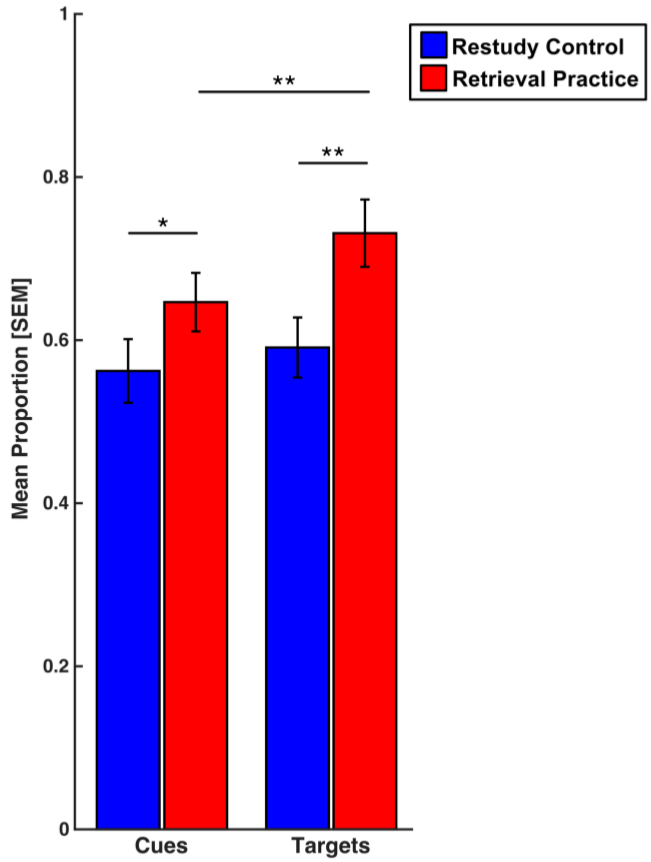


**Figure 5**

*Proportion of Hits Given High Confidence Results*

*Note.* Error bars based on SEM; *indicates p < .05, **indicates p < .01.

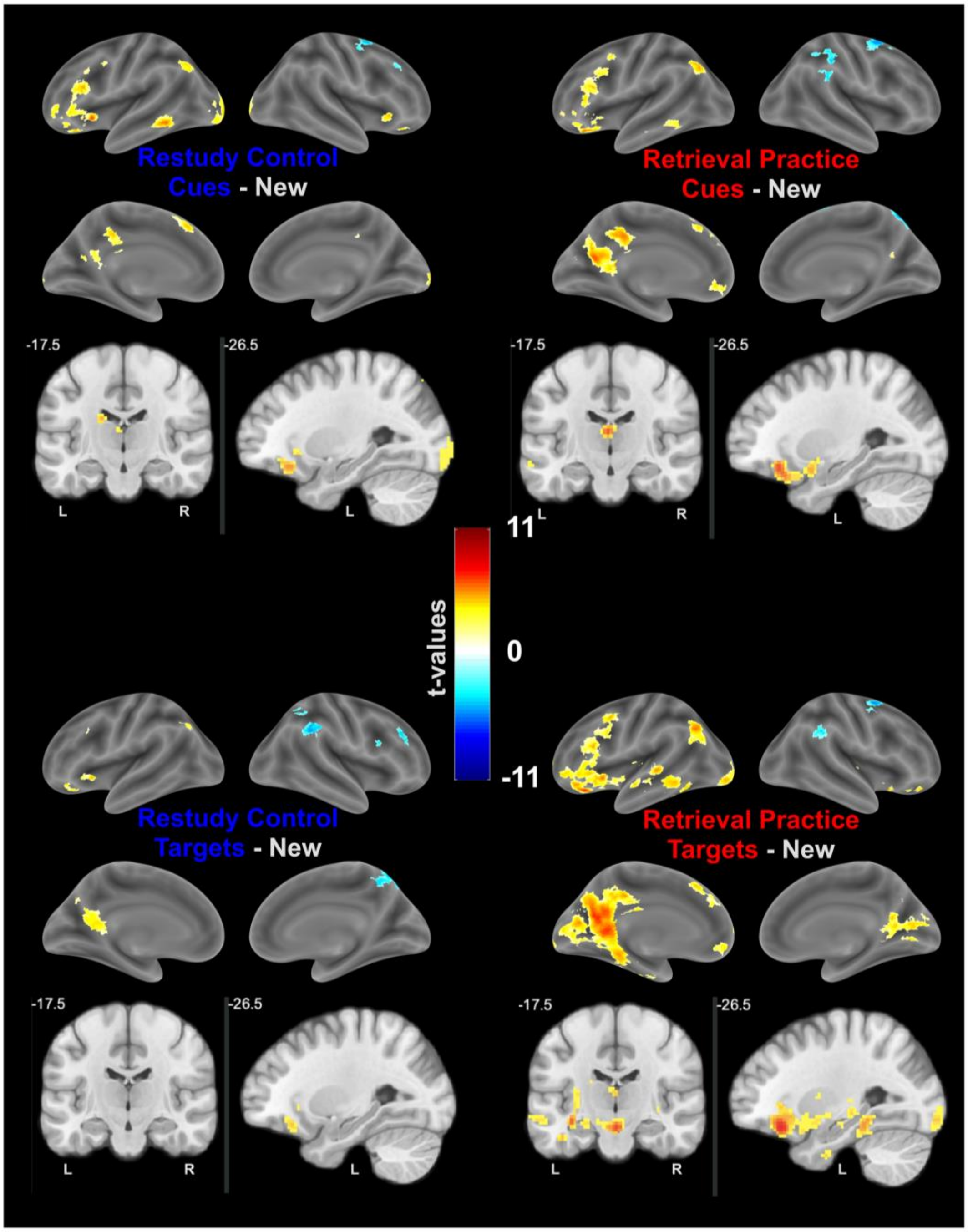


**Figure 6**

*Old Words versus New Words with Combined Trials*

*Note.* Contrasts of old word conditions (e.g., retrieval practice cue words) relative to the new word condition with all trials (i.e., incorrect and correct trials) after correcting for multiple comparisons with a $p < .05$.

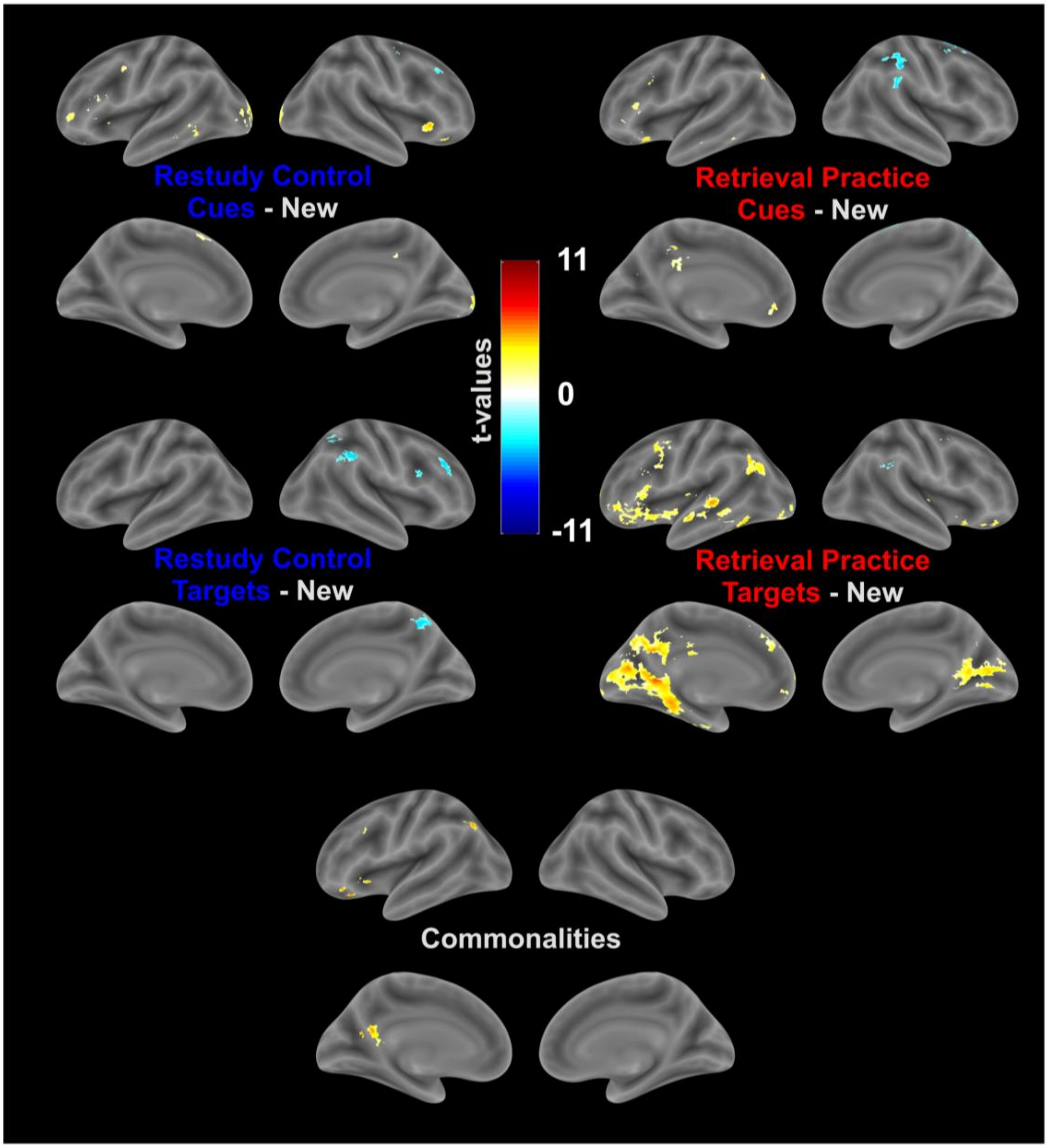


**Figure 7**

*Unique and Common Masks from Combined Trials*

*Note.* Unique and common (bottom) masks created from contrasts between old word conditions and new word condition, with activity unique to each condition and a single mask depicting activity observed across all old conditions relative to the new condition.

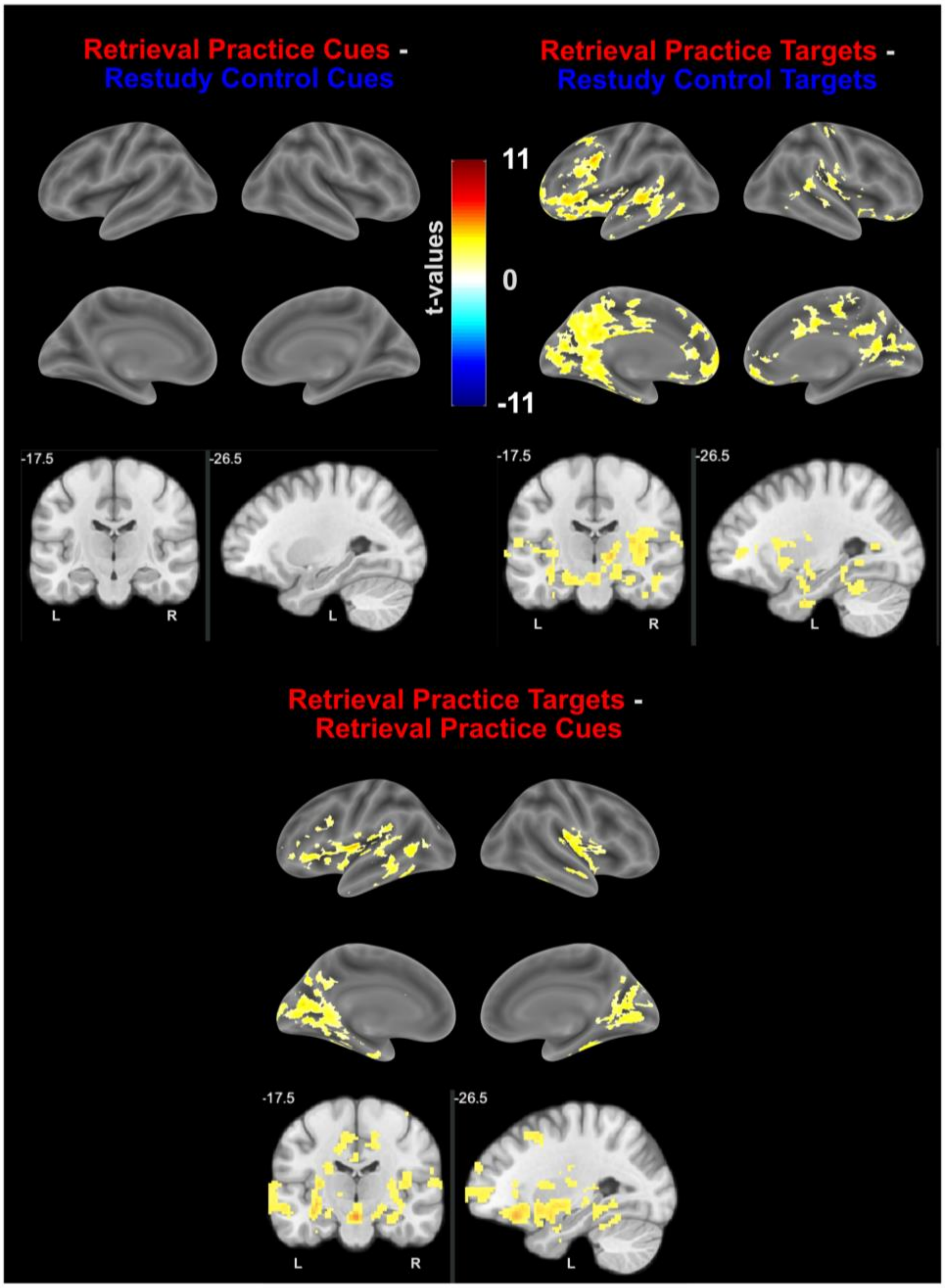


**Figure 8**

*Direct Contrasts of Old Word Conditions with Combined Trials*

*Note.* The results from direct contrasts between old conditions only with all trials after correcting for multiple comparisons using a $p < .05$.

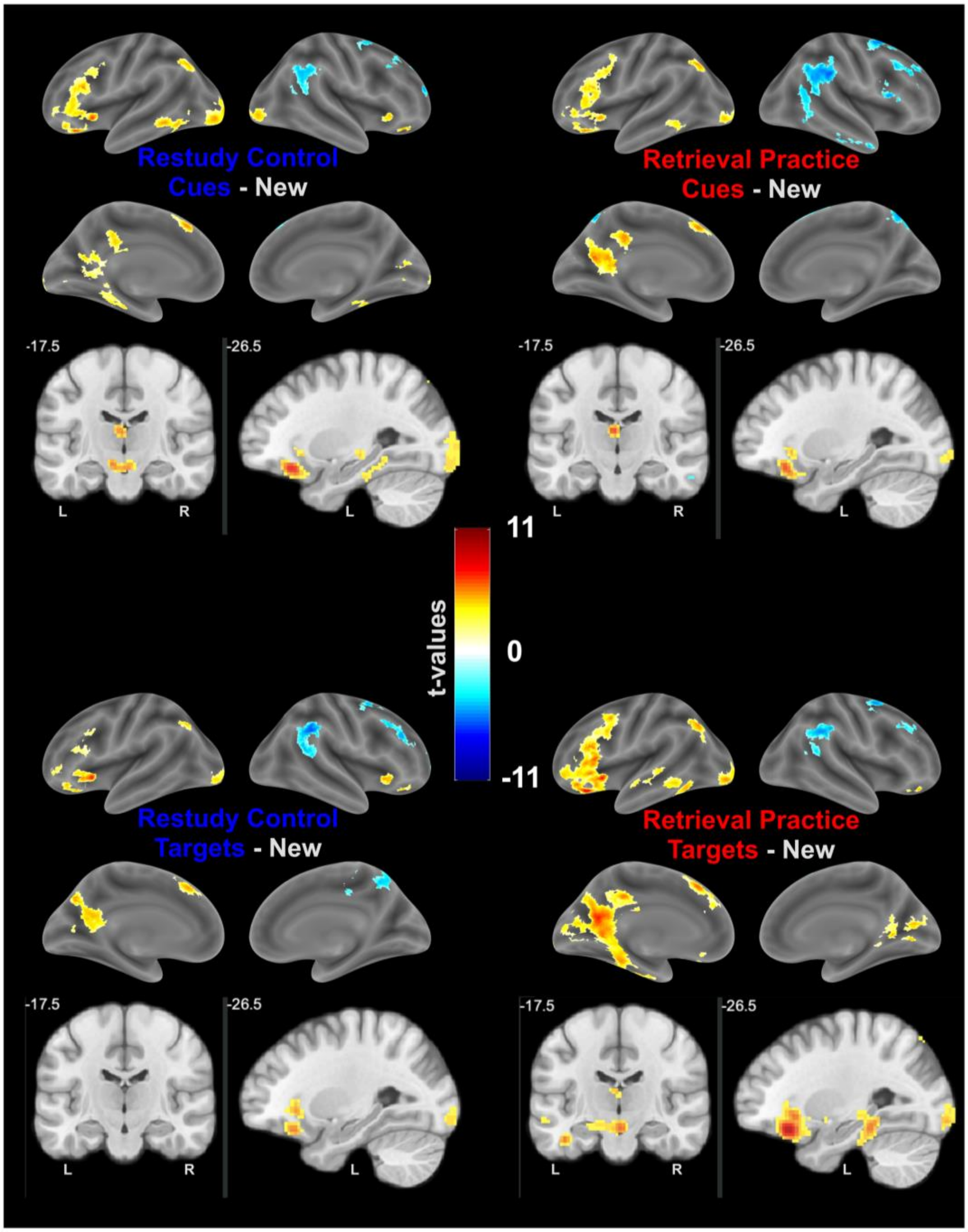


**Figure 9**

*Old Words versus New Words with Correct Only Trials*

*Note.* Contrasts between old word conditions relative to the new word condition using correct only trials after correcting for multiple comparisons with a p < .05.

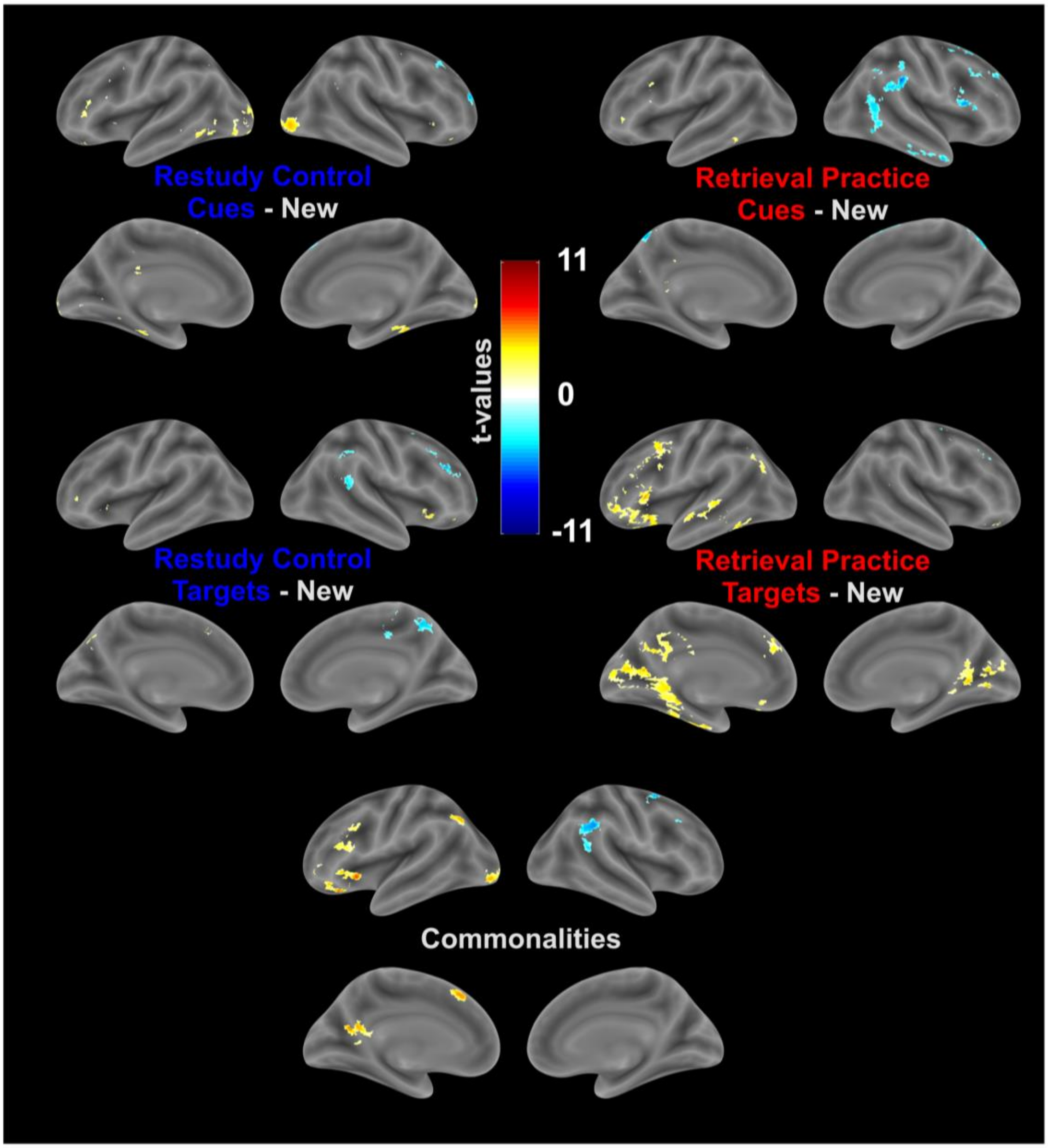


**Figure 10**

*Unique and Common Masks from Correct Only Trials*

*Note.* Unique and common (bottom) masks between old word conditions relative to the new word condition for correct only trials.

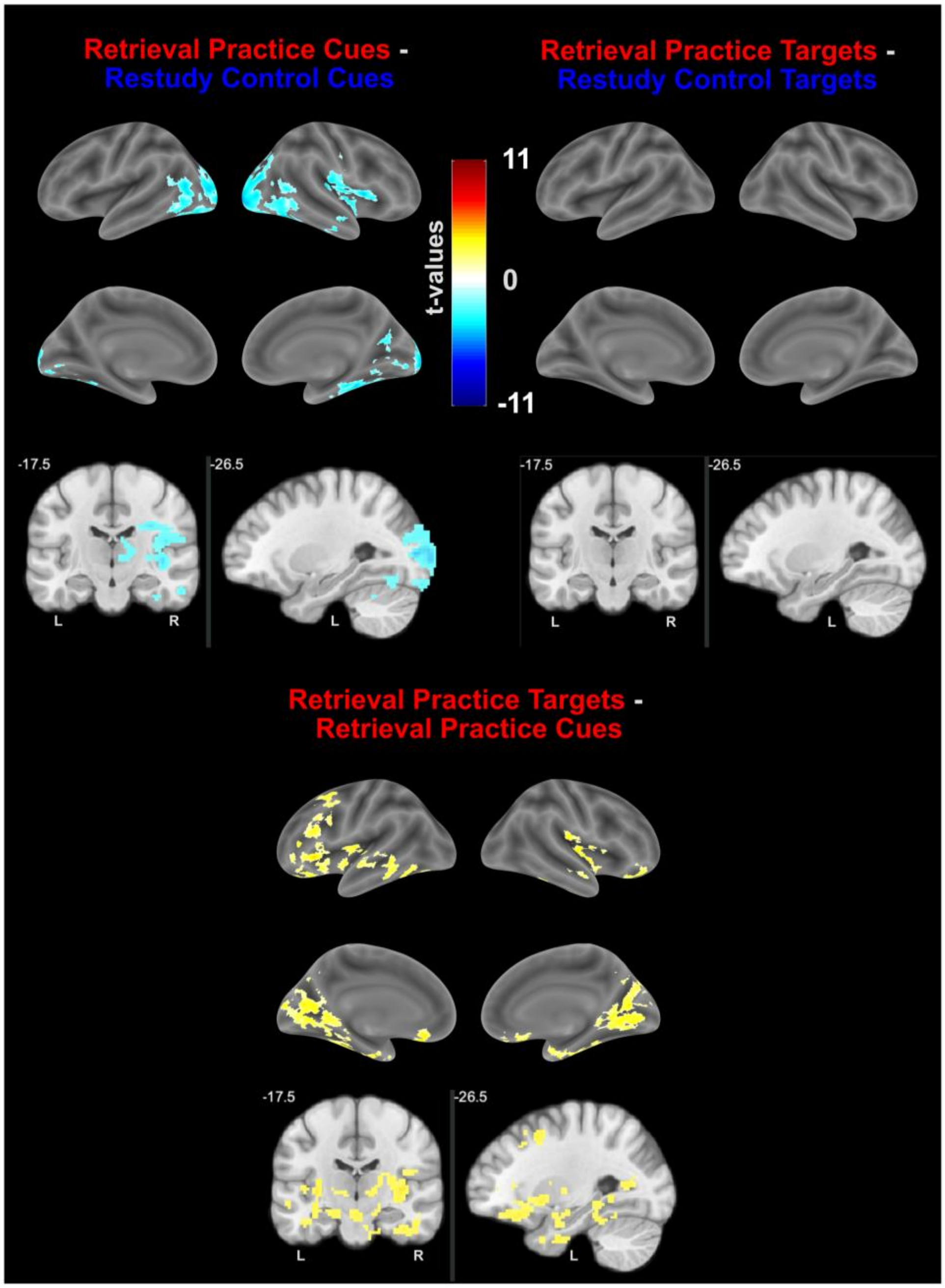


**Figure 11**

*Direct Contrasts of Old Word Conditions with Combined Trials*

*Note.* The results from direct contrast between old conditions for correct only trials after correcting for multiple comparisons using a $p < .05$.

# CHAPTER IV. GENERAL DISCUSSION

The overarching goal of this dissertation was to investigate the extent to which the retrieval practice effect depends upon retrieval. To achieve this goal, we carried out three experiments that explored the behavioral and neurological differences in memory for cue and target words learned across retrieval practice and restudy conditions. In the retrieval practice condition, participants were required to recall the target word from memory but were provided with the cue word. In the restudy control condition, participants were also provided with both the cue and target words. Using a final recognition memory test, we were able to compare participants' memory separately for cue and target words learned in both conditions. We hypothesized that if we found a retrieval practice effect for both cue and target words in the retrieval practice condition, then the effect would be partly attributable to deeper levels of engagement with the stimuli and not on factors related to recall alone. A second aim of this dissertation was to connect the results from these experiments to different hypotheses accounting for the mechanisms that support the retrieval practice effect. In this discussion, I will first summarize the findings across the three experiments carried out in Chapters 2 and 3. Then, I will connect those findings to the broader theoretical frameworks about the mechanisms supporting the retrieval practice effect.

In Chapter 2, we carried out two behavioral experiments. We measured retention of cue and target words learned in either the retrieval practice or the restudy control conditions using a final recognition memory test. The results from both experiments revealed a retrieval practice effect for both cue and target words. This finding made clear that retrieval practice effects do not depend on factors related to recall alone. Instead, the process involved in actively retrieving missing information can increase retention of all presented information, perhaps due to overall deeper task engagement. We additionally fit the data with two ROC signal detection models, the

UVSD and DPSD models, to elucidate differences in memory strength and variability between cue and target words. While both models determined different memory effects occurring for retrieval practice cue words and retrieval practice target words, I will focus specifically on findings from the UVSD model. The results from the UVSD model made clear that retrieval practice improved memory for target words more so than cue words. Specifically, retrieval practice target words received a *larger and more variable* memory improvement relative to restudy control target words. In a somewhat similar manner, the UVSD model depicted a *smaller yet consistent* memory boost for retrieval practice cue words relative to restudy control cue words. Additional results revealed a stronger correlation between intervention performance and final test performance for retrieval practice target words relative to retrieval practice cue words. We also measured response times during the final testing phase in Experiment 2 and found that participants employed faster decision-making processes when tested on retrieval practice target words compared to retrieval practice cue words. Lastly, we found that participants provided a higher proportion of correct high confident memory responses for retrieval practice target words compared to retrieval practice cue words. These findings from the two experiments collectively led us to conclude that the memory benefits for retrieval practice cue and target words were supported by different cognitive processes.

In Chapter 3, we aimed to determine whether the retrieval practice effect for both cue and target words were supported by similar neural processes. To accomplish this aim, we employed an additional approach, namely functional neuroimaging. Similar to the procedures in Chapter 2, we had participants learn word pairs in either a retrieval practice condition or a restudy control condition. However, unlike in Chapter 2, we administered the final recognition memory test in an MRI scanner. The behavioral results replicated findings from Chapter 2; we observed a retrieval practice effect for both cue and target words. Parallel to the findings from Chapter 2, there was also a greater memory advantage provided to retrieval practice target words compared to retrieval practice cue words. In addition to replicating these behavioral findings, the fMRI results revealed more neural correlates of memory-related activity for retrieval practice target words relative to retrieval practice cue words. Multiple patterns of neural activity seem to support retention of retrieval practice target words, including the left prefrontal cortex, the medial temporal lobe, the left lateral temporal cortex, and the medial occipital lobe. These findings suggest that the retrieval practice effects for cue and target words are supported by different neural processes.

Taken together, the findings from all three experiments indicated that recalling both cue and target words was not necessary to observe a retrieval practice effect. Instead, asking one to actively try and remember some missing information improved memory for both the to-be-recalled target and the provided cue information. However, our data suggest that the behaviorally similar memory benefits provided to cue and target words from retrieval practice can be attributed to differing cognitive and neuronal processes. These differing processes will be discussed next.

**Depth of Engagement Improving Memory**

We found that cue words learned in the retrieval practice condition were better remembered than those learned in the restudy condition. Since cue words from both conditions are provided during initial study and the intervention, one might expect that automatic recognition memory processes on the final test would be similar for cue words from both conditions. Because similar recognition memory performance on the final test was not found between conditions for cue words, we suspected that memory testing during a retrieval practice intervention impacted a participant's overall level of arousal, attention, or engagement with the tested material. Because cue words in the retrieval practice trials were presented on the screen like they were in the restudy trials, the memory differences for cue words between these two conditions must have relied upon deeper levels of processing or engagement with the tested stimuli (Craik & Lockhart, 1972; Craik & Tulving, 1975). Thus, we attributed the retrieval practice effect, primarily for cue words, as being a result of the depth of engagement involved during the retrieval process. Previous research on the generation effect helps support this interpretation.

**Parallels Between the Retrieval Practice Effect and the Generation Effect**

We did not consider the generation effect when developing our hypotheses, *a-priori* because our study design did not have a generation component. However, we cannot help but notice parallels between that effect and our observed results. In this paradigm, participants are presented with semantically-related word pairs (e.g., "MORNING-AFTERNOON") or a single word (e.g., "MORNING-???????"). When presented with a single word, the participant is required to generate a potential response word that is semantically related to the one provided. A subsequent memory test is then provided after a varying delay period, in which words that were generated were more consistently reported as being better remembered relative to words that were simply studied (for a review see Bertsch et al., 2007). This deeper level of engagement with the word pairs through active generation improves memory, an effect termed the generation effect (Clark, 1995; Jacoby, 1978).

Like the retrieval practice effect, the generation effect is also robust and found to occur whether using a within- or between-subject experimental design. The main contrast between generation effect paradigms and retrieval practice effect paradigms is that the latter paradigms require one to first study all the stimuli in an intact format. Only after initial study is complete is a more engaging intervention phase administered where the participant must actively remember a missing word or is once again presented with intact word pairs to restudy. In contrast, generation effect paradigms recruit deeper engagement with the self-generated stimuli during the initial study phase. This deeper level of engagement with the self-generated stimuli leads to memory improvements relative to less engaging activity with the provided stimuli. Therefore, we hypothesized that the observed retrieval practice effect for cue words and target words resulted from eliciting deeper overall engagement from participants during the retrieval practice intervention. However, as we have seen from the findings in the experiments in this dissertation, recalling the target word during the retrieval practice intervention tended to provide an additional and more variable memory benefit compared to retrieval practice cue words.

**Additional Factors Improving Memory for Retrieval Practice Target Words**

The finding of a retrieval practice effect for target words is not novel. Rather this effect is well known and recognized in the retrieval practice literature (e.g., Rowland, 2014; Yang et al., 2021). The novelty of our experiments, however, is that we established that factors related to memory recall is not solely necessary for a retrieval practice effect to occur. Nevertheless, the results from the three experiments revealed that the retrieval practice effects for cue and target words may not rely on similar memory processes. The differences in memory retention between retrieval practice target and cue words meant that there was an additional memory improvement for target words than from just the benefit of deeper engagement. Therefore, the retrieval practice effect for target words was supported by both deeper engagement with the stimuli and memory improvement from recall. In the following section, we will explore how our findings relate to or diverge from the major hypotheses concerning the mechanisms underlying the retrieval practice effect.

**Connecting our Findings with Prior Retrieval Practice Hypotheses**

A second aim of this dissertation was to connect the findings with pre-existing hypotheses that account for the mechanisms supporting retrieval practice effects. We will focus our discussion on the three categories of hypotheses referred to as the effortful retrieval hypothesis, the elaborative association hypothesis, and the episodic reactivation hypothesis. According to the effortful search hypothesis (e.g., Kornell et al., 2011), the retrieval practice effect should only benefit the target word since only the target word required greater mental effort or exertion to retrieve. In contrast, retrieval practice cue words did not require mental exertion since the words were already provided on the screen. Consequently, there would be no memory improvement for retrieval practice cue words. Prior research has typically been unable to clearly operationalize what mental effort implies. Although, one interpretation of mental effort was described as an indicator of the reprocessing of the memory trace that takes place during retrieval (Roediger & Butler, 2011). The effortful search hypothesis built upon earlier ideas related to the retrieval hypothesis (Bjork, 1975) wherein more difficult initial tests resulted in greater memory benefits (Bjork & Bjork, 1992; Halamish & Bjork, 2011). Nevertheless, because there was a retrieval practice effect that occurred for both target and cue words in all three experiments, we did not find evidence to support this hypothesis as initially proposed. In contrast to the behavioral findings, we did find evidence supporting this hypothesis from the fMRI results. Specifically, there were more neural correlates of memory-related activity for retrieval practice target words relative to retrieval practice cue words. Moreover, the greater cognitive memory benefit we observed for retrieval practice target words relative to cue words would also lend some credibility to this hypothesis. Nonetheless, according to the authors' interpretation of this hypothesis, there should not be a memory benefit provided to retrieval practice cue words, meaning we must rule out this hypothesis.

The elaborative association hypothesis (e.g., Carpenter, 2011; Pyc & Rawson, 2012) suggests a greater benefit for retrieval practice cue words relative to target words. This benefit is achieved by establishing associations with the provided cue word that then directly links to the missing target word. The premise is to connect the provided cue information with information that has already been acquired and established (i.e., prior knowledge). By building connections with prior knowledge, it helps one to retrieve the to-be-recalled target information. Furthermore, a provided cue word that decreases the accessibility to the target word leads to better memory by creating a greater need for the elaborative process (Carpenter, 2009; Carpenter & DeLosh, 2006). For instance, Carpenter (2009) found that memory was better for word pairs that bore a weak related semantic association (e.g., "BASKET-BREAD") relative to word pairs with a stronger related semantic association (e.g., "TOAST-BREAD"). It is from this elaboration-with-the-cue process that supports subsequent memory for word pairs because this process creates additional retrieval routes (Carpenter & DeLosh, 2006). One caveat to this hypothesis was that the benefit was provided in a unidirectional manner, stemming from the cue word to the associated information and then to the target word (Carpenter & Yeung, 2017). Consequently, the memory benefit focuses on retrieval practice cue words because encountering retrieval practice target words alone on a final test would not allow the opportunity to utilize the enhanced directional associations. Nevertheless, in our three experiments the behavioral results found similar, or near similar, retrieval practice effects in memory accuracy for cue and target words. In contrast, the

UVSD model and the fMRI results depicted a greater memory improvement for retrieval practice target words. Thus, the findings from this dissertation also do not support this hypothesis.

The episodic reactivation hypothesis (e.g., Rowland & DeLosh, 2014; Karpicke et al., 2014) posits that retrieval practice reactivates the initial study episode through mental time travel (Tulving, 2005). Should the initial study episode be retrieved using any provided cues, then that memory is integrated with the current intervention episode (Whiffen & Karpicke, 2017). The combination of these two episodes then benefits later memory search. This memory combination should lead to an equal retrieval practice effect for both cue and target words, given that both words are equally associated with reinstatement of the initial study episode (Whiffen & Karpicke, 2017). Our finding of a retrieval practice effect for both cue and target words from memory accuracy across all three experiments would support this hypothesis. However, our fMRI results do not support this hypothesis. The results from the UVSD models further suggest that there were differences in memory processes between retrieval practice words. Considering these findings, it appears that reinstatement of the initial study episode does not provide an equal memory benefit to both retrieval practice words, meaning we must also rule out this hypothesis.

**Revising Old Hypotheses Based on Current Findings**

The predictions from each of these three categories of hypotheses do not fully account for our behavioral and neural findings. The original authors of these ideas formulated their hypotheses based on studies investigating the impact of retrieval practice on memory for target words. Nonetheless, partial aspects of these hypotheses can be applied to the findings in this dissertation. Accordingly, I will now revisit, refine, and combine aspects of these hypotheses to create a theoretical framework around our findings.

We proposed that some, though not all, of the retrieval practice effect was supported by factors related to deeper levels of engagement. Intentional and conscious active searching or reflecting on prior knowledge or experience results in deeper engagement (Craik & Lockhart, 1972). Based on this framework, the memory search process required during retrieval practice, by itself, can trigger deeper levels of engagement with the tested stimuli. This process can be supported by ideas taken from all three hypotheses, wherein active engagement involves the amalgamation of greater effort, elaboration, and episodic reactivation. These ideas may not necessarily be mutually exclusive from one another (Carpenter & Yeung, 2017; Karpicke, 2017) but can be incorporated together. For instance, the elaborative association and the episodic reactivation hypotheses mention that the memory benefit of retrieval practice begins with a search-like process based on a provided cue. Although, additional processes continue based on either elaboration or episodic reactivation, both processes suggest a subsequent memory improvement based on this initial memory search. Accordingly, either hypothesis suggests one retains better memory of the cue word by connecting it with prior knowledge or experience. For instance, findings from Whiffen and Karpicke (Experiments 2 & 3, 2017) revealed that separate tasks involving elaboration and episodic reinstatement during a retrieval practice intervention led to near similar levels of final recall test performance. Further analyses revealed that these different tasks led to different patterns of recall organization. Thus, the recruitment of the cognitive process appeared dependent upon the task requirement during retrieval.

Furthermore, people will be more engaged and have better memory performance if the task recruits more mental effort to connect the provided cue information with prior knowledge or experience. Roediger and Butler (2011) suggested that retrieval effort during retrieval practice could also entail elaboration. Indeed, findings from research on the elaborative association hypothesis revealed that harder to retrieve information led to greater retrieval practice results (Carpenter & DeLosh, 2006; Carpenter, 2009). The authors suggested that this resulted because of the greater demand on the elaboration process and we suspect that our results stemmed from the more difficult tasks requiring greater depths of engagement with the stimuli. Accordingly, this deeper engagement can lead to increased attention, arousal, or vigilance. Therefore, this search process associated with retrieval recruits deeper engagement with the stimuli and thereby provides a memory boost to both the retrieval practice cue and target words. Like the episodic reactivation and elaborative association hypotheses, this idea would lead to a retrieval practice effect for both cue words and target words.

We proposed that the retrieval practice effect for target words was attributed to a combination of deeper engagement and memory recall. Because the target word must be overtly retrieved, it requires a greater amount of mental effort to recall during the intervention phase compared to the cue word. As such, retrieval provides an even greater memory boost for the item-specific information (Bjork, 1994). This interpretation is supported by aspects of the effortful search hypothesis. Additionally, this idea shares similarities to the elaborative association hypothesis, in which the greater difficulty needed to retrieve an item, the better the memory boost (Carpenter, 2009; Carpenter & DeLosh, 2006). This memory boost results in the reactivation of the memory trace but specifically focuses on the item-specific information (i.e., target word).

**Retrieval Practice Effects on a Molecular Basis**

The reactivation of a memory trace can reinforce the memory of an event or for specific information through a process known as reconsolidation. Prior research on reconsolidation detailing the molecular factors associated with memory retrieval and re-encoding in non-human animal models may help to explain our behavioral, cognitive, and neural findings (e.g., Szapiro et al., 2002; Dudai, 2004; Tronson & Taylor, 2007). Reconsolidation triggers a new round of *de novo* protein synthesis. This round of protein synthesis is associated with stabilizing the labile memory through synaptic consolidation (Dudai, 2004). Previous researchers have also suggested this reconsolidation process supports retrieval practice effects (Antony et al., 2017; Roediger & Karpicke, 2011). Moreover, this reconsolidation process may help reduce the rate of forgetting for retrieval practice cue and target words relative to restudy control words (Carpenter et al., 2008). However, the act of retrieving item-specific information (i.e., the target word), and if successful, would reinforce this reconsolidation process and result in even greater memory stabilization. This item-specific reconsolidation process differs from the memory trace reconsolidation process from deeper engagement, because the *de novo* protein synthesis during the reconsolidation process for the retrieved item-specific information would be more impactful but also more selective. Consequently, reconsolidation from recall would result in an even better reduced rate of forgetting for the retrieval practice target word. Thus, the memory benefit provided to retrieval practice target words is what led to the greater and more variable memory results observed across the three experiments in this dissertation. Given that we could not measure these factors in humans, future research would greatly benefit from distinguishing the reconsolidation process between these two proposed types of memory enhancements further in non-human animal models.

**Future Explorations of Memory Strength**

Experiment 2 attempted to explore the idea of increased memory strength. The results from that experiment demonstrated that memory performance from two restudy opportunities does not match performance from a single instance of retrieval practice. The findings made clear that relative to this additional restudy opportunity, retrieval practice provides a unique and disproportionate memory advantage. Moreover, the results from memory accuracy in Chapter 3 were nearly identical for retrieval practice cue and target words. Despite this result, the UVSD model and the fMRI results demonstrated clear difference between retrieval practice cue and target words. Thus, the results from all three experiments taken together are consistent with the idea that different processes support memory performance for retrieval practice cue and target words. Future research could explore further how memory strength from retrieval practice might qualitatively differ from memory strength from other means, such as three restudy opportunities.

**Conclusion**

In conclusion, retrieval practice effects are not solely produced from memory factors related to recall. Retrieval practice appears to prompt deeper engagement with the tested stimuli, a process that does not occur in the same manner during a restudy opportunity. However, the memory performance improvement for retrieval practice cue and target words are not supported by the same processes. Rather, recall during retrieval practice appears to also influence memory strength and variability. Findings from two behavioral experiments and one fMRI experiment made this point clear.

**Chapter I AND IV. References**

Abott, E. E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements*, *11*(1), 159-177.

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences*, *21*(8), 573-576.

Bäuml, K. H. T., & Trißl, L. (2022). Selective memory retrieval can revive forgotten memories. *Proceedings of the National Academy of Sciences*, *119*(8), 1-6.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & cognition*, *35*, 201-210.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In *J. Metcalfe, A. Shimamura, (Eds.), Metacognition: Knowing About Knowing*, *185*-205.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. From *Learning processes to cognitive processes: Essays in honor of William K. Estes*, *2*, 35-67.

Brame, C. J., & Biel, R. (2015). Test-enhanced learning: The potential for testing to promote greater learning in undergraduate science courses. *CBE—Life Sciences Education*, *14*(2), 1-12.

Buchin, Z. L., & Mulligan, N. W. (2017). The testing effect under divided attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(12), 1934-1947.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118-1133.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918-928.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563-1569.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1547-1552.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*, 268-276.

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin and Review*, *13*(5), 826-830.

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*(2), 438-448.

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language*, *92*, 128-141.

Chan, J. C. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, *61*(2), 153-170.

Chan, J.C. (2010). Long-term effects of testing on the recall of nontested materials. *Memory, 18*(1), 49-57.

Chan, J. C., McDermott, K. B., & Roediger III, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*(4), 553-571.

Cho, K. W., Neely, J. H., Crocco, S., & Vitrano, D. (2017). Testing enhances both encoding and retrieval for both tested and untested items. *Quarterly Journal of Experimental Psychology*, *70*(7), 1211-1235.

Clark, S. E. (1995). The generation effect and the modeling of associations in memory. *Memory & Cognition*, *23*(4), 442-455.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-428.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268-294.

Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology*, *21*(6), 919-940.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.*, *55*, 51-86.

Eriksson, J., Kalpouzos, G., & Nyberg, L. (2011). Rewiring the brain with repeated retrieval: A parametric fMRI study of the testing effect. *Neuroscience Letters*, *505*(1), 36-40.

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 801-812.

Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, *10*(2), 169-207.

Hinze, S. R., & Wiley, J. (2011). Testing the limits of testing effects using completion tests. *Memory, 19*(3), 290-304.

Hong, M. K., Polyn, S. M., & Fazio, L. K. (2019). Examining the episodic context account: Does retrieval practice enhance memory for context? *Cognitive Research: Principles and Implications*, *4*(1), 1-9.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269-299.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*(6), 649-667.

Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4-5), 528-558.

Karpicke, J. D. (2017). Retrieval-Based Learning: A Decade of Progress. *Learning and Memory: A Comprehensive Reference, 2,* 487-514.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772-775.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In *Psychology of Learning and Motivation*, *61*, 237-284. Academic Press.

Karpicke, J. D., & Roediger III, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151-162.

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, *67*(1), 17-29.

Keresztes, A., Kaiser, D., Kovács, G., & Racsmány, M. (2014). Testing promotes long-term learning via stabilizing activation patterns in a large network of brain areas. *Cerebral Cortex*, *24*(11), 3025-3035.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85-97.

LaPorte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. *Journal of Educational Psychology*, *67*(2), 259-266.

Lehman, M., & Karpicke, J. D. (2016). Elaborative retrieval: Do semantic mediators improve memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1573-1592.

Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, *120*(1), 155-190.

Liu, X. L., Liang, P., Li, K., & Reder, L. M. (2014). Uncovering the neural mechanisms underlying learning from tests. *PLoS One*, *9*(3), 1-7.

Lozito, J. P., & Mulligan, N. W. (2006). Exploring the role of attention during memory retrieval: Effects of semantic encoding and divided attention. *Memory & Cognition*, *34*, 986-998.

Malmberg, K. J., Lehman, M., Annis, J., Criss, A. H., & Shiffrin, R. M. (2014). Consequences of testing memory. *Psychology of Learning and Motivation*, *61*, 285-313**.**

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*(4-5), 494-513.

Mulligan, N. W., & Picklesimer, M. (2016). Attention and the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(6), 938-950.

Nielson, K. A., & Powless, M. (2007). Positive and negative sources of emotional arousal enhance long-term word-list retention when induced as long as 30 min after learning. *Neurobiology of Learning and Memory*, *88*(1), 40-47.

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, *74*(1), 18-22.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437-447.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*(6002), 335-335.

Pyc, M. A., & Rawson, K. A. (2012). Why is test–restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 737-746.

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.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255.

Roediger III, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. *Psychology of Learning and Motivation*, *55*, 1-36.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432-1464.

Rowland, C. A., & DeLosh, E. L. (2014). Benefits of testing for nontested information: Retrieval-induced facilitation of episodically bound material. *Psychonomic Bulletin & Review*, *21*, 1516-1523.

Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory*, *22*(7), 784-802.

Szapiro, G., Galante, J. M., Barros, D. M., Levi de Stein, M., Vianna, M. R., Izquierdo, L. A., ... & Medina, J. H. (2002). Molecular mechanisms of memory retrieval. *Neurochemical Research*, *27*, 1491-1498.

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*(4), 252-257.

Tronson, N.C., & Taylor, J.R. (2007). Molecular mechanisms of memory reconsolidation. *Nature Reviews Neuroscience*, *8*(4), 262-275.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human. *The Missing Link in Cognition: Origins of Self-reflective Consciousness*, 3-56.

van den Broek, G. S., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *Neuroimage*, *78*, 94-102.

van den Broek, G., Takashima, A., Wiklund-Hörnqvist, C., Wirebring, L. K., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the “testing effect”: A review. *Trends in Neuroscience and Education*, *5*(2), 52-66.

Vannest, J., Eaton, K. P., Henkel, D., Siegel, M., Tsevat, R. K., Allendorfer, J. B., ... & Szaflarski, J. P. (2012). Cortical correlates of self-generation in verbal paired associate learning. *Brain Research*, *1437*, 104-114.

Vestergren, P., & Nyberg, L. (2014). Testing alters brain activity during subsequent restudy: Evidence for test-potentiated encoding. *Trends in Neuroscience and Education*, *3*(2), 69-80.

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*(6), 571-580.

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.

Wiklund-Hörnqvist, C., Andersson, M., Jonsson, B., & Nyberg, L. (2017). Neural activations associated with feedback and retrieval success. *NPJ Science of Learning*, *2*(1), 12-19.

Wing, E. A., Marsh, E. J., & Cabeza, R. (2013). Neural correlates of retrieval-based memory enhancement: An fMRI study of the testing effect. *Neuropsychologia*, *51*(12), 2360-2370.

Wirebring, L. K., Wiklund-Hörnqvist, C., Eriksson, J., Andersson, M., Jonsson, B., & Nyberg, L. (2015). Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. *Journal of Neuroscience*, *35*(26), 9595-9602.

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399-436.

**Appendix A**

| Word | Concreteness | Familiarity | Imageability | Num. Letters |
|---|---|---|---|---|
| ACROBAT | 566 | 431 | 583 | 7 |
| ADMIRAL | 528 | 436 | 543 | 7 |
| AGENCY | 426 | 420 | 366 | 6 |
| AIRSHIP | 585 | 332 | 545 | 7 |
| AISLE | 509 | 503 | 528 | 5 |
| ALGAE | 545 | 317 | 424 | 5 |
| ALGEBRA | 453 | 483 | 510 | 7 |
| ALIMONY | 454 | 415 | 456 | 7 |
| ALLEY | 557 | 526 | 572 | 5 |
| AMMONIA | 563 | 524 | 536 | 7 |
| ANCHOR | 595 | 458 | 561 | 6 |
| ANGLE | 467 | 518 | 503 | 5 |
| ANTIQUE | 492 | 484 | 549 | 7 |
| ARTICLE | 479 | 533 | 421 | 7 |
| ARTIST | 554 | 547 | 600 | 6 |
| ASPIRIN | 574 | 546 | 542 | 7 |
| ATHLETE | 545 | 482 | 591 | 7 |
| ATTACK | 411 | 542 | 501 | 6 |
| AVENUE | 539 | 529 | 564 | 6 |
| BADGE | 561 | 473 | 519 | 5 |
| BALLOT | 455 | 453 | 437 | 6 |

| | | | | |
|---|---|---|---|---|
| BANDIT | 547 | 388 | 562 | 6 |
| BANKER | 547 | 524 | 565 | 6 |
| BANNER | 567 | 381 | 569 | 6 |
| BATON | 570 | 424 | 516 | 5 |
| BATTLE | 564 | 537 | 597 | 6 |
| BEAST | 564 | 456 | 558 | 5 |
| BIBLE | 585 | 496 | 594 | 5 |
| BIOLOGY | 400 | 526 | 441 | 7 |
| BIRTH | 471 | 527 | 532 | 5 |
| BISCUIT | 574 | 521 | 571 | 7 |
| BISHOP | 587 | 467 | 524 | 6 |
| BLADE | 584 | 517 | 568 | 5 |
| BLIND | 443 | 531 | 485 | 5 |
| BLOCK | 558 | 544 | 483 | 5 |
| BOARD | 565 | 546 | 592 | 5 |
| BOATING | 494 | 519 | 533 | 7 |
| BOOTH | 556 | 444 | 486 | 5 |
| BORDER | 444 | 489 | 453 | 6 |
| BOUQUET | 566 | 473 | 599 | 7 |
| BRAWL | 456 | 418 | 485 | 5 |
| BRISKET | 456 | 364 | 370 | 7 |
| BRISTLE | 558 | 461 | 562 | 7 |
| BROWNIE | 535 | 386 | 553 | 7 |

| | | | | |
|---|---|---|---|---|
| BRUTE | 462 | 430 | 481 | 5 |
| BUCKET | 594 | 506 | 586 | 6 |
| BUCKLE | 568 | 474 | 587 | 6 |
| BUFFER | 509 | 402 | 459 | 6 |
| BUREAU | 547 | 395 | 497 | 6 |
| BURLAP | 583 | 412 | 488 | 6 |
| BURNER | 500 | 518 | 488 | 6 |
| BURROW | 426 | 421 | 444 | 6 |
| BUTCHER | 556 | 473 | 596 | 7 |
| CABIN | 596 | 523 | 582 | 5 |
| CABINET | 593 | 472 | 524 | 7 |
| CABLE | 544 | 492 | 469 | 5 |
| CAMEL | 597 | 421 | 561 | 5 |
| CANDLE | 565 | 544 | 594 | 6 |
| CANON | 541 | 424 | 493 | 5 |
| CANTEEN | 587 | 490 | 540 | 7 |
| CAPITAL | 448 | 538 | 518 | 7 |
| CAPITOL | 529 | 529 | 545 | 7 |
| CAPSULE | 540 | 505 | 594 | 7 |
| CAPTAIN | 534 | 498 | 497 | 7 |
| CAPTIVE | 516 | 415 | 518 | 7 |
| CARAVAN | 539 | 304 | 562 | 7 |
| CARBON | 511 | 470 | 445 | 6 |

| | | | | |
|---|---|---|---|---|
| CAROL | 535 | 489 | 499 | 5 |
| CARPET | 581 | 508 | 538 | 6 |
| CAVERN | 534 | 400 | 548 | 6 |
| CELLAR | 572 | 467 | 572 | 6 |
| CHAIN | 595 | 513 | 559 | 5 |
| CHANNEL | 527 | 482 | 508 | 7 |
| CHAPEL | 587 | 471 | 560 | 6 |
| CHART | 532 | 533 | 531 | 5 |
| CHASSIS | 561 | 327 | 386 | 7 |
| CHEEK | 565 | 533 | 561 | 5 |
| CHEST | 580 | 543 | 556 | 5 |
| CHIEF | 503 | 482 | 545 | 5 |
| CHISEL | 597 | 469 | 567 | 6 |
| CHOIR | 567 | 526 | 567 | 5 |
| CHORAL | 460 | 411 | 490 | 6 |
| CHROME | 555 | 385 | 491 | 6 |
| CHUCKLE | 429 | 505 | 497 | 7 |
| CITIZEN | 455 | 535 | 445 | 7 |
| CLANG | 423 | 431 | 445 | 5 |
| CLASP | 498 | 441 | 456 | 5 |
| CLEAT | 546 | 333 | 481 | 5 |
| CLIFF | 591 | 479 | 599 | 5 |
| CLOAK | 543 | 423 | 518 | 5 |

| | | | | |
|---|---|---|---|---|
| CLOVE | 565 | 395 | 446 | 5 |
| COARSE | 445 | 506 | 494 | 6 |
| COAST | 562 | 541 | 588 | 5 |
| COCKPIT | 576 | 481 | 594 | 7 |
| COLONEL | 523 | 482 | 552 | 7 |
| COLUMN | 520 | 519 | 491 | 6 |
| CONCERN | 509 | 519 | 353 | 7 |
| CONTACT | 456 | 543 | 449 | 7 |
| CONVENT | 537 | 458 | 559 | 7 |
| CORAL | 572 | 425 | 561 | 5 |
| COSTUME | 544 | 456 | 538 | 7 |
| COUNCIL | 435 | 508 | 405 | 7 |
| COURSE | | | | |
| COURT | 509 | 549 | 552 | 5 |
| COUSIN | 502 | 515 | 478 | 6 |
| COWHIDE | 587 | 346 | 528 | 7 |
| CRADLE | 587 | 478 | 592 | 6 |
| CRAFT | 464 | 487 | 429 | 5 |
| CRANIUM | 558 | 313 | 460 | 7 |
| CRAWL | 408 | 519 | 488 | 5 |
| CROOK | 520 | 467 | 526 | 5 |
| CROWD | 546 | 523 | 548 | 5 |
| CRUISER | 571 | 490 | 553 | 7 |

| | | | | |
|---|---|---|---|---|
| CRUMB | 541 | 524 | 497 | 5 |
| CRYSTAL | 587 | 510 | 579 | 7 |
| CUISINE | 497 | 335 | 541 | 7 |
| CURFEW | 426 | 475 | 471 | 6 |
| CURLER | 600 | 501 | 521 | 6 |
| CURVE | 447 | 521 | 520 | 5 |
| CUSTARD | 549 | 460 | 515 | 7 |
| CYMBAL | 513 | 375 | 494 | 6 |
| DANCE | 502 | 550 | 510 | 5 |
| DANCER | 558 | 535 | 551 | 6 |
| DECOY | 486 | 434 | 453 | 5 |
| DELTA | 494 | 359 | 499 | 5 |
| DEPOSIT | 417 | 532 | 413 | 7 |
| DESIGN | 444 | 538 | 407 | 6 |
| DEVICE | 444 | 500 | 391 | 6 |
| DIMPLE | 547 | 442 | 518 | 6 |
| DINER | 515 | 442 | 497 | 5 |
| DONOR | 409 | 479 | 406 | 5 |
| DOORWAY | 578 | 532 | 548 | 7 |
| DRAIN | 591 | 510 | 540 | 5 |
| DRAPE | 532 | 396 | 464 | 5 |
| DREAMER | 442 | 517 | 507 | 7 |
| DRESSER | 560 | 526 | 556 | 7 |

| | | | | |
|---|---|---|---|---|
| DRIZZLE | 558 | 516 | 582 | 7 |
| DUCHESS | 568 | 416 | 525 | 7 |
| DUNGEON | 562 | 428 | 579 | 7 |
| DYNASTY | 406 | 407 | 386 | 7 |
| EASEL | 580 | 349 | 532 | 5 |
| EDITION | 439 | 499 | 373 | 7 |
| EMBRACE | 449 | 433 | 597 | 7 |
| EMPEROR | 527 | 379 | 502 | 7 |
| EMPIRE | 429 | 479 | 470 | 6 |
| ENGINE | 586 | 543 | 595 | 6 |
| ENTREE | 418 | 441 | 413 | 6 |
| ERRAND | 411 | 441 | 440 | 6 |
| ESTATE | 541 | 498 | 474 | 6 |
| FABLE | 459 | 477 | 477 | 5 |
| FABRIC | 565 | 477 | 544 | 6 |
| FELLOW | 502 | 475 | 435 | 6 |
| FEVER | 492 | 454 | 563 | 5 |
| FIDDLE | 582 | 465 | 555 | 6 |
| FIGURE | 472 | 534 | 526 | 6 |
| FILTH | 467 | 492 | 517 | 5 |
| FLANNEL | 574 | 499 | 520 | 7 |
| FLARE | 467 | 508 | 509 | 5 |
| FLAVOR | 449 | 538 | 472 | 7 |

| | | | | |
|---|---|---|---|---|
| FLEET | 520 | 465 | 510 | 5 |
| FLICKER | 415 | 470 | 506 | 7 |
| FLOCK | 477 | 434 | 516 | 5 |
| FLOOD | 553 | 523 | 598 | 5 |
| FLUTE | 587 | 496 | 581 | 5 |
| FRAME | 562 | 494 | 508 | 5 |
| FRIAR | 543 | 343 | 497 | 5 |
| FRILL | 523 | 379 | 497 | 5 |
| FROWN | 454 | 502 | 589 | 5 |
| FURNACE | 600 | 406 | 586 | 7 |
| GABLE | 449 | 355 | 424 | 5 |
| GALLON | 488 | 519 | 525 | 6 |
| GARMENT | 552 | 440 | 507 | 7 |
| GASKET | 525 | 428 | 487 | 6 |
| GAVEL | 558 | 383 | 539 | 5 |
| GESTURE | 403 | 477 | 432 | 7 |
| GIANT | 515 | 469 | 562 | 5 |
| GINGER | 522 | 455 | 430 | 6 |
| GIRDLE | 570 | 488 | 559 | 6 |
| GLACIER | 590 | 409 | 580 | 7 |
| GLARE | 439 | 458 | 536 | 5 |
| GLOBE | 535 | 477 | 583 | 5 |
| GOBLET | 592 | 306 | 577 | 6 |

| | | | | |
|---|---|---|---|---|
| GOSPEL | 403 | 437 | 440 | 6 |
| GRAPH | 553 | 524 | 535 | 5 |
| GRATE | 432 | 431 | 424 | 5 |
| GRAZE | 409 | 453 | 470 | 5 |
| GRIND | 441 | 486 | 485 | 5 |
| GROAN | 432 | 508 | 506 | 5 |
| GROVE | 538 | 374 | 470 | 5 |
| GUARD | 517 | 504 | 530 | 5 |
| GUIDE | 468 | 524 | 482 | 5 |
| GUTTER | 498 | 467 | 506 | 6 |
| HAIRPIN | 584 | 441 | 569 | 7 |
| HALTER | 550 | 374 | 453 | 6 |
| HAMSTER | 599 | 467 | 581 | 7 |
| HARPOON | 592 | 423 | 531 | 7 |
| HARVEST | 535 | 466 | 562 | 7 |
| HERMIT | 508 | 407 | 537 | 6 |
| HEXAGON | 559 | 387 | 527 | 7 |
| HIGHWAY | 575 | 488 | 581 | 7 |
| HOSTAGE | 526 | 448 | 536 | 7 |
| HOUND | 583 | 433 | 596 | 5 |
| HUNTER | 535 | 428 | 564 | 6 |
| HURDLE | 572 | 437 | 600 | 6 |
| ICICLE | 569 | 485 | 526 | 6 |

| | | | | |
|---|---|---|---|---|
| IMPRINT | 421 | 439 | 332 | 7 |
| INCOME | 429 | 521 | 475 | 6 |
| INFANT | 579 | 513 | 600 | 6 |
| INFERNO | 487 | 401 | 572 | 7 |
| INSECT | 593 | 542 | 586 | 6 |
| INVADER | 485 | 402 | 419 | 7 |
| INVOICE | 538 | 434 | 510 | 7 |
| IODINE | 576 | 396 | 508 | 6 |
| JANITOR | 532 | 463 | 508 | 7 |
| JINGLE | 437 | 409 | 497 | 6 |
| JOURNAL | 563 | 486 | 509 | 7 |
| JUDGE | 506 | 539 | 558 | 5 |
| JUGGLER | 547 | 425 | 553 | 7 |
| KABOB | 536 | 337 | 430 | 5 |
| KEEPER | 459 | 464 | 421 | 6 |
| KERNEL | 559 | 477 | 542 | 6 |
| KNUCKLE | 586 | 491 | 520 | 7 |
| LAUNDRY | 576 | 502 | 559 | 7 |
| LAWYER | 569 | 520 | 557 | 6 |
| LEVEL | 422 | 504 | 385 | 5 |
| LEVER | 572 | 518 | 515 | 5 |
| LIGHTER | 400 | 538 | 458 | 7 |
| LINEN | 581 | 515 | 551 | 5 |

| | | | | |
|---|---|---|---|---|
| LINK | 454 | 468 | 454 | 5 |
| LITER | 482 | 462 | 412 | 5 |
| LOCKER | 586 | 538 | 569 | 6 |
| LODGE | 538 | 429 | 464 | 5 |
| LOTION | 534 | 479 | 497 | 6 |
| LYMPH | 501 | 330 | 383 | 5 |
| MACHINE | 578 | 549 | 575 | 7 |
| MADMAN | 470 | 407 | 545 | 6 |
| MAGNET | 550 | 526 | 543 | 6 |
| MAIDEN | 545 | 374 | 554 | 6 |
| MAKER | 426 | 487 | 379 | 5 |
| MAMMAL | 549 | 458 | 541 | 6 |
| MAPLE | 534 | 518 | 511 | 5 |
| MARGIN | 472 | 499 | 494 | 6 |
| MARKET | 551 | 518 | 583 | 6 |
| MAX | 600 | 550 | 600 | 7 |
| MAYOR | 507 | 443 | 523 | 5 |
| MEASLES | 568 | 487 | 582 | 7 |
| MEDAL | 571 | 494 | 529 | 5 |
| MERMAID | 494 | 391 | 578 | 7 |
| MICROBE | 482 | 339 | 394 | 7 |
| MIN | 400 | 300 | 332 | 5 |
| MINER | 551 | 521 | 569 | 5 |

| | | | | |
|---|---|---|---|---|
| MINERAL | 527 | 454 | 432 | 7 |
| MIXER | 574 | 518 | 536 | 5 |
| MONARCH | 525 | 428 | 572 | 7 |
| MONKEY | 566 | 531 | 588 | 6 |
| MONSOON | 508 | 336 | 498 | 7 |
| MORTAL | 406 | 454 | 402 | 6 |
| MOTOR | 565 | 545 | 521 | 5 |
| MOVIE | 590 | 523 | 571 | 5 |
| MUCUS | 565 | 447 | 570 | 5 |
| MURAL | 515 | 451 | 515 | 5 |
| MUSCLE | 573 | 540 | 553 | 6 |
| MUSTARD | 595 | 532 | 599 | 7 |
| MUZZLE | 585 | 369 | 513 | 6 |
| NAPKIN | 585 | 495 | 582 | 6 |
| NATION | 415 | 508 | 436 | 6 |
| NATIVE | 471 | 508 | 507 | 6 |
| NAVEL | 583 | 485 | 546 | 5 |
| NECTAR | 556 | 344 | 523 | 6 |
| NEPHEW | 541 | 452 | 443 | 6 |
| NEWBORN | 452 | 383 | 532 | 7 |
| NOMAD | 512 | 342 | 516 | 5 |
| NOODLE | 588 | 499 | 530 | 6 |
| NOVEL | 529 | 530 | 547 | 5 |

| | | | | |
|---|---|---|---|---|
| NOZZLE | 555 | 412 | 513 | 6 |
| NURSERY | 528 | 461 | 542 | 7 |
| NUTMEG | 586 | 354 | 526 | 6 |
| OATMEAL | 552 | 471 | 558 | 7 |
| OFFICER | 550 | 549 | 593 | 7 |
| OPENING | 455 | 542 | 462 | 7 |
| ORCHID | 599 | 456 | 597 | 6 |
| ORGAN | 596 | 510 | 576 | 5 |
| OUTFIT | 515 | 489 | 487 | 6 |
| OUTPOST | 462 | 368 | 378 | 7 |
| OXIDE | 462 | 397 | 337 | 5 |
| OXYGEN | 484 | 529 | 430 | 6 |
| OYSTER | 573 | 453 | 521 | 6 |
| PACKAGE | 580 | 497 | 529 | 7 |
| PALETTE | 565 | 301 | 437 | 7 |
| PARADE | 523 | 526 | 578 | 6 |
| PARCEL | 525 | 503 | 509 | 6 |
| PASSAGE | 558 | 525 | 525 | 7 |
| PASTOR | 567 | 461 | 488 | 6 |
| PASTURE | 562 | 414 | 562 | 7 |
| PATENT | 400 | 426 | 375 | 6 |
| PATIENT | 487 | 538 | 526 | 7 |
| PAYMENT | 432 | 527 | 472 | 7 |

| | | | | |
|---|---|---|---|---|
| PEARL | 597 | 508 | 590 | 5 |
| PEASANT | 550 | 422 | 540 | 7 |
| PETAL | 586 | 466 | 508 | 5 |
| PHYSICS | 406 | 483 | 425 | 7 |
| PIONEER | 467 | 346 | 416 | 7 |
| PISTON | 586 | 409 | 526 | 6 |
| PLACARD | 448 | 300 | 382 | 7 |
| PLANET | 523 | 457 | 578 | 6 |
| PLANK | 592 | 483 | 598 | 5 |
| PLATTER | 593 | 478 | 576 | 7 |
| PODIUM | 546 | 455 | 508 | 6 |
| POINT | 464 | 538 | 481 | 5 |
| PORCH | 596 | 455 | 586 | 5 |
| POSTER | 592 | 545 | 600 | 6 |
| POWDER | 513 | 521 | 524 | 6 |
| PRAIRIE | 575 | 416 | 569 | 7 |
| PRIEST | 561 | 484 | 568 | 6 |
| PRISON | 570 | 462 | 593 | 6 |
| PRIZE | 474 | 508 | 517 | 5 |
| PROFILE | 510 | 445 | 572 | 7 |
| PRONG | 530 | 409 | 499 | 5 |
| PROPHET | 450 | 503 | 467 | 7 |
| PUDDING | 593 | 510 | 588 | 7 |

| | | | | |
|---|---|---|---|---|
| PUDDLE | | | | |
| PULPIT | 562 | 415 | 551 | 6 |
| PUPIL | 570 | 547 | 572 | 5 |
| PURSE | 572 | 533 | 567 | 5 |
| PUZZLE | 449 | 486 | 510 | 6 |
| QUACK | 459 | 467 | 446 | 5 |
| QUAIL | 600 | 376 | 505 | 5 |
| QUAKE | 440 | 474 | 463 | 5 |
| QUILL | 595 | 409 | 530 | 5 |
| RABBI | 572 | 515 | 557 | 5 |
| RACKET | 562 | 486 | 530 | 6 |
| RACQUET | 513 | 480 | 522 | 7 |
| RANGE | 417 | 515 | 413 | 5 |
| RAPID | 415 | 524 | 387 | 5 |
| RATTLE | 549 | 448 | 554 | 6 |
| REBEL | 439 | 448 | 497 | 5 |
| RECEIPT | 474 | 498 | 432 | 7 |
| RECITAL | 476 | 468 | 495 | 7 |
| REFEREE | 554 | 534 | 564 | 7 |
| REGION | 441 | 511 | 459 | 6 |
| RELIC | 528 | 413 | 481 | 5 |
| REPTILE | 578 | 490 | 579 | 7 |
| RESIDUE | 510 | 401 | 487 | 7 |

| | | | | |
|---|---|---|---|---|
| RESORT | 499 | 523 | 523 | 6 |
| REVOLT | 400 | 497 | 502 | 6 |
| RHYME | 434 | 480 | 475 | 5 |
| RIBBON | 600 | 480 | 563 | 6 |
| RIDDLE | 404 | 489 | 455 | 6 |
| RIDGE | 547 | 430 | 543 | 5 |
| ROBBER | 545 | 493 | 549 | 6 |
| ROCKER | 583 | 474 | 520 | 6 |
| ROSARY | 535 | 410 | 464 | 6 |
| ROSEBUD | 593 | 369 | 586 | 7 |
| ROUTE | 440 | 515 | 447 | 5 |
| RUBBER | 596 | 547 | 599 | 6 |
| SAINT | 458 | 463 | 394 | 5 |
| SALUTE | 471 | 479 | 538 | 6 |
| SAMPLER | 426 | 397 | 378 | 7 |
| SATCHEL | 593 | 364 | 580 | 7 |
| SAUCE | 576 | 522 | 569 | 5 |
| SCALE | 475 | 523 | 463 | 5 |
| SCENE | 408 | 526 | 432 | 5 |
| SCENT | 462 | 501 | 421 | 5 |
| SCHOLAR | 450 | 489 | 451 | 7 |
| SCOOTER | 565 | 468 | 569 | 7 |
| SCOUT | 562 | 452 | 578 | 5 |

| | | | | |
|---|---|---|---|---|
| SCREAM | 479 | 522 | 589 | 6 |
| SCROLL | 593 | 350 | 572 | 6 |
| SEGMENT | 485 | 451 | 480 | 7 |
| SELLER | 444 | 459 | 427 | 6 |
| SENDER | 482 | 464 | 370 | 6 |
| SERVANT | 515 | 437 | 508 | 7 |
| SETTLER | 533 | 384 | 528 | 7 |
| SEVER | 404 | 429 | 439 | 5 |
| SEWER | 564 | 489 | 538 | 5 |
| SHADOW | 457 | 536 | 565 | 6 |
| SHEAR | 456 | 428 | 484 | 5 |
| SHELL | 597 | 524 | 581 | 5 |
| SHIELD | 576 | 464 | 556 | 6 |
| SHIVER | 455 | 517 | 578 | 6 |
| SHOPPER | 497 | 528 | 519 | 7 |
| SHOVEL | 581 | 528 | 538 | 6 |
| SHRIEK | 481 | 458 | 515 | 6 |
| SHRUB | 588 | 446 | 556 | 5 |
| SHRUG | 421 | 481 | 497 | 5 |
| SHUTTER | 562 | 398 | 533 | 7 |
| SIGNAL | 464 | 507 | 513 | 6 |
| SINGER | 553 | 548 | 575 | 6 |
| SIREN | 538 | 431 | 578 | 5 |

| | | | | |
|---|---|---|---|---|
| SKATE | 562 | 534 | 563 | 5 |
| SKILLET | 579 | 432 | 586 | 7 |
| SLEEVE | | | | |
| SLICE | 443 | 540 | 507 | 5 |
| SLIDE | 504 | 529 | 490 | 5 |
| SLIPPER | 585 | 494 | 595 | 7 |
| SNAIL | 579 | 489 | 577 | 5 |
| SNEEZE | 563 | 544 | 562 | 6 |
| SNORT | 424 | 401 | 472 | 5 |
| SOCCER | 524 | 521 | 570 | 6 |
| SODIUM | 511 | 403 | 423 | 6 |
| SOPRANO | 497 | 373 | 535 | 7 |
| SPADE | 565 | 513 | 578 | 5 |
| SPARK | 526 | 505 | 539 | 5 |
| SPASM | 439 | 422 | 486 | 5 |
| SPATULA | 586 | 407 | 517 | 7 |
| SPEAR | 584 | 513 | 545 | 5 |
| SPECK | 484 | 477 | 503 | 5 |
| SPHERE | 489 | 457 | 562 | 6 |
| SPICE | 590 | 518 | 592 | 5 |
| SPIKE | 572 | 471 | 573 | 5 |
| SPIRE | 576 | 309 | 541 | 5 |
| SPLIT | 417 | 514 | 445 | 5 |

| | | | | |
|---|---|---|---|---|
| SPOKE | 526 | 532 | 466 | 5 |
| SPONGE | 597 | 538 | 577 | 6 |
| SPOOK | 417 | 475 | 449 | 5 |
| SPOOL | 565 | 499 | 552 | 5 |
| SPOUT | 468 | 433 | 466 | 5 |
| SPRAY | 514 | 521 | 559 | 5 |
| SPRINT | 411 | 461 | 526 | 6 |
| SPURT | 417 | 417 | 491 | 5 |
| SQUEAK | 461 | 506 | 492 | 6 |
| SQUINT | 456 | 528 | 515 | 6 |
| SQUIRE | 502 | 323 | 459 | 6 |
| SQUIRT | 453 | 492 | 496 | 6 |
| STABLE | 562 | 519 | 537 | 6 |
| STADIUM | 569 | 526 | 586 | 7 |
| STAIN | 535 | 534 | 533 | 5 |
| STAKE | 540 | 460 | 506 | 5 |
| STALK | 474 | 413 | 440 | 5 |
| STATUE | 600 | 444 | 562 | 6 |
| STING | 509 | 480 | 553 | 5 |
| STOOL | 592 | 531 | 584 | 5 |
| STOUT | 413 | 511 | 521 | 5 |
| STOVE | 591 | 525 | 592 | 5 |
| STRAND | 442 | 500 | 431 | 6 |

| | | | | |
|---|---|---|---|---|
| STRIDE | 409 | 438 | 462 | 6 |
| STROKE | 463 | 511 | 481 | 6 |
| STRUT | 429 | 368 | 437 | 5 |
| STUMBLE | 433 | 536 | 485 | 7 |
| STUMP | 540 | 447 | 490 | 5 |
| SUCKER | 404 | 492 | 419 | 6 |
| SUITE | 534 | 453 | 487 | 5 |
| SWALLOW | 547 | 531 | 554 | 7 |
| SWAMP | 570 | 438 | 600 | 5 |
| SWARM | 406 | 463 | 488 | 5 |
| SWEAT | 569 | 545 | 560 | 5 |
| SWEEP | 476 | 495 | 513 | 5 |
| SWORD | 577 | 444 | 597 | 5 |
| SYMBOL | 402 | 536 | 447 | 6 |
| TAILOR | 535 | 417 | 499 | 6 |
| TALES | 445 | 529 | 466 | 5 |
| TARNISH | 443 | 455 | 456 | 7 |
| TEMPLE | 565 | 450 | 547 | 6 |
| THIMBLE | 529 | 397 | 570 | 7 |
| THINKER | 403 | 474 | 405 | 7 |
| THORN | 586 | 454 | 600 | 5 |
| THROW | 400 | 548 | 477 | 5 |
| THYME | 587 | 336 | 470 | 5 |

| | | | | |
|---|---|---|---|---|
| TICKLE | 473 | 524 | 492 | 6 |
| TOASTER | 579 | 520 | 580 | 7 |
| TOKEN | 467 | 473 | 416 | 5 |
| TOURIST | 533 | 536 | 577 | 7 |
| TOWER | 585 | 463 | 596 | 5 |
| TRACK | 547 | 514 | 499 | 5 |
| TRACTOR | 590 | 518 | 585 | 7 |
| TRAIL | 511 | 508 | 525 | 5 |
| TRAILER | 597 | 528 | 587 | 7 |
| TRAIN | 592 | 548 | 593 | 5 |
| TRAITOR | 467 | 467 | 447 | 7 |
| TRAVEL | 402 | 550 | 506 | 6 |
| TREMOR | 487 | 401 | 491 | 6 |
| TRIAL | 446 | 509 | 516 | 5 |
| TRIPOD | 577 | 363 | 574 | 6 |
| TROLLEY | 579 | 449 | 585 | 7 |
| TROOP | 509 | 470 | 498 | 5 |
| TRUNK | 596 | 485 | 529 | 5 |
| TUMBLE | 433 | 490 | 461 | 6 |
| TUNNEL | 555 | 541 | 578 | 6 |
| TWEEZER | 540 | 496 | 587 | 7 |
| TWIST | 423 | 510 | 529 | 5 |
| TYPHOON | 542 | 331 | 536 | 7 |

| | | | | |
|---|---|---|---|---|
| TYRANT | 467 | 387 | 494 | 6 |
| UMPIRE | 581 | 542 | 572 | 6 |
| UNIFORM | 550 | 484 | 591 | 7 |
| URCHIN | 568 | 332 | 529 | 6 |
| UTENSIL | 567 | 494 | 534 | 7 |
| VALLEY | 575 | 515 | 600 | 6 |
| VAPOR | 499 | 455 | 493 | 6 |
| VAULT | 550 | 445 | 550 | 5 |
| VEHICLE | 558 | 534 | 593 | 7 |
| VELVET | 580 | 515 | 569 | 6 |
| VERSE | 500 | 483 | 489 | 5 |
| VESSEL | 571 | 461 | 525 | 6 |
| VILLAGE | 576 | 524 | 578 | 7 |
| VIOLET | 541 | 461 | 560 | 6 |
| VOTER | 548 | 519 | 486 | 5 |
| WAFER | 536 | 430 | 484 | 5 |
| WAIST | 563 | 540 | 530 | 5 |
| WARRIOR | 525 | 368 | 553 | 7 |
| WEAPON | 560 | 517 | 546 | 6 |
| WEDDING | 509 | 507 | 594 | 7 |
| WHACK | 409 | 350 | 486 | 5 |
| WHARF | 573 | 330 | 463 | 5 |
| WHEAT | 594 | 510 | 577 | 5 |

| | | | | |
|---|---|---|---|---|
| WHIFF | 413 | 426 | 461 | 5 |
| WHIRL | 402 | 423 | 499 | 5 |
| WHISKER | 560 | 489 | 542 | 7 |
| WHISPER | 490 | 550 | 567 | 7 |
| WHISTLE | 579 | 505 | 574 | 7 |
| WIDOW | 547 | 485 | 505 | 5 |
| WILLOW | 589 | 425 | 565 | 6 |
| WIZARD | 473 | 473 | 551 | 6 |
| WOUND | 561 | 474 | 570 | 5 |