Examining the Episodic Context Account of the Testing Effect: Does episodic recall enhance memory for context?

By

Clair Min Kyung Hong

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

August 13, 2021

Nashville, Tennessee

Approved:

Lisa K. Fazio, Ph.D.

Sean M. Polyn, Ph.D.

Dan T. Levin, Ph.D.

Melissa C. Duff, Ph.D.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

The Testing Effect

Testing one's knowledge or retrieval practice is one of the most effective techniques for learning. A wealth of evidence from experimental studies conducted in the last 100 years shows that recalling studied materials from memory reliably improves our ability to recall the information again in the future. As one of the earliest studies, Abbot (1909) showed that retrieval practice benefits memory regardless of whether the study materials make semantic sense or not using words and nonsense syllables. Shortly after, Gates (1917) demonstrated benefits of retrieval practice on long-term memory compared to restudying and introduced the term *testing effect*. Since then, benefits of retrieval practice on subsequent memory have been firmly established through a number of studies (for reviews, see Rowland, 2014; McDermott, 2021). This benefit, often referred to as retrieval-based learning or the testing effect, has been found consistently in both laboratory and classroom contexts (e.g., McDaniel, Cahill, Bugg, & Meadow, 2011; Roediger & Karpicke, 2006a). The effect is reliable and insensitive to the type of study materials. From verbal materials such as single words (Carpenter & DeLosh, 2006; Zaromb & Roediger, 2010), foreign vocabulary (Pyc & Rawson, 2009), cue-target associated word pairs (Toppino & Cohen, 2009; Pyc & Rawson, 2010), prose passages (Glover, 1989; Roediger & Karpicke, 2006c) to nonverbal materials such as pictures (Wheeler & Roediger, 1992) and films (Kubit & Janata, 2021), the testing effect has been ubiquitously observed.

1.1 Theories of the Testing Effect

While an impressive body of research to date shows the testing effect, the underlying mechanism is still unclear. With every act of retrieval, it is assumed that there is *some* change that improves one's ability to retrieve that knowledge in the future. The exact nature of that change, however, is still unknown. Several theories have been proposed to explain the role of retrieval practice on learning (e.g., Bjork, 1994 Carpenter & DeLosh, 2006; Pyc & Rawson, 2009; Rickard & Pan, 2018; Hopper & Huber, 2018) but to date, but none are currently widely accepted by the field: some theories simply re-describe the effect and fail to operationalize the mechanism underlying the benefit, some have been refuted with empirical evidence, and some are supported by equivocal evidence from indirect testing of the hypothesis.

The following section summarizes and reviews some of the most well-known accounts of the testing effect. We first begin by reviewing descriptive accounts of the testing effect – retrieval effort, dual memory, bifurcation – that offer conceptual explanations as to why recalling something from memory leads to a better recall restudying the same material. Next, we review accounts that attempt to further identify the mechanism – primary and convergent retrieval, elaborative retrieval, episodic context reinstatement – by addressing *how* the act of retrieval causes this enhancement. We conclude the section by delving into the theories behind the episodic context account and its predictions.

1.1.1 Descriptive Accounts of the Testing Effect

1.1.1.1 Retrieval Effort Hypothesis

As one of the earliest accounts of testing effect, the retrieval effort hypothesis by Bjork (1975) suggests that the act of retrieval affects the strength of memory traces. Here, memory is defined as stored representations or "traces" of information from an episodic event. When an item is retrieved from memory, these traces are strengthened in some manner: the magnitude of strengthening corresponds to the effort required to successfully retrieve the item. The retrieval effort theory provides an intuitive explanation and it is also consistent with findings on desirable difficulty: the greater the difficulty, the better the memory (Gardiner et al., 1973; Pyc & Rawson, 2009). Yet, one critical shortcoming of this account is the lack of operationalization of term *retrieval effort.* The theory does not provide any specific details on what the "effort" entails and how it leads to the observed mnemonic effect. As such, the theory has been criticized for its ambiguity (McDermott, 2021; Karpicke et al., 2014). Since the account remains vague as to what constitutes as retrieval effort, it is also unclear how it could be measured in order to be tested.

1.1.1.2 Dual Memory Theory

A contemporary theory of testing effect is the dual memory theory by Rickard & Pan (2018). The theory proposes that the initial encoding creates study memory (e.g., memory A), and this study memory is strengthened when the items are restudied (e.g., memory A+); in contrast, testing after encoding not only strengthens the study memory (e.g., memory A+) but also creates a new separate test memory (e.g., memory B). As a result, tested items are supported with two different traces of memories (i.e., memory A+ & memory B) whereas restudied items are guided by only one trace of memory (i.e., memory A+), giving rise to the comparative benefits of retrieval practice on memory as opposed to a restudy event even though the amount of time exposed to the study material is the same in both conditions. This theory is consistent with the idea behind encoding variability as an explanation for the testing effect: when items are studied or experienced multiple times, more variable encoding experiences lead to more distinct recall cues (Melton, 1970). Across ten experiments, Rickard & Pan (2018) tested their predictions based on a simplified quantitative model of the dual trace theory and demonstrated a successful model fit. However, as the authors pointed out, the model is limited in a way that it cannot account for a variety of circumstances under which the testing effect is observed such as free recall during initial test where there aren't as various recall cues. Most importantly, similarly from above, the account does not delineate what constitutes a trace (i.e., does it include the entire experience during retrieval? Or is it limited to what was encoded intentionally?) or what would be a necessary condition for a feature to be a trace. In sum, while dual memory theory is an account that can explain the observed effect, the current version of the theory cannot fully explain the mechanism underlying the testing effect.

1.1.1.3 Bifurcation Account

Along the same line of strength of memory trace, Kornell, Bjork, & Garcia (2011) proposed a bifurcation account which explains both positive and negative testing effects from retrieval practice without feedback. Negative testing effect refers to a phenomenon in which restudy produces better memory than retrieval practice: when memory is tested after a shorter retention interval, restudy produces a higher recall accuracy than retrieval practice. According to Kornell, Bjork and Garcia (2011), retrieval practice without feedback produces a "bifurcated" distribution of memory where retrieval practice leads to high rates of boost in memory strength for items with retrieval success but zero boost for those with retrieval failure (Hogan & Kintsch, 1971; Roediger & Karpicke, 2006b). In contrast, after a restudy, all of the restudied items receive equal amounts of boost in memory. While the degree of this boost is smaller than that of retrieval success items, it temporarily results in more items above the retrieval threshold after a short delay. After a longer delay, however, all items are assumed to be weakened at the same rate at which point most of restudied items fall below the retrieval threshold; retrieval success items, on the other hand, remain above the threshold and therefore more items that underwent retrieval practice are recalled than those that were restudied. The bifurcation account delineates a more detailed mechanism of how retrieval practice produces the benefit and it also explains why the benefits of retrieval practice relative to restudy vary as a function of delay. Furthermore, a recent study by (Rowland & DeLosh, 2015) showed that high recall success during retrieval practice can eliminate the negative testing effect by showing that a practice test with high recall success leads to better memory than restudy even after shorter delays. This makes sense in terms of the bifurcation account. However, similarly to the retrieval effort theory above, the bifurcation account fails to define the cognitive mechanism behind the memory boost – it assumes that a successful recall produces more learning than restudy, but it does not specify why or whether these different levels of difficulty in learning reflect different processes.

1.1.2 Mechanistic Accounts of the Testing Effect

1.1.2.1 Primary and Convergent Retrieval

Built based on the bifurcation account by Kornell, Bjork, and Garcia (2011), Hopper and Huber (2018) proposed the primary and convergent retrieval model of recall. According to this account, retrieval practice, unlike restudy, strengthens associations within an item among its features. Retrieval cues activate an initial memory state (i.e., primary retrieval): at this point, however, there is not enough support for a successful recall. Only after a subsequent process of convergent retrieval, which activates features within an item, a successful recall takes place. Such step-wise activation of item features during retrieval leads to strong intra-item associations. In contrast, restudy produces less intra-item learning because all item features are presented simultaneously. This account not only provides a more detailed mechanism of how retrieval practice enhances retention but also predicts faster recall in the retrieval practice condition from stronger intra-item associations. Through a series of experiments, Hopper and Huber (2018) confirmed that retrieval practice leads to faster recalls than restudy even under circumstances in which the negative testing effect is observed (i.e., better performance from restudy due to shorter delay between practice and a final test). The primary retrieval and convergent account made considerable contributions toward the testing effect literature by uncovering a novel mechanism of how retrieval practice uniquely differs from restudy. However, one key limitation of the account lies in its ambiguity on the types of item features that are activated during retrieval. For instance, would items with more intra-item features show stronger testing effects than those with fewer features because its overall bond is stronger? The authors of the account broadly suggest that associations formed from retrieval would include (1) item to temporal context, (2) item to retrieval cues, and (3) item to item itself, but no further details – such as what necessitates a feature to be a retrieval cue – are provided. While the primary convergent retrieval model successfully builds on older theories of the testing effect and suggests a novel mechanism, more work is needed to fully explain how episodic memory retrieval enhances memory.

1.1.2.2 Elaborative Retrieval Hypothesis

Elaborative retrieval hypothesis (Carpenter & DeLosh, 2006; Carpenter, 2009, 2011) is one account that has received the most considerable attention in recent years. According to this account, retrieval cues lead participants to generate or activate items that are semantically related to the target response. For example, when presented with a target word 'bread' paired with a weak cue such as 'basket' instead of a stronger cue such as 'butter', several words that are associated with the cue such as 'eggs' and 'fruit' (i.e., things that could also fit in a basket) are thought to be activated, which facilitate later recall of the pair 'bread – basket' (Carpenter & Delosh, 2016). In other words, elaborating on semantically related items during retrieval practice serves to strengthen the relationship between the retrieval cue and its target. While this theory provides a plausible account of the testing effect, Karpicke & Blunt (2011) showed that elaboration alone does not yield the same learning benefits as testing by directly comparing participants' recall performances after elaboration (concept mapping) and retrieval practice. In addition, the elaborative retrieval hypothesis is challenged by the principle of cue overload, which states that increasing the number of items associated with a single retrieval cue decreases the probability of recovering the target item (Raaijmakers & Shiffrin, 1981). Given that the semantically related items generated during retrieval are not provided during the final test, simply generating more items related to the cue is unlikely to aid retention of the word but rather make it more difficult to recover the original target.

1.1.2.3 Episodic Context Account

The episodic context account is one recent theory of why retrieval practice is beneficial, proposed by Karpicke, Lehman, & Aue (2014). The authors suggest that benefits of retrieval practice stem from reinstating the episodic context associated with the target (or to-be-retrieved) memory. During retrieval practice, the reinstated context representation is strengthened and updated to include the context at the moment of recall. Since the updated context includes both the study context and recall context, it provides more cues to guide one's memory search on a future retrieval attempt (Siegel & Kahana, 2014). To illustrate, if a subset of study items were studied in context A and were later retrieved in context D, their associated context representation would be a composite of A and D features; during recall, one can restrict their search only to items associated with both A and D contexts. In other words, memory retrieval of an item or event involves reinstating its prior learning context and when successful, the representation of context is updated to mix the retrieved context with the current context. Retrieving this same information in the future becomes more effortless and efficient as its associated context is a unique combination of unique combination of contexts at study and retrieval.

While the episodic context account offers a feasible mechanism of how retrieval practice strengthens memory, the details of the account remain to be clarified. First, it is unclear *what* the context entails: do all types of context reinstatement produce the retrieval practice benefits? Or is the benefit limited to only certain types of contexts? Context is broad term and can refer to multiple features of an event (e.g., passage of time during an experiment, semantic meaning of the word, the modality of presentation, the speaker's voice, the presentation font), as well as background features (e.g., the current location, the participants' mood, background noises) (Bjork & Richardson-Klavehn, 1989; Geiselman & Bjork, 1980). Yet currently, the account does not specify which aspects of context are or are not encoded and reinstated during retrieval.

1.1.3 Discussion

In sum, several theories have been proposed to explain the testing effect. Intuitively, it is not surprising that retrieval practice, an active form of learning, leads to a better memory performance than restudying, a more passive approach. While it is clear that retrieval practice involves more effort, the key to identifying the underlying mechanism lies in defining what the *effort* entails. Of several theories that had been proposed to explain the testing effect, only two accounts, the episodic context account and the elaborative retrieval hypothesis, provide a specific mechanism that operationalizes what constitutes as effort during retrieval. However, as mentioned above, the elaborative retrieval hypothesis has been refuted through studies that directly compared the effects of elaboration and retrieval practice, suggesting that pure elaboration does not produce the testing effect. On the other hand, the episodic context account is one that has received considerable attention in recent years and has also been tested directly and not yet been fully refuted. In the next section, we discuss underlying mechanisms of the episodic context account and implications from studies that directly tested the account.

1.2 A Closer Examination of the Episodic Context Account

1.2.1 Effects of Context Reinstatement during Recall

As one of the primary reasons for its appeal, the episodic context account is consistent with the role of context as a facilitator of memory search and recall as discussed in context-based memory models (e.g., Howard & Kahana, 2002; Lehman & Malmberg, 2013; Mensink & Raaijmakers, 1988). Studies have shown that participant's recall performance is associated with their ability to reconstruct and organize their memory according to certain features of the study context such as temporal and semantic features (Sederberg et al., 2010; Spillers & Unsworth, 2011). For instance, Healey, Crutchley, & Kahana (2014) found that individuals with high recall success rates also exhibited stronger tendencies to organize their recall based on temporal information from the initial study episode (e.g., grouping recalls by items that were studied together or in temporal proximity).

The proposed causal link between context reinstatement (and strengthening) and boost in memory during retrieval is in line with the two well-known frameworks: the encoding specificity principle (Thomson & Tulving, 1970; Tulving & Thompson, 1973) and the transfer-appropriate processing framework (Morris et al., 1977; Roediger et al., 2002). Both frameworks would suggest that the match between encoding and testing is key to a successful memory retrieval as increasing the levels of overlap with the encoding context during retrieval aids memory performance (Tulving & Thomson, 1973). To date, a series of studies have demonstrated the importance of contextual match between the study and test context in recall performance (see DuBrow et al., 2017, for a review). Early on, Godden & Baddeley (1975) and Smith, Glenberg, & Bjork (1978) showed that words studied underwater are better recalled under water than on a dry land, and vice versa. More recently, Shin, Masís-Obando, Keshavarzian, Dáve, & Norman (2021) created two virtual reality environments – underwater and Mars – and replicated these findings.

However, several studies have also failed to observe the benefits of context reinstatement on memory. As reviewed by Shin and colleagues (2021), contexts have been manipulated in various ways such as background colors (Weiss & Margolius, 1954; Isarida &Isarida, 2007) and physical rooms (Eich, 1985; Fernandez & Glenberg, 1985), but these manipulations do not always lead to a context-reinstatement effect (Isarida & Isarida, 2007; Fernandez & Glenberg, 1985). Furthermore, a recent study by Racsmány, Bencze, Pajkossy, Szőllősi, & Marián (2021) showed that when the context is not directly relevant to the target item (e.g., randomly assigned backgrounds scenes), reinstatement of context, in fact, impairs one's mnemonic discrimination ability (i.e., increased false recognition rates).

To date, only two studies — Brewer and colleagues (2010) and Akan, Stanley, and Benjamin (2018) — have directly examined the effects of retrieval practice on non-temporal context memory. Both studies examined participants' memory for incidental details from the original study phase after retrieval practice, but their results differ substantially.

First, Brewer and colleagues (2010) found that retrieval practice did not improve memory for context features of the items themselves. In their study, participants encoded two lists of words. Each word was presented visually and spoken by a male or female voice. After each list, participants either recalled the items or completed a distractor task. When asked to recall who presented each word on a final test, participants who had practiced retrieving the items were no better at identifying if the voice was male or female than participants in the control condition. However, a recent study by Akan, Stanley, and Benjamin (2018) did find evidence that retrieval practice improves memory for context features of the items themselves. Participants studied word pairs in one of eight locations circling the screen. During retrieval practice, participants were presented with one word of the pair in the center of the screen and were asked to recall the second word. On a final test, participants were more accurate at identifying the word pair's original spatial location when they had previously retrieved the items as compared to when they restudied them. Thus, the current evidence surrounding the episodic context account is mixed. There is some evidence that retrieval practice does enhance memory for temporal context, but mixed results for contextual aspects of the item itself.

1.2.2 Limitations

As briefly mentioned above, while the episodic context account offers a clear explanation for why and how retrieval benefits memory, the details of the account remain to be clarified. First, it is unclear what is reinstated during the memory retrieval. Context is a broad term and can refer to multiple features of an event (e.g., the modality of presentation, the speaker's voice, the presentation font), as well as background features (e.g., the current location, the participants' mood, background noises) (Geiselman & Bjork, 1980; Bjork & Richardson-Klavehn, 1989). Many memory models highlight temporal context as being especially important for memory retrieval. Temporal context is assumed to change slowly over time so that items that are encoded nearby in time share similar temporal contexts (Polyn, Norman, & Kahana, 2009; Kahana, 1996). Karpicke and colleagues (2014) emphasize the importance of temporal context, but no further details are given about what aspects of context are or are not encoded and reinstated during retrieval.

1.3 Direct testing of the Episodic Context Account

A recent study by Whiffen & Karpicke (2017) is one of few studies that were designed to directly test the episodic context account. They hypothesized that (1) thinking back to the original study episode would enhance memory, and (2) participants in the retrieval practice condition would be more likely to recall the words in its original study order during the final recall test than those in other conditions. Whiffen & Karpicke (2017) tested these predictions through a series of experiments by comparing learning from different conditions in a between-subjects design. In one experiment, participants studied two lists of words separated by a brief distracter task, and they were shown each word from both lists mixed together to either restudy or recall in which of the two lists each word had appeared during the initial study phase (i.e., list discrimination task). After this learning phase, participants in both conditions were asked to recall all the words they had studied (from both lists) in any order. Engaging in a list discrimination task rather than restudying led to better memory for both items and their list memberships, consistent with previous findings in the literature (Chan & McDermott, 2007). In another experiment, the authors compared the effects of restudying, list discrimination, and elaborative encoding on learning. Procedures were identical to the prior experiment save the elaborative encoding condition in which participants made pleasantness judgements of each word. While the same benefits were observed in the experimental conditions, engaging in the list discrimination task did not benefit memory any more than the elaborative encoding task did.

One critical limitation to these findings is that participants were asked to retrieve the list information, not the target words encoded during the initial study phase. This is important to note because the episodic context account proposes that retrieval of an item must reinstate its prior encoding context. However, in both studies, participants were not retrieving the target items while naturally retrieving the context; instead, they were instructed to think back to the original study episode. This may not be representative of what happens during memory retrieval in a setting where people are not instructed to actively think back to their previous learning episode. While Whiffen & Karpicke (2017) provide clear evidence that reinstatement of temporal information from the initial study episode enhances later retention, these findings neither support nor discredit the central claim of the episodic context account that context reinstatement occurs during retrieval practice.

CHAPTER 2

Testing the Episodic Context Account

Background and overview of Experiments 1, 2, 3, & 4

Across four experiments [1], we tested predictions from the episodic context account: would retrieval practice of an item also enhance memory for its context? We addressed this question by examining participants' memory for the initial study context after successful retrieval practice and comparing it to the memory of participants who restudied the same items. Specifically, we examined memory for source information (font colors). During the initial study phase, cue-target pairs were presented in four different colors. Participants then either practiced retrieving the targets or restudied them and were later asked to recall the font color in which the word pair had originally been presented. During the initial study phase, cue-target pairs were presented in four different colors. Participants then either practiced retrieving the targets or restudied them and were later asked to remember the original font color they saw during study.

If the episodic context account is true, and retrieval practice benefits memory by reinstating and updating the entirety of episodic context from the encoding phase, items that were successfully retrieved during practice should show enhanced memory for context details (i.e., font color) compared to those that were restudied. The successful retrieval practice should require participants to reinstate the prior episodic context and strengthen participants' memory for that context. Therefore, we expect better color memory performance for items that were successfully practiced than those that were restudied.

[1] Experiments 1, 2, and 3 in this chapter are adapted from "Examining the Episodic Context Account: Does retrieval practice enhance memory for context?" published in *Cognitive Research: Principles and Implications*, and have been reproduced with permission.
Hong, M., 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), 46

2.1 Experiment 1

2.1.1 Method

Participants

Sixty-two adults participated in exchange for course credit or $10 (14 male; mean age 19.6 years). Participation was limited to students who indicated that they were not color blind on a pre-screening questionnaire. One participant was excluded for not following the instructions during the study phase, leaving 61 participants in the analysis: 31 in retrieval practice condition and 30 in restudy condition. Participants were randomly assigned to one of the two conditions and were tested individually or in small groups of up to four people (each on their own computer).

Stimuli

Participants studied 40 word pairs selected from a pool of Jacoby's (1996) norms. Each item in the wordpool includes a cue, a fragment and two possible target words that complete the fragment (one typical and one atypical). For example, the cue-fragment pair "Flower - _a_sy", could be completed with *daisy* (typical) or *pansy* (atypical). To ensure that participants were actively recalling from the original study episode during retrieval practice, only the atypical associates were selected as target words in our stimuli. As a result, guessing based on cues (e.g., Flower - _a_sy) would frequently lead to an incorrect response (e.g., daisy) as participants are likely to complete the fragment with the more typical associate.

Out of the 104 items in the norms, we selected 40 cue - atypical target word pairs that fit our two selection criteria; (1) the difference in completion base rate of the typical (e.g., daisy) and atypical (e.g., pansy) alternative was greater than 0.3 and (2) the atypical target had a completion base rate less than or equal to 0.2. This led to our atypical target words having a mean completion base rate of 0.09 compared to a base rate of 0.62 for the typical alternatives. By only using atypical items we aimed to ensure that correctly identifying a target word could only be achieved via episodic memory recall, not by guessing or generating

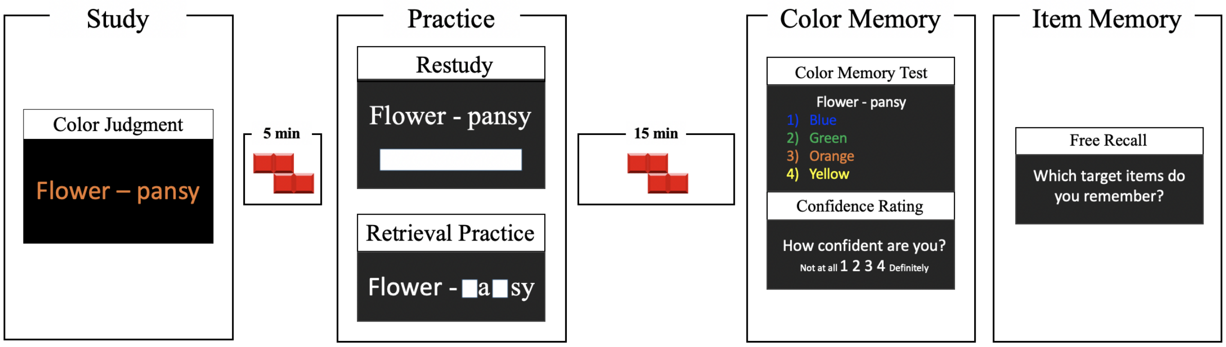

Figure 2.1: A schematic depiction of the design of Experiment 1. See text for details.

from semantic memory.

Procedure

A visual overview of the procedure is shown in Figure 2.1. Participants were first given a brief overview of the experiment by doing a quick training round consisting of the study, practice and final test phases with four word pairs that were not used in the experiment. The color memory test was not included in the training round to make it less likely that participants would intentionally try to remember the font colors for the later test.

The main experiment began with the study phase, which included 40 cue-target pairs. Each word pair was presented in one of four colors (blue, green, orange, or yellow), with ten cue-target pairs in each color. The order of the word pairs was randomized for each participant. Each pair was presented for five seconds with a one second interstimulus interval. During presentation, participants indicated the font color of each word pair by pressing a corresponding key on the keyboard (color stickers were placed over four of the keys).

A 5-min distraction period followed the study phase, during which participants played a video game (Tetris) on the computer. The key manipulation occurred during the next phase (practice). For participants in the restudy condition, the 40 cue-target pairs were presented in a scrambled order and they were asked to type the target word into a blank line under the target word. For the retrieval practice condition, participants were shown the cue word and the target fragment (again in a scrambled order). Participants were asked to complete the fragment by recalling the word they had seen earlier. For both the restudy and retrieval practice conditions, all words were presented in white font color on a black background.

After the practice phase, participants again played a video game (Tetris) on the computer for 15 minutes. Then, in the color memory test, participants were shown each of the pairs again and were asked to indicate in which of the four colors the pair was presented during the study phase. After each response, they indicated their confidence on a four-point scale with one indicating "not confident at all" and four indicating "definitely confident".

Finally, participants were asked to recall as many of the target words as possible, in any order (free recall phase). This final free recall test occurred immediately after participants re-experienced all of the word pairs on the color test and thus is not a pure measure of the effects of retrieval practice on memory. Our primary goal was to assess participants' memory for aspects of the study context. Thus, it was important to test color memory before free recall. Given that the color memory test involves another presentation of both the restudied and retrieval practiced items, it will tend to work against finding a canonical retrieval practice effect. In Experiment 3, we reverse the order of the free recall and source memory tests to assess the robustness of our findings. In Experiment 4, participants were given a cued recall test (instead of the free recall test) prior to the source memory test.

2.1.2 Results

Study phase

All participants successfully identified the font color of each word pair shown during study by pressing a key that matched in color (color indication accuracy $M = .99$, $SD = .03$).

Practice phase

On average, fragment completion was successful on 51% of the trials ($SD = .18$).

Color memory

According to the episodic context account, successfully recalling an item from a past study episode requires reinstating the entire context associated with the study event and (as a result) strengthening it. In our paradigm, the episodic context account predicts enhanced memory for font color when the items have previously been successfully retrieved. Thus, we compared color memory across the retrieval practice and restudy groups. In order to examine the effects of successful retrieval, we limited the analysis within the retrieval practice group to items that were successfully retrieved during the practice phase. While this analysis may suffer from item effects (the successfully retrieved items are likely the items that are easiest to remember), this should work in favor of the episodic context account – improving color memory for the items that were previously retrieved.

In contrast, we found that color memory for items that were successfully retrieved during practice ($M = .37$, $SD = .14$) did not differ from color memory in the restudy condition ($M = .40$, $SD = .11$), $t(59) = 0.98, p = .33, d = 0.25$, 95% CI [-0.10, 0.03] (Figure 2.2a). Color memory performance for both groups was well above chance levels (0.25). We conducted a Bayesian analysis to quantify the relative support for two hypotheses: one in which memory for font color is higher for successfully retrieved items as compared to restudied items (as predicted by the episodic context account), and a null hypothesis in which there is no difference between the groups (using JASP; JASP Team, 2018). Using a standard Cauchy prior width of 0.707, we found an estimated Bayes factor (null/alternative) of $BF_{01} = 6.87$. In other words, the data were 6.87 times more likely to occur under the null hypothesis than the alternative hypothesis. A Bayes factor of 3 is typically considered moderate evidence and a factor of 10 is considered strong evidence (Lee & Wagenmakers, 2014).

Confidence ratings

Participants' confidence ratings on the font color task were similar across participants who restudied the items ($M = 2.21$, $SD = 0.48$) and for the successfully retrieved items in the

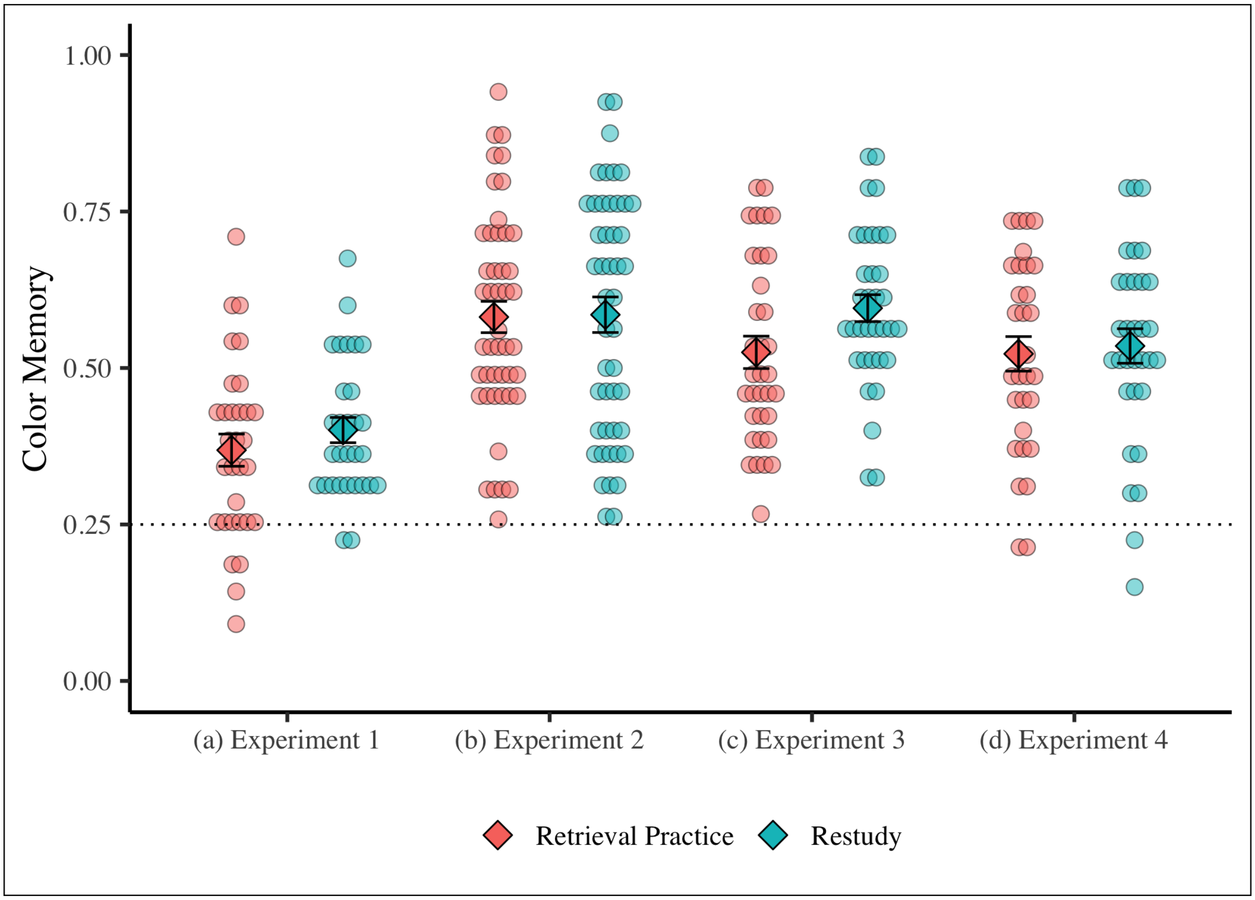

Figure 2.2: Proportion correct on the color memory test from (a) Experiment 1, (b) Experiment 2, (c) Experiment 3 & (d) Experiment 4. Across all four experiments, successful retrieval practice did not result in improved memory for source details from the original study event (font color). Each dot or square represents one participant. Each bar shows the overall group's performance on recalling colors of items they successfully retrieved (left) or restudied (right) during the practice. The mean accuracy for each condition is marked by the diamond. Error bars indicate standard errors of the mean and the dashed line indicates chance performance (0.25).

retrieval practice condition ($M = 2.15,\ SD = 0.65$), $t(59) = 0.44, p = .66, d = 0.11$, 95% CI [-0.36, 0.23].

Free recall

When scoring the free recall data, simple typos were ignored (e.g., *thief* and *theif* were both counted as correct responses). Items that were successfully retrieved during the practice phase were later better remembered than items in the restudy condition (retrieval practice success $M = .35,\ SD = .15$ vs. restudy $M = .24,\ SD = .12$), $t(59) = 3.02, p = .004, d = 0.77$, 95% CI [0.04, 0.18]. Items that were not successfully retrieved during the practice phase were less likely to be remembered during the free recall phase than items that were (retrieval practice failure $M = 0.19,\ SD = 0.14$), $t(30) = 5.94, p < .001, d = 1.07$, 95% CI [0.11, 0.22] (Figure 2.3a).

2.1.3 Discussion

Participants in Experiment 1 showed no evidence of increased memory for context following retrieval practice. According to the episodic context account, broad contextual details from the initial study phase are reinstated during retrieval. Since participants were shown the words in different colors during the initial study phase, a globally reinstated context should include source details such as font color. However, participants did not show enhanced memory for the original study context following successful retrieval. Overall, the results from Experiment 1 were inconsistent with the episodic context account of retrieval practice. We not only failed to observe enhanced memory for context for successfully retrieved items, but the overall trend was in the opposite direction with the restudy group showing better memory for context. Nevertheless, a more specific or targeted version of the episodic context account may still be true, in which a narrower slice of study context is reinstated to support retrieval of the target word.

In Experiment 1, participants pressed a key to indicate the font color of the studied item. Encoding the font color was technically obligatory, but this encoding may have been shallow,

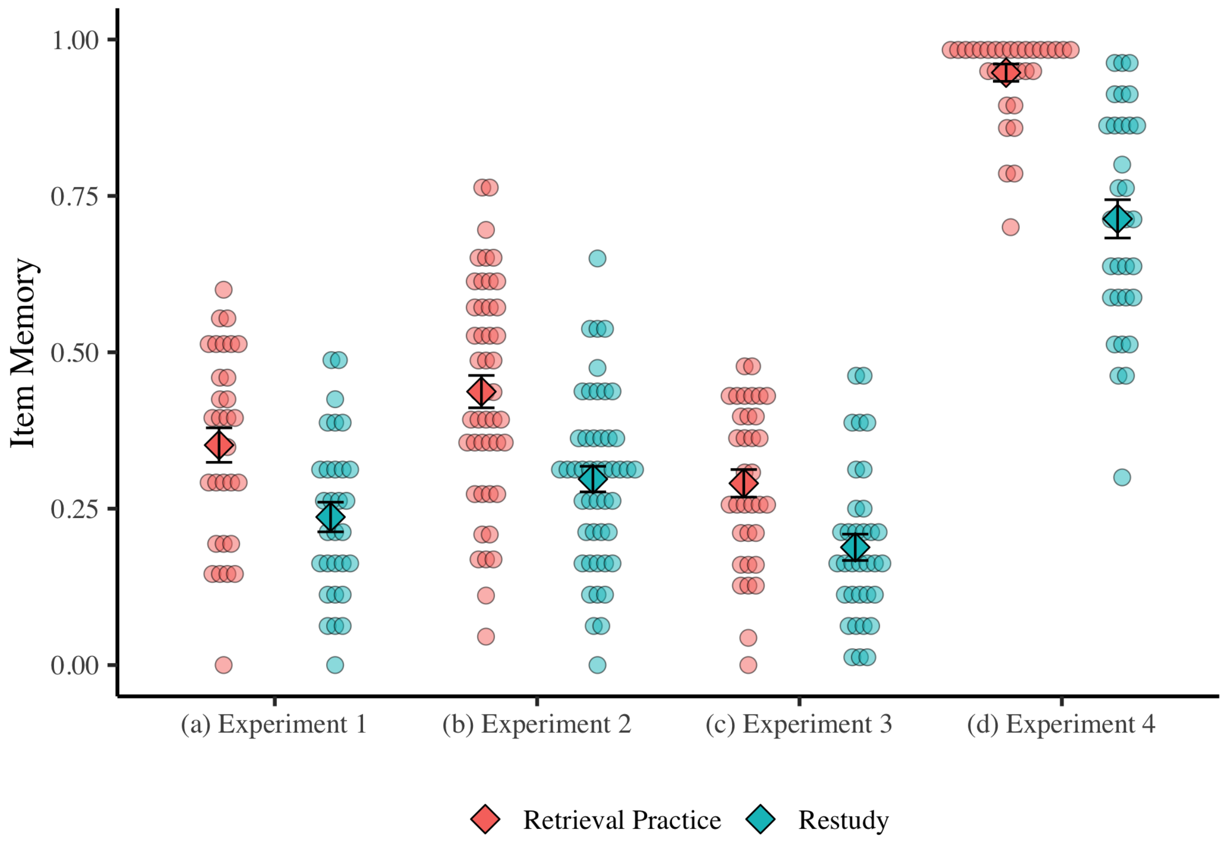

Figure 2.3: Proportion correct on the item memory test from (a) Experiment 1, (b) Experiment 2, (c) Experiment 3 & (d) Experiment 4.

and did not relate the font color to the identity of the study item. It is possible that details of the episodic context were indeed reinstated during retrieval practice, but that font color was poorly integrated into the study context, and as such was not preferentially reinstated. Thus, in Experiment 2, we used a deeper encoding task that required integrating the target word and the font color. This was designed to force font color to be a more integral part of the study context. If the episodic context account is true, we should see improved memory for font color in the retrieval practice condition.

2.2 Experiment 2

2.2.1 Method

Participants

Ninety-four subjects participated in exchange for course credit or $10 (30 male, $M_{age} = 20.67$ years). None of the subjects had participated in Experiment 1 and all indicated that they were not color blind on a prescreening questionnaire. One participant was excluded due to a software error, leaving 93 participants in the analysis: 46 in retrieval practice condition and 47 in restudy condition. The participants were tested individually or in small groups of up to four people (each on their own computer).

Procedure

The procedure was identical to Experiment 1, except for the study phase. Instead of simply indicating the color of each word pair, participants were asked to judge how consistent the font color was with the target word. Consistency was rated on a four-point scale with a response of one indicating "not consistent at all" and four indicating "highly consistent". For example, "LAKE – pool" presented in blue was commonly rated as highly consistent.

2.2.2 Results

Consistency ratings & practice phase

On average, participants rated the words as being moderately consistent with their colors ($M = 2.43$, $SD = 0.56$). In addition, the responses were well distributed along the rating scale. On average, each participant rated 26% of the word pairs as a 1, 26% as 2s, 26% as 3s, and 22% as 4s. During the practice phase, fragment completion was successful on 62% of the trials ($SD = .15$).

Color memory

The use of the integrative encoding task caused source memory performance to increase overall relative to Experiment 1 (Figure 2.2b). However, there was again no difference in color memory between items that were successfully retrieved during practice ($M = .58$, $SD = .17$) and items from the restudy condition ($M = .59$, $SD = .19$), $t(91) = 0.10, p = .92, d = 0.02$, 95% CI [-0.08, 0.07]. As in Experiment 1, color memory performance was well above chance levels (0.25) for both conditions. A Bayesian analysis with the same parameters as in Experiment 1 (H_1: font color is higher for successfully retrieved items as compared to restudied items vs. H_0: no difference) indicated a moderate support for the null hypothesis, $BF_{01} = 4.93$.

Confidence ratings

Consistent with their performance on the color memory test, participants' confidence ratings for their font color responses did not differ between items that were previously successfully retrieved ($M = 2.82$, $SD = 0.43$) and those that were restudied ($M = 2.81$, $SD = 0.52$), $t(91) = 0.10, p = .92, d = 0.02$, 95% CI [-0.19, 0.21].

Free recall

Items that were successfully retrieved during the practice phase were later better remembered than items in the restudy condition (retrieval practice success $M = .43$, $SD = .19$ vs. restudy $M = .30$, $SD = .14$), $t(91) = 3.75, p < .001, d = 0.78$, 95% CI [0.06, 0.20]. Items that were not successfully retrieved during the practice phase were less likely to be remembered during the free recall phase than items that were (retrieval practice failure $M = 0.24$, $SD = 0.23$), $t(45) = 6.11, p < .001, d = 0.90$, 95% CI [0.13, 0.25] (Figure 2.3b).

2.2.3 Discussion

We again found no evidence of enhanced memory for global episodic context following retrieval practice. Overall, compared to our results from Experiment 1, the only difference between the two experiment findings lies in the strength of the memories. Participants from Experiment 2 showed enhanced performance in both the retrieval practice and restudy conditions relative to Experiment 1 across all four measures (retrieval practice success, free recall, color memory, and confidence). This improvement in performance can be explained by the deeper processing of the study items with the modified encoding task. Despite the overall enhancement of memory for font color, retrieval practice success did not selectively enhance memory for this aspect of the study context. Memory for the original font color was equivalent across items that were successfully retrieved during retrieval practice and items in the restudy condition. This pattern of results is again inconsistent with the episodic context account.

In Experiment 3, we switched the sequence of the two final tests, color memory test and free recall test, to better match the typical design of a retrieval practice experiment.

2.3 Experiment 3

2.3.1 Method

Participants

Sixty-eight subjects participated in exchange for course credit or \$10 (21 male, $M_{age} = 19.3$ years). None of the subjects had participated in Experiment 1 & 2 and all indicated that they were not color blind on a prescreening questionnaire. The participants were tested individually or in small groups of up to four people (each on their own computer).

Procedure

The procedure was identical to Experiment 2, save the final phase where the task order was switched. Therefore, participants first recalled all of the target words they learned throughout the experiment (free recall) and then moved onto answering questions about the font color of each cue-target pair from the initial study phase (color memory test).

2.3.2 Results

Consistency ratings & practice phase

On average, participants rated the words as being moderately consistent with their colors ($M = 2.45$, $SD = 0.50$). In addition, the responses were well distributed along the rating scale. On average, each participant rated 26% of the word pairs as a 1, 26% as 2s, 25% as 3s, and 23% as 4s. During the practice phase, fragment completion was successful on 56% of the trials ($SD = 0.15$).

Color memory

Contrary to the episodic context account, participants who restudied the target items were better at recalling the original font colors (restudy: $M = .60$, $SD = .13$) as compared to the successfully retrieved items within the retrieval practice group (retrieval practice success: $M = .53$, $SD = .15$), $t(66) = 2.10, p = .04, d = 0.51$, 95% CI [0, 0.14] (Figure 2.2c). A Bayesian analysis with the same parameters as in previous experiments found strong support for the null hypothesis, $BF_{01} = 11.28$

Confidence ratings

Participants' confidence ratings for their font color responses did not differ between successfully retrieved items ($M = 2.64$, $SD = 0.50$) and restudied items ($M = 2.79$, $SD = 0.39$), $t(66) = 1.41, p = .164, d = 0.34$, 95% CI [-0.37, 0.06].

Free recall

Items that were successfully retrieved during the practice phase were better remembered than items in the restudy condition (retrieval practice success $M = .29$, $SD = .13$ vs. restudy $M = .19$, $SD = .12$), $t(66) = 3.36$, $p = .001$, $d = 0.81$, 95% CI [0.04, 0.16]. Items that were not successfully retrieved during the practice phase were less likely to be remembered during the free recall phase than items that were (retrieval practice failure $M = 0.05$, $SD = 0.08$), $t(33) = 11.62$, $p < .001$, $d = 1.99$, 95% CI [0.20, 0.28] (Figure 2.3c).

2.3.3 Discussion

Consistent with what we found in Experiments 1 & 2, we again found no evidence of improvements in memory for the study context (font color) following successful retrieval practice. In fact, participants who restudied the items showed better color memory. It is clear that successful retrieval does not improve memory for all aspects of the encoded episodic context.

In Experiment 4, we changed the item memory test from a free recall to a cued recall test. This change was important because here, participants in both retrieval practice and restudy conditions are prompted with retrieval practice of all 40 items prior to the color memory test. If the episodic context account is true, the cued recall test would presumably necessitate all participants to think back to the original study episode and "reinstate" the presentation of each target word in one of four colors. In Experiment 4, all participants are given at least a single opportunity for item retrieval (once for restudy condition and twice for retrieval practice condition) before the color memory test. Therefore, if the episodic context account is true, the effects of context reinstatement and strengthening from retrieval should be even more pronounced in the retrieval practice condition from Experiment 4.

2.4 Experiment 4

2.4.1 Method

Participants

Sixty-three subjects participated in exchange for course credit or \$10 (23 male, $M_{age} = 19.7$ years). None of the subjects had participated in Experiment 1, 2, & 3 and all indicated that they were not color blind on a prescreening questionnaire. The participants were tested individually or in small groups of up to four people (each on their own computer).

Procedure

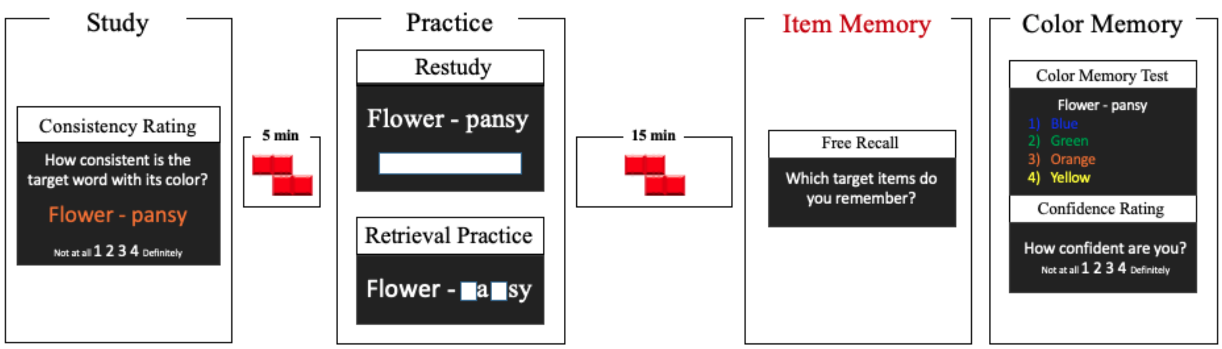

Figure 2.4: A schematic depiction of the design of Experiment 4. See text for details.

The procedure was identical to Experiment 3, save the item memory test phase where the free recall task was replaced by a cued recall task (Figure 2.4). During the cued recall test, participants were shown all of 40 cue-target fragment pairs in a scrambled order and were asked complete the fragment by recalling the correct target. Then, participants moved onto answering questions about the font color of each cue-target pair from the initial study phase (color memory test).

2.4.2 Results

Consistency ratings & practice phase

On average, participants rated the words as being moderately consistent with their colors ($M = 2.46$, $SD = 0.58$). In addition, the responses were well distributed along the rating scale. On average, each participant rated 25.2% of the word pairs as a 1, 27.5% as 2s, 23.3% as 3s, and 24% as 4s. During the practice phase, fragment completion was successful on 60% of the trials ($SD = 0.22$).

Color memory

Once again, contrary to the episodic context account, there was no difference in color memory between participants who restudied the target items ($M = .54$, $SD = .16$) and those who successfully retrieved them during the practice phase ($M = .52$, $SD = .15$), $t(61) = 0.32, p = .749, d = 0.08$, 95% CI [-0.09, 0.07] (Figure 2.2d). A Bayesian analysis with the same parameters as in previous experiments found moderate support for the null hypothesis, BF_{01} = 3.70.

Confidence ratings

Participants' confidence ratings for their font color responses did not differ between successfully retrieved items ($M = 2.35$, $SD = 0.64$) and restudied items ($M = 2.57$, $SD = 0.57$), $t(61) = 1.49, p = .143, d = 0.37$, 95% CI [-0.53, 0.08].

Cued recall

Items that were successfully retrieved during the practice phase were better remembered than items in the restudy condition (retrieval practice success $M = .95$, $SD = .19$ vs. restudy $M = .71$, $SD = .17$), $t(61) = 6.87$, $p < .001$, $d = 1.73$, 95% CI [0.17, 0.30]. Items that were not successfully retrieved during the practice phase were less likely to be remembered during the free recall phase than items that were (retrieval practice failure $M = 0.10$, $SD = 0.08$), $t(29) = 42.06$, $p < .001$, $d = 7.68$, 95% CI [0.80, 0.88] (Figure 2.3d).

2.4.3 Discussion

In Experiment 4, all participants engaged in a cued retrieval of all 40 items from the study list immediately prior the color memory test. Consistent with what we found in Experiments

1, 2, and 3, we again found no evidence of improvements in memory for the source context (i.e., original font colors) following successful retrieval practice.

2.5 Discussion of Experiments 1, 2, 3, & 4

Our findings suggest some limitations and constraints on the episodic context account proposed by Karpicke and colleagues (2014). Their account suggests that the benefits of recall practice stem from reinstating the episodic context of original study event. Successful retrieval strengthens the context memory associated with the target item, and the retrieved context is updated to include features of the recall context. Because this strengthened and updated context memory becomes an effective retrieval cue for later recall attempts, items that receive retrieval practice become more memorable than those that were simply restudied.

While this account provides a tenable explanation as to why retrieval practice benefits memory, the account does not specify the type of information reinstated from the episodic context. Context, in general, is a broad concept, and can refer to many aspects of the encoded event including characteristics of the target item itself such as typeface, stimuli color, and mode of presentation, as well as other features of the episode such as the testing room, outside noises, one's mood or internal thoughts. In the current studies, we found no evidence that contextual details of the studied words themselves were strengthened following retrieval even though these contextual details were part of the encoding process and were remembered at above chance levels during the color memory test. It remains possible, however, that a more limited version of the episodic context account may be correct; one in which some but not all aspects of the episodic context are reinstated during retrieval practice.

Our results suggest that contextual aspects of the item itself (e.g., its color) are not strengthened by retrieval practice. This is consistent with the results from (Brewer et al., 2010) who found that retrieval practice did not affect participants' memory for a different aspect of the item (the voice of the speaker). Building on the results of Brewer and colleagues, we found that these contextual details were not strengthened even when they were emphasized as a part of encoding task. Thus, neither the original item color nor the presenting voice is retrieved and strengthened during retrieval practice.

According to the episodic context account, memory retrieval improves retention through reinstating its associated context. Retrieval practice reinstates and strengthens its context memory, which serves as a retrieval cue that facilitates and guides recall. As a result, if the episodic context account is true, one would observe a signature of context reinstatement through memory organization. For example, if font colors were part of the reinstated episodic context in our experiments, we would expect participants in the retrieval practice condition to group their recalls by font colors when asked to free recall all the words in any order (i.e., recalling all of the items presented in orange together). We therefore measured participants' font color-based grouping during free recalls by computing an Adjusted Ratio of Clustering (ARC) score (Roenker et al., 1971) from each participant in Experiments 1, 2, & 3 (Experiment 4 was excluded as the item memory test was a cued recall test where participants could not decide on the order of recalls). ARC quantifies the extent to which subjects tend to cluster responses according to taxonomic categories and ARC scores can range in value from -1.0 to 1.0, where 0 indicates that the amount of clustering is no greater than the expected by chance alone, and 1.0 indicates perfect clustering. From all three of our experiments, however, we not only failed to observe better color memory in the retrieval practice condition, but we also did not see any recall clustering pattern by the font colors. On average, participants' mean ARC score was below 0.1 in all three experiments, confirming that there was no presence of clustering by font colors overall.

As mentioned earlier, however, there is evidence that memory for certain kinds of context are strengthened during retrieval practice. Retrieval practice has been shown to improve memory for temporal context, specifying on which list an item appeared on, and spatial context, specifying where an item appeared on the screen (Akan et al., 2018; Brewer et al., 2010; Chan & McDermott, 2007). Finally, aspects of the context that are explicitly retrieved are also strengthened. Participants in the Brewer studies above who were asked to recall both the word and the speaker's gender during retrieval practice did show enhanced memory for the speaker's gender on a later test (Brewer et al., 2010).

The episodic context account suggests that the benefit of retrieval practice is due to the reinstatement of study period contextual details and the association of these details with practice-phase contextual details. However, it seems that certain contextual characteristics of the study event (voice gender and font color) are not more memorable following retrieval practice. These contextual details are clearly part of the memory for the study event, as they are remembered at levels well above chance when tested directly. In addition, retrieval practice failed to increase context memory both when the encoding of context was incidental (experiment 1; Brewer et al., 2010) and when contextual details were an active part of encoding task (experiments 2 and 3).

One possible interpretation of these results is that the "context" defined in the episodic context account is more limited than previously suggested. It may be the case that retrieval practice only enhances memory for those aspects of the event that are specifically targeted by the retrieval probe or are automatically reinstated during retrieval. The current research suggests that memory for item-related characteristics of a study event, such as voice gender and font color, are less likely to benefit from retrieval practice than spatial and temporal characteristics. Bjork & Richardson-Klavehn (1989) proposed that there are two different types of contextual features that may be reinstated to enhance recall performance: intraitem context feature and extraitem context feature. An intraitem context feature refers to any physical attributes of a presented item such as font color and voice, while an extraitem context feature refers to any physical aspects of the environment surrounding the stimuli such as background noise, mood, or one's physiological state (Geiselman & Bjork, 1980). Bjork & Richardson-Klavehn (1989) argued that these two types of contexts interact with target items and create different types of contexts, two of which are integral context and incidental context. As the name suggests, integral context refers to features that are essential to encoding. An example of an integral *intraitem* context would be the language of the target word, and an example of integral *extraitem* context would be a feature of an environment that is directly associated with the target word as in the method of loci. In contrast, incidental context refers to non-essential aspects of the target item. Examples of incidental intraitem context would be print color or letter case, and those for incidental *extraitem* context would be room temperature or background noise. Bjork & Richardson-Klavehn (1989) suggested that when the relationship between context and the target item is integral, physical reinstatement of context would enhance memory no matter the what type of context it is; on the other hand, when the relationship is incidental, whether physical reinstatement would aid recall depends on various other factors such as uniqueness of the cue, and number of targets associated with the same cue. While font colors in Experiments 1, 2, 3 and 4 can be considered as incidental inter-item context, the encoding tasks in all three experiments required participants to pay attention to the colors, making colors an "integral" part of the encoding process. Therefore, if the reinstated context during retrieval were to encompass the entirety of the encoding context, items that were successfully recalled during practice should have preserved color information better relative to those that were restudied.

However, it is also possible that the retrieval practice advantage is unrelated to the successful retrieval of episodic contextual details. It could be that episodic contextual information from the study phase is sometimes recalled during the practice phase, but this has no bearing on the enhanced memorability of items that are successfully retrieved during practice. By this account, certain episodic contextual details (such as spatial position or list context) may simply be details that are more likely to be retrieved despite not explicitly being cued, and other contextual details (such as voice gender or font color) are less likely to be retrieved spontaneously. In other words, while the observation of enhanced memory for source details is consistent with the episodic context account of retrieval practice, it is also consistent with alternative theories in which source details are sometimes retrieved, but are not mechanistically involved in the retrieval practice advantage.

Regardless of the true cognitive mechanism underlying the retrieval practice advantage, the current experiments provide an important demonstration that retrieval practice does not strengthen all aspects of episodic context are less likely to benefit from retrieval practice than spatial and temporal characteristics.

CHAPTER 3

Temporal and Semantic Context Retrieval during Episodic Recall

In light of our new findings from Experiments 1, 2, 3, & 4 showing that contextual features such as font colors are not reinstated and strengthened during recall, we aimed to examine context features that are automatically encoded during recall in Experiments 5 & 6. Unlike external features related to stimuli presentation such as the speaker's voice or presented font colors, encoding of temporal or semantic context do not require much extra effort or attention. Therefore, it is plausible that temporal or semantic context would be used effectively as a recall cue to guide one's memory search (e.g., organizing one's recalls by the order they were studied or grouping recalls by category). Therefore, the goals for Experiment 5 and Experiment 6 were testing for reinstatement of temporal context (Experiment 5) and semantic context (Experiment 6) during recall.

3.1 Temporal Context as a Guide for Memory Search

There is some evidence that temporal context is strengthened during retrieval practice. In a multiple list learning study, retrieval practice of items had been shown to boost participants' ability to identify each item's list membership (i.e., on which list was the item presented?) in addition to the enhancement in item memory (Brewer et al., 2010; Chan & McDermott, 2007). In other words, retrieval practice of studied items leads to better memory for their associated temporal context.

Furthermore, predictions of temporal context reinstatement during recall are consistent with its known role as a *guide* for memory search (Kahana, 1996; Howard & Kahana, 1999). In a standard memory study, participants study a list of words and are later asked to recall those in any order. Even though there is no special instructions on how they should orient their focus during study or recall, participants often organize their recalls by the order in which items were studied (Healey, Long, & Kahana, 2019). This phenomenon, known as the contiguity effect, has been replicated and firmly established through a large data set from over 30 behavioral experiments as well as simulations from memory models (Healey et al., 2019). Context Maintenance and Retrieval (CMR) model by Polyn, Norman, & Kahana (2009) characterizes the temporal context associated with each item as a part of internally maintained retrieval cues. Temporal context reflects the passage of time and is assumed to change slowly in a gradient-like manner as shown in Figure 3.1(a). Because of this gradual rate of change, temporal context is highly similar for items that are presented close in time: conversely, it is highly dissimilar for items that are studied far apart. During recall, it is hypothesized that we search through our memories by revisiting the past, often referred to as *mental time travel* (Tulving, 1993): as we travel back to the original event, key characteristics related to the target memory are often revisited or reactivated (Tulving, 2002; Polyn & Sederberg, 2014). Therefore, when an item memory is targeted to be reactivated and recalled, there is a high chance that memories of non-target items that appeared nearby in time would also be activated. Behaviorally, this effect is observed from participants' recall transitions: after recalling an item, the next recall is likely to be chosen from its temporal neighbors due to the similarities in temporal context. For example, suppose a participant is asked to study words 'apple', 'cat', 'boat', and 'dog' as illustrated in Figure 3.1(b). Later during a free recall test, if a participant recalls the word 'boat', the next recalled item is most likely to be items studied right before or after 'boat', as shown by the highest conditional response probabilities for 'cat' and 'dog'. Notably, the conditional response probability is higher for 'dog' (lag +1) than 'cat' (lag -1). This is most likely due to the forward-direction bias observed in the contiguity effect (Polyn, Norman, and Khana, 2009): our tendency to recall events from the past in a temporally forward manner.

To date, many studies have established the reliability of the temporal contiguity effect (for a review, see Healey, Long, & Kahana 2019). It has been shown to be insensitive to various experimental manipulations such as amount of practice, participant's age, IQ, type of encoding task, category structure, and so on. While some factors can modulate the size

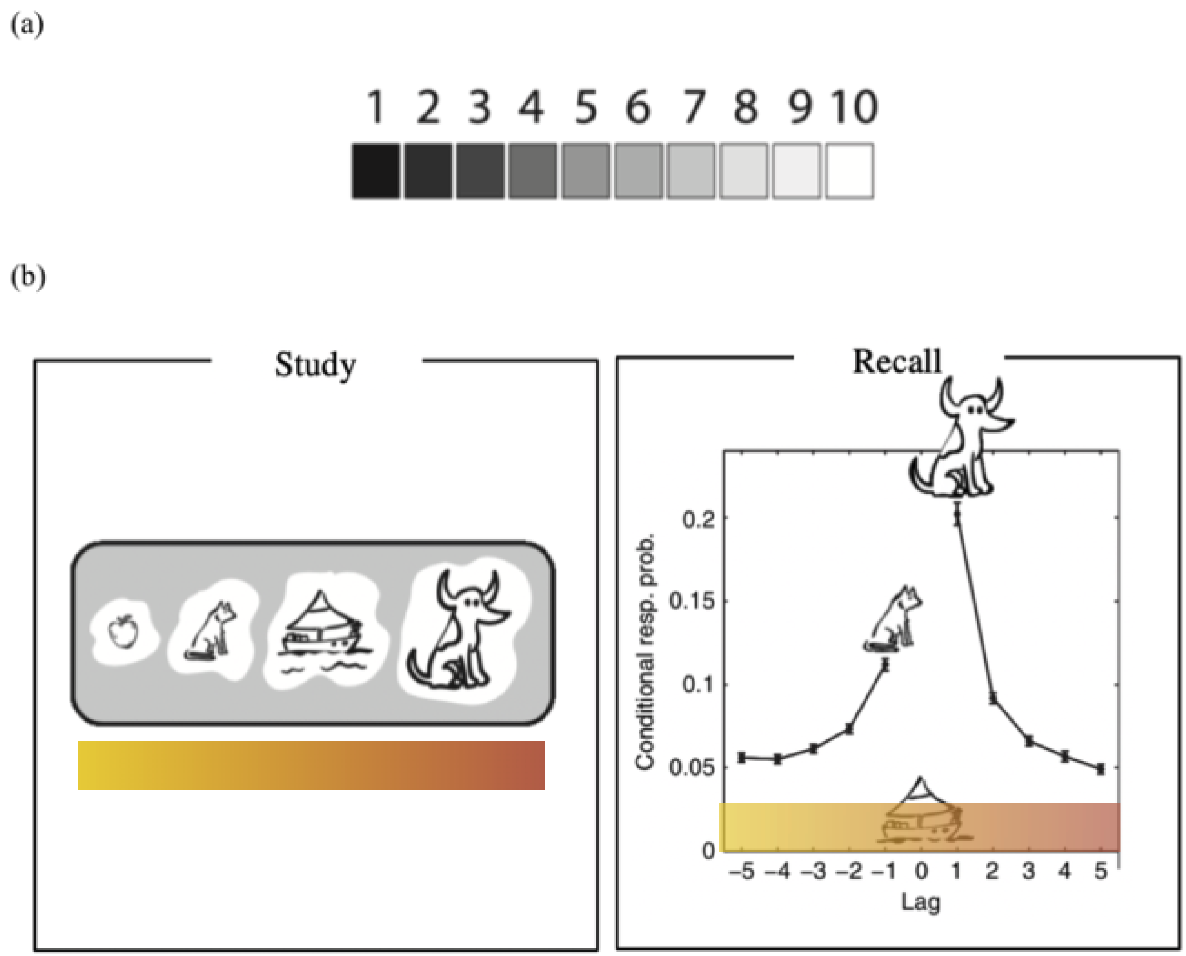

Figure 3.1: Context Maintenance and Retrieval Model. (a) An example of gradient-like context representation from Siegel & Kahana (2014b). Each number represents an item during study and the grayscale shade represents the associated temporal context. (b) An illustration of the conditional–response probability as a function of study position or relative lag adapted from Khana, Howard, & Polyn (2009). Lag of 0 represents a recall response made during free recall (i.e. 'boat'). Of all the recall transitions that can be made for the next recall, participants are most likely to recall an item that was studied right before (lag of -1, i.e., 'cat') or right after (lag of +1, i.e., 'dog') the recalled item. (c) Recalling an item and reinstating its temporal context can cue its (temporal) neighbors into memory because the associated temporal context for temporally contiguous items are highly similar.

of the effect (e.g., amount of practice, higher contiguity effect observed on the 12^{th} block than on the 1^{th} block), no single factor has been shown to abolish the contiguity effect. Such ubiquity of the temporal contiguity effect illustrates its role as a reliable guide for memory search and recall. As a result of this memory search process, it is reasonable to hypothesize that recalling an item would also reinstate its temporal context.

Notably, there is one condition under which temporal contiguity is shown to be nearly eliminated: the presence of semantic context. After studying a list containing exemplars from several distinct semantic categories, participants tend to show reduced temporal contiguity during free recall Healey & Uitvlugt (2019); Polyn et al. (2011): instead of organizing their recalls based on the temporal information, participants group their recalls by semantic categories. Such finding is This makes sense under our hypothesis that context reinstatement results from the memory search process. It is possible that the type of reinstated context depends on which context features are used to guide the recall process. Therefore, it is reasonable to predict that recalling items from a study list with strong semantic associations would result in semantic contiguity, instead of temporal.

3.2 Context Reinstatement during Episodic Retrieval: Facilitation vs. Forgetting

3.2.1 Retrieval-Induced Facilitation from Context Reinstatement

If a retrieved item's associated context is reinstated and strengthened during recall, would other un-recalled items sharing similar contexts also benefit from this retrieval? Based on the memory search process proposed by the Context Maintenance and Retrieval model (Polyn et al., 2009), recalling an item and reinstating its temporal context would cue its temporal neighbors into memory (Figure 3.2, Prediction A). In other words, temporal context reinstatement during recall of an item could induce indirect retrieval practice for its temporal neighbors. The underlying concept of retrieval-induced facilitation is similar to the testing effect as it focuses on memory enhancement from retrieval practice; however, unlike the testing effect, it also focuses the effects of retrieval on non-target memory that are not directly retrieved or practiced, yet still benefit from the retrieval event. For example, in one experiment by Jonker, Dimsdale-Zucker, Ritchey, Clarke, & Ranganath (2018), participants studied association between scene contexts and objects where each scene was paired with two different objects. During practice, some items were selected for retrieval practice while others were restudied. Results from participants' final free recall test showed that retrieval practice enhanced memory for not only the practiced practiced items (i.e., the testing effect) but also for non-practiced item sharing the same scene context as compared to the restudy condition. Jonker and colleagues (2018) suggested that this effect was associated with reactivation of the shared encoding context during retrieval practice.

3.2.2 Retrieval-Induced Forgetting from Context Reinstatement

On the flip side, reinstating context upon item retrieval could also *impair* memory for items that are associated sharing similar contexts. Studies on retrieval-induced forgetting show that selective cued recall practice improves recall for the retrieved items but impairs memory of the non-practiced items that share the same recall cue (for a recent review, see Bäuml et al., 2017)(Figure 3.2, Prediction B). As one of the seminal studies of this effect, Anderson and colleagues (1994) asked participants to study a series of category-exemplar word pairs where each category membership was shown multiple times with a different exemplar (e.g., Metals - Nickel, Metals – Lithium, Fruits - Apple, Fruits - Orange). Participants then performed cued recall test for half of the categories where they were asked to recall the target word given the category membership and the first two letters of the exemplar as a cue. Critically, participants only performed the cued recall test for only some of the exemplars from the study list (e.g., Metals - li_ _ _ _ _). Results from the final recall test showed impaired memory for unpracticed exemplars from these categories (e.g., Metals - Nickel) relative to associations from unpracticed categories (e.g., Fruits - Apple, Fruits - Orange). In other words, selective retrieval practice led to a poorer memory for unpracticed items, not just compared to those practiced but also to those that were not practiced at all. Research on

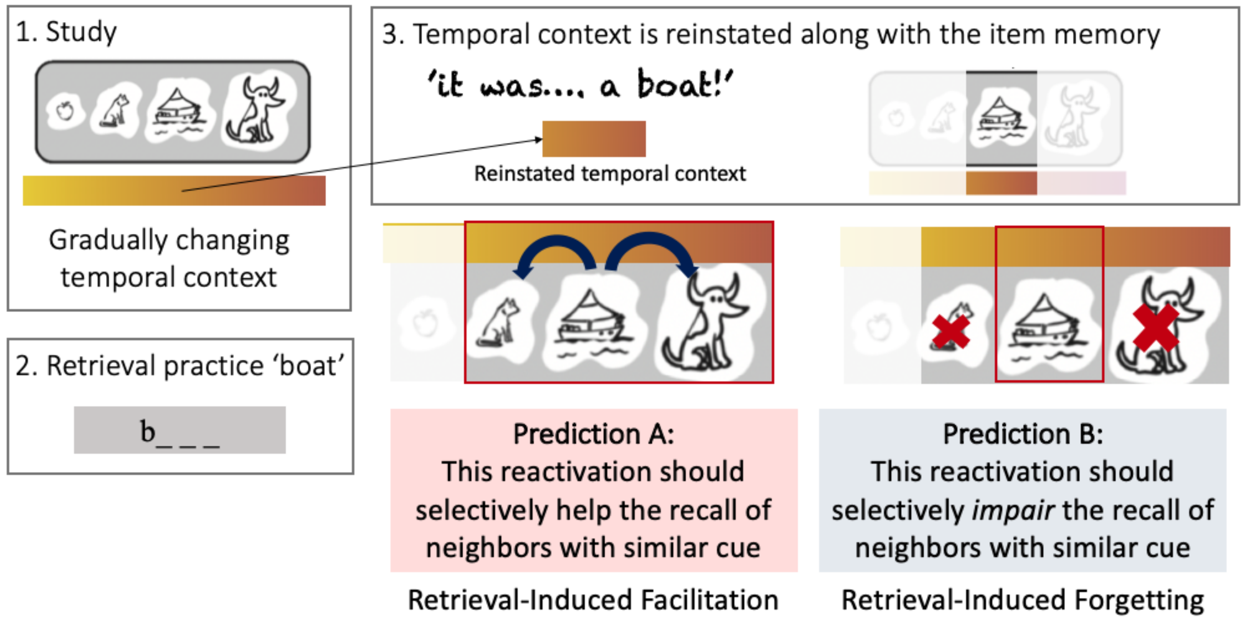

Figure 3.2: Predicted effects of retrieval practice on non-practiced items

retrieval-induced forgetting has also shown that impoverished memory for unpracticed items happens only at the presence of *retrieval* such that restudy of selective items do not produce the same effect (Bäuml, 2002; Bäuml & Aslan, 2004). This phenomenon makes sense based on how recall competition is described in the Context Maintenance and Retrieval model. ...

3.2.3 Facilitation or Forgetting: Factors that modulate the effects of reinstated context on non-retrieved items

In sum, we have two contrasting lines of research showing how retrieval practice influences memory for unpracticed, but contextually related items. Context reactivation during retrieval could *enhance* the memory for unpracticed items from indirect retrieval practice, or *impoverish* the memory as they are unsupported during recall competition against the retrieved target item. This discrepancy is puzzling as the proposed mechanisms for both are based on the same assumption: context is reinstated during retrieval. What could be the key factor that modulates this effect? While there's no clear answer yet, several potential mechanisms have been suggested and reviewed below.

First, the facilitation or forgetting could be mediated by the type of context that functions as a recall cue. To date, retrieval-induced memory literature focuses primarily on two types of contexts: temporal context and semantic context. Extensive body of work in the retrieval induced forgetting literature suggest that selective retrieval of an item from one semantic category would inhibit later recall of other items from the same category (Anderson, 2003; Bäuml, 2002). In contrast, the opposite might happen when temporal context is used as a recall cue: upon retrieval, memory for items sharing similar temporal context may be boosted while memory for items with dissimilar temporal context (i.e., studied far apart in time) may be inhibited. As reviewed previously, memory models suggest that episodic memory is organized around temporal context. This idea is supported by the nearly universal phenomenon in free recall experiments, temporal contiguity effect, where retrieval of one item facilitates recall of other items that were studied nearby in time (Howard & Kahana, 2002). Likewise, retrieval of an item may indirectly enhance one's memory for items that were studied near by in time, producing an indirect testing effect (Rowland & DeLosh, 2014). This boost in memory for a certain temporal context then ties these items together under the shared context (Liu & Ranganath, 2021). Consequently, it is possible that later recall for items that do not share any similarity in temporal context with the retrieved item (i.e., items that were studied far before or after the retrieved item) may be reduced as memory search is focused primarily around those boosted subset and items with similar temporal context.

It is also possible that the boundaries between facilitation vs. forgetting depend on the saliency of recall cue or how strongly the cue is tied to the target item. Retrieval-induced forgetting had been observed in studies where non-retrieved items share a recall cue that is already strongly associated with the retrieval-practiced item, not just associated via random pairing in the experimental task (revise but main point: there's pre-exisitng strong association between cue-target, not like some new association that were forged from the experiment stimuli structure): when prompted with a recall cue, all associated items are cued into memory, compete, and the strongest memory (the retrieved item) survives while others are suppressed. For example, a number of studies have shown that retrieval practice of an item from one semantic category inhibits retention of other items from the same semantic category as retrieval of items that share similar semantic context are suppressed due to higher competition. Based on this assumption, if context reinstatement is present during episodic memory retrieval, we would predict impoverished recall for items with the shared "cue" context: when this cue context is reinstated, items associated with the similar/same cue will be suppressed or inhibited since the retrieval-practiced item has the strongest association to the reinstated cue.

While we have two predictions pointing to opposite directions, both hypotheses are useful for testing the presence of context reinstatement during retrieval. If there is *any* context reinstatement during retrieval, one should expect to observe either improved or impoverished memory for contextually related items of previously retrieved targets. Conversely, in the absence of context reinstatement, the memory for the contextually related items would not differ from items in that were not practiced.

CHAPTER 4

Testing the Episodic Context Account: Memory for non-retrieved temporal and semantic neighbors

4.1 Overview of Experiments 5 and 6

In Experiments 5, we tested the presence of context reinstatement by examining how retrieval practice of an item affects later memory for non-practiced items that share highly similar context. The episodic context account assumes that an item memory is activated through its reinstated context. During this activation process, we hypothesize that non-practiced items whose associated context that is highly similar to the retrieved context are also cued into memory along with the practiced item. As a result, retrieval practice would not only activate and boost memory for the target item, but it will also enhance memory for items with similar associated context that were also unintentionally cued into memory. Taken together, we can test the presence of context reinstatement during recall by examining memory for items whose associated context is similar to the one that is retrieved during retrieval practice. If the episodic context account is true such that each item retrieval is supported through reinstating the associated temporal context or semantic, benefits of retrieval practice on memory should extend to non-target items with similar associated context.

How do we examine temporal context? As the name suggests, temporal context reflects the passage of time. Here, we assume that temporal context changes gradually for each study event: the order in which each item is studied directly reflects the temporal context associated with each item. As a result, the associated temporal context is highly similar for items studied close in time. Conversely, the temporal context for an item from the beginning of the study list is highly dissimilar to an item studied at the end of the list. In addition, semantic context in our current experiment will be measured through each word's category membership such that items from the same category are assumed to share highly similar semantic context.

To be concrete, suppose a participant studies a list of semantically-distinct lists of words as shown in 4.1 and is prompted to recall one of the items, 'willow'. When the temporal context associated with the item 'willow' is reinstated, its neighboring items 'clarinet' and 'nylon', whose associated temporal context is highly similar to the one that is reinstated, are also likely to be cued into memory along with the target item 'willow'. In other words, if the episodic context account is true such that each item retrieval is supported through reinstating the associated temporal context, then retrieval practice should benefit memory for not only the target items but also those that share similar context.

What happens when both temporal and semantic context features are encoded during study? Suppose a participant studies a list of words drawn from two distinct categories, musical instruments and flowers as shown in 4.6. If the semantic context (i.e. category membership) is reinstated upon retrieval of item 'Freesia', other flower items on the list will have a higher chance of being cued into memory than semantically dissimilar instrument items such as clarinet and harp. However, if temporal context is reinstated upon realling 'A3', then the temporal-contextually similar items 'B2' and 'B4' will be more likely to be reactivated than temporal-contextually dissimilar items 'A1' and 'A5'.

To sum up, Experiment 5 was designed to test for presence of temporal context reinstatement and Experiment 6 was designed to examine reinstatement of temporal and/or semantic context during recall. The presence of context reinstatement in both experiments are measured by examining memory for the retrieved-context neighbors. For temporal context, a retrieved-context neighbor item is one that was studied right before or after the retrieved item whose associated temporal context is highly similar to the retrieved item. For semantic context, a retrieved-context neighbor is one that shares semantic features with the retrieved item such as an item from the same category. If context reinstatement is present during episodic memory retrieval, selective retrieval of an item will result in stronger memory of the item itself as well as its retrieved-context neighbors (temporally or semantically). That is, if temporal context is reinstated during recall, the testing effect would be spread to items that were studied right before or after the retrieved item. Similarly, if semantic context is reinstated during recall, items that are from the same semantic category as the practiced target will benefit from the retrieval practice event even though they were not directly practiced. Conversely, in the absence of context reinstatement, the memory for retrieved-context neighbors would not differ from the non-contiguous items.

4.2 Experiment 5

4.2.1 Method

The study design, sample size, and analyses plan were pre-registered through aspredicted.com (#35380, https://aspredicted.org/blind.php?x=vc2su9).

Participants

To date, no study has examined the effects of memory retrieval on the retrieved-context neighbors. Therefore, we used an estimate from a prior work on the effect of retrieval practice on memory for context by Akan, Stanley, and Benjamin, (2018). To achieve 80% power to detect an effect size of $d = 0.4$ at alpha level of 0.05 for paired-samples t-test we had planned to recruit a minimum of 54 participants. However, due to COVID 19 pandemic, participant recruitment was interrupted, leaving 44 participants in our sample. Two individuals excluded from the analysis for incomplete data, leaving 42 participants in the analysis.

Stimuli

Materials were selected from stimuli used in Polyn, Erlikhman, & Kahana (2011). In their study, forty-nine categories were chosen from the wordpools developed by Battig & Montague (1969) and Van Orden et al. (2003). Polyn and colleagues (2011) developed the category norms by asking participants to freely generate exemplars from a given category. Based on participants' responses from the norming experiment, highly common (generated by over 50% of the participants) and highly unusual (generated by three or fewer people) category exemplars were excluded from the wordpool to limit memory bias from semantic frequency. From this set, study lists for our current experiment were generated through randomly sampling a list of 25 category-target word pairs with several restrictions. First, each study list contained seven filler and 18 critical items, each of which was sampled from a unique category. Furthermore, each critical item's target word started with a unique first letter such that the 25 target words on the study list consisted of 19 or more unique first letters.

Design

We implemented a within-subject design with two conditions, selective retrieval practice and control. The experiment was divided into four blocks, two of which were randomly selected to include selective retrieval practice in between study and free recall; the rest were assigned to control conditions. During each block, participants studied a list of 25 categorized word pairs. Five filler items appeared at the beginning (Study 01 - 05), followed by 18 critical items (Study 06 - 23), and the list ended with two filler items (Study 24 - 25). These filler word pairs were placed in order to control for recency or primacy effects during later recall. While the seven filler items were excluded from data analyses, distinction between the critical and filler items was not revealed to participants.

Procedure

A visual overview of the procedure is shown in Figure 4.1. Participants were first given a brief overview of the experiment and completed a short practice round consisted of a study, retrieval practice, and a final test trial with three word pairs that were not used in the experiment. The main experiment began with the study phase, which included 25 category-exemplar pairs. Each word pair was presented on the screen for five seconds with one interstimulus interval. On each screen, each item's category membership was presented on top in small uppercase letters and its exemplar was presented in large lowercase letters. Participants were instructed to focus on the exemplars as they would be asked to recall those for a later memory test.

A 5-min distraction period followed the study phase, during which participants played a video game (Tetris) on the computer. The key manipulation occurred during the next phase. During the blocks assigned for selective retrieval condition, participants were prompted to recall 7^{th}, 13^{th}, and 19^{th} items they saw from the previous study list (selective retrieval practice). The first letter of each target item appeared as a recall cue along with blanks corresponding to its word length (e.g. "w_ _ _ _ _" as a retrieval cue for "willow"). Participants were allowed to spend as much time as they needed to respond and no feedback was given. After completing all three selective retrieval practice prompts, participants played Tetris again for five minutes before moving onto the final phase of the experiment. During the control condition blocks, participants played Tetris for ten minutes. At the end of each block, participants were given a free recall test in which they were asked to recall all of the exemplars they had studied within the block in any order.

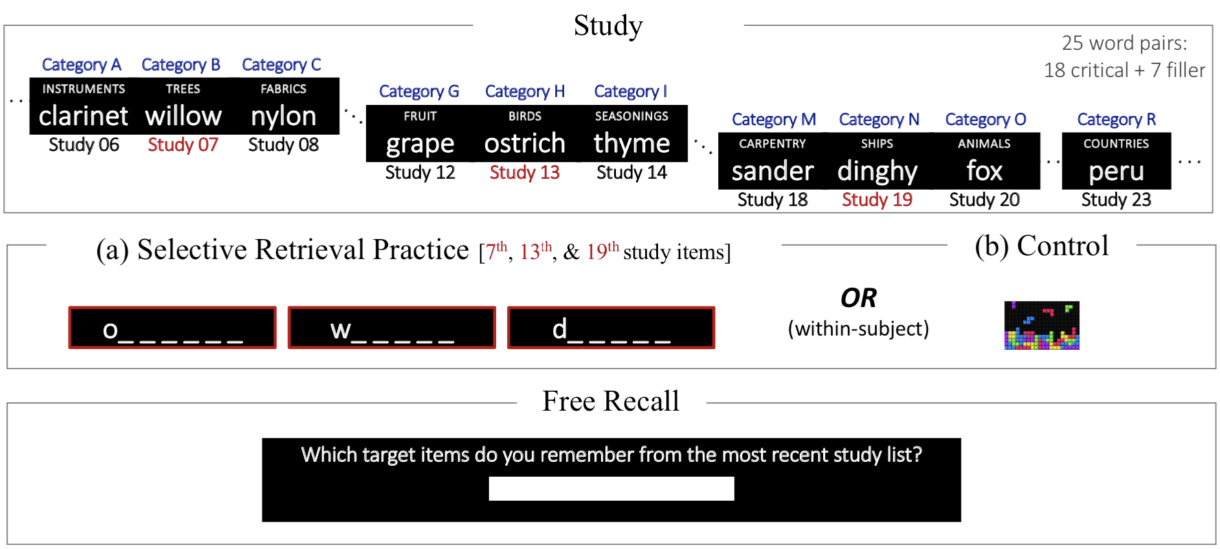

Figure 4.1: A schematic depiction of a single block from Experiment 5. In each block, participants studied a total of 25 words (18 critical items). Participants were given one of two tasks based on the assigned condition within each block. (a) Selective retrieval practice condition: after five minutes of tetris, participants were prompted to recall 7^{th}, 13^{th}, and 19^{th} items from the previous study list. No feedback was given and participants played tetris again for five minutes after answering all three retrieval practice prompts. (b) Control condition: participants played tetris for ten minutes. Each block ended with a free recall phase.

4.2.2 Simulated results

Using a modified version of the Context Maintenance and Retrieval (CMR) model, we simulated the recall probabilities for the selective retrieval and control condition. In CMR, there are two key representations: a feature layer, and a context layer. These two layers are interconnected to reflect how context is integrated with item features during encoding. Simply put, when an item is presented during study, its feature representation is activated and at the same time, the current state of context is retrieved.

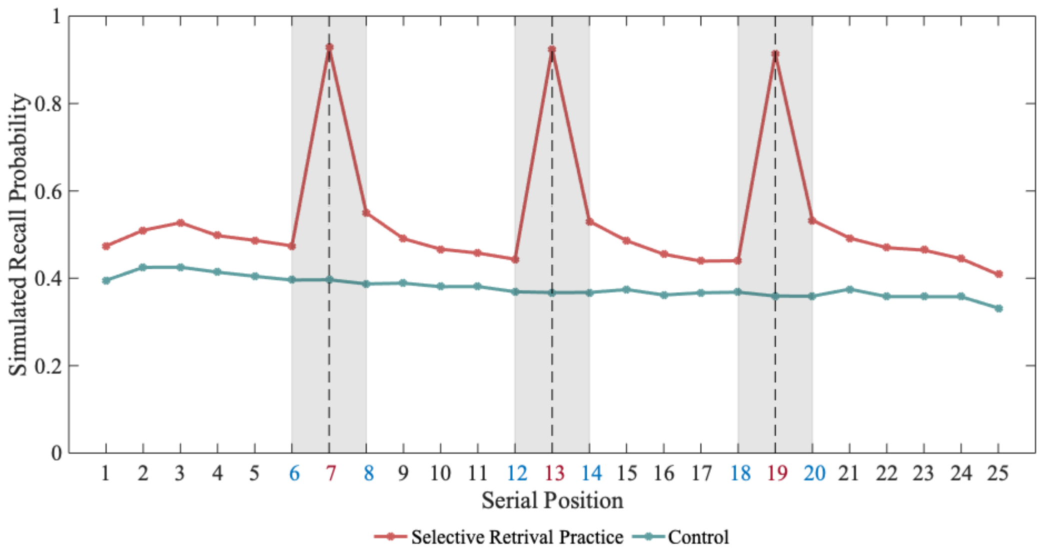

Figure 4.2: Simulated results from CMR illustrating recall probabilities for each studied item. 7th, 13th, and 19th items from the study list are selected for retrieval practice in the experimental condition (in red). This simulation shows that the selective retrieval practice of three items will boost memory for all items in teh study list. Furthermore, it shows that retrieved-context neighbors in the forward direction (+1) will benefit the most from the retrieval practice event.

Shown below in Figure 4.2 is a simulated recall probabilities for each studied item during free recall. Because the model parameters were set to assume high levels of learning from retrieval practice (i.e. high activation of the feature representation and retrieval of the context), our simulated results suggested an overall boost in recall for all items in the selective retrieval practice condition relative to the control condition. Consistent with predictions based on testing effect, simulated results displayed high recall rates for items selected for retrieval practice (7^{th}, 13^{th} and 19^{th} items).

Furthermore, our simulated results implied that the retrieved context neighbors in the forward direction (+1 transition, 8^{th}, 14^{th}, and 20^{th} items) would benefit the most from selective retrieval. This is most likely due to the previously mentioned forward-direction bias observed in the contiguity effect (Polyn, Norman, and Khana, 2009), a tendency to recall events from the past in a temporally forward manner. This phenomenon is explained to be due to the nature of our memory search process in which recalling an event from the past involves traveling back to the study episode and searching for the target item through the temporal sequence of events. In order to take this forward asymmetry into account, we pre-registered a secondary analysis in which the recall rates for the forward (+1) and backward (-1) neighbors are separately examined.

4.2.3 Results

To preview, we first examined retrieval practice success in the selective retrieval condition. Then we compared participants' free recall data from the selective retrieval condition to control condition and examined the following: free recall of all critical items, memory target items (items selected for retrieval practice), and memory for temporal neighbors of the target items (items presented right before or after those selected for retrieval). Furthermore, as mentioned above, our preliminary modeling of the retrieval practice effect on temporal neighbors led us to believe that the effect may be stronger in the forward direction. Thus, we examined differences in recall for only the forward temporal neighbors.

Selective retrieval practice phase

Recall that only three target items out of 18 critical items were selected for retrieval practice. On average, selective retrieval was successful 36% of the trials (SD = .26).

Free recall

Recall that the first five and last two items presented during study were excluded from analyses in order to control for the primacy and recency effects. Participants' memory for 18 critical items in the selective retrieval condition ($M = .33, SD = .21$) did not differ from that in the control condition ($M = .30, SD = .20$), $t(41) = 1.28, p = .208, d = 0.20$, 95% $CI[-0.02, 0.09]$.

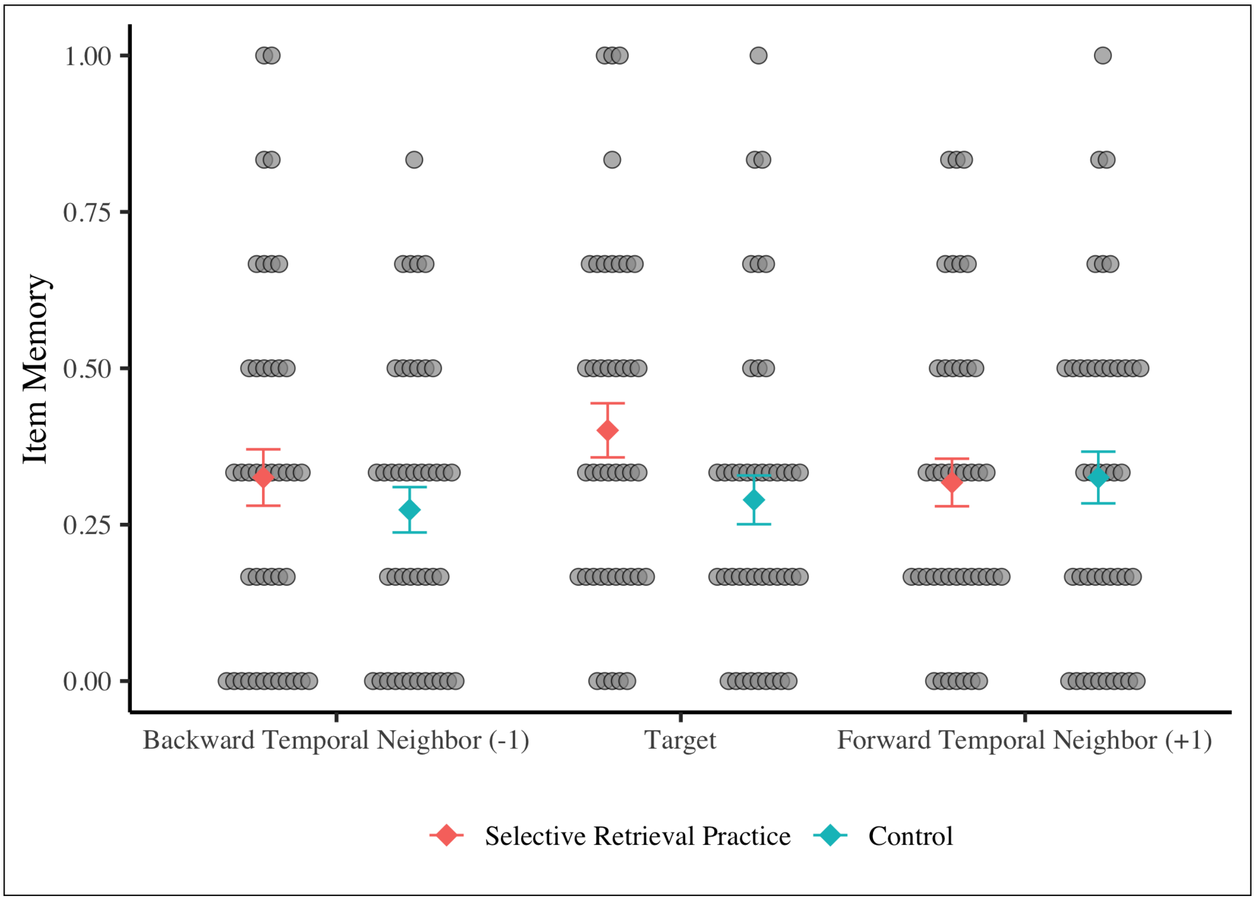

Figure 4.3: Proportion of backward neighbor, target, and forward neighbor items recalled during free recall from Experiment 5. Each dot represents one participant. For each criteria, participants recalled a minimum of zero (0%) and maximum of six items (100%): three backward neighbors (6^{th}, 12^{th}, and 18^{th}), targets (7^{th}, 13^{th}, and 19^{th}), and forward neighbors (8^{th}, 14^{th}, and 20^{th}) per block, two blocks total. For the target items, items that underwent selective retrieval practice were recalled significantly better than those in the control condition. However, selective retrieval practice did not result in better or worse item memory for both backward and forward temporal neighbors of the target items.

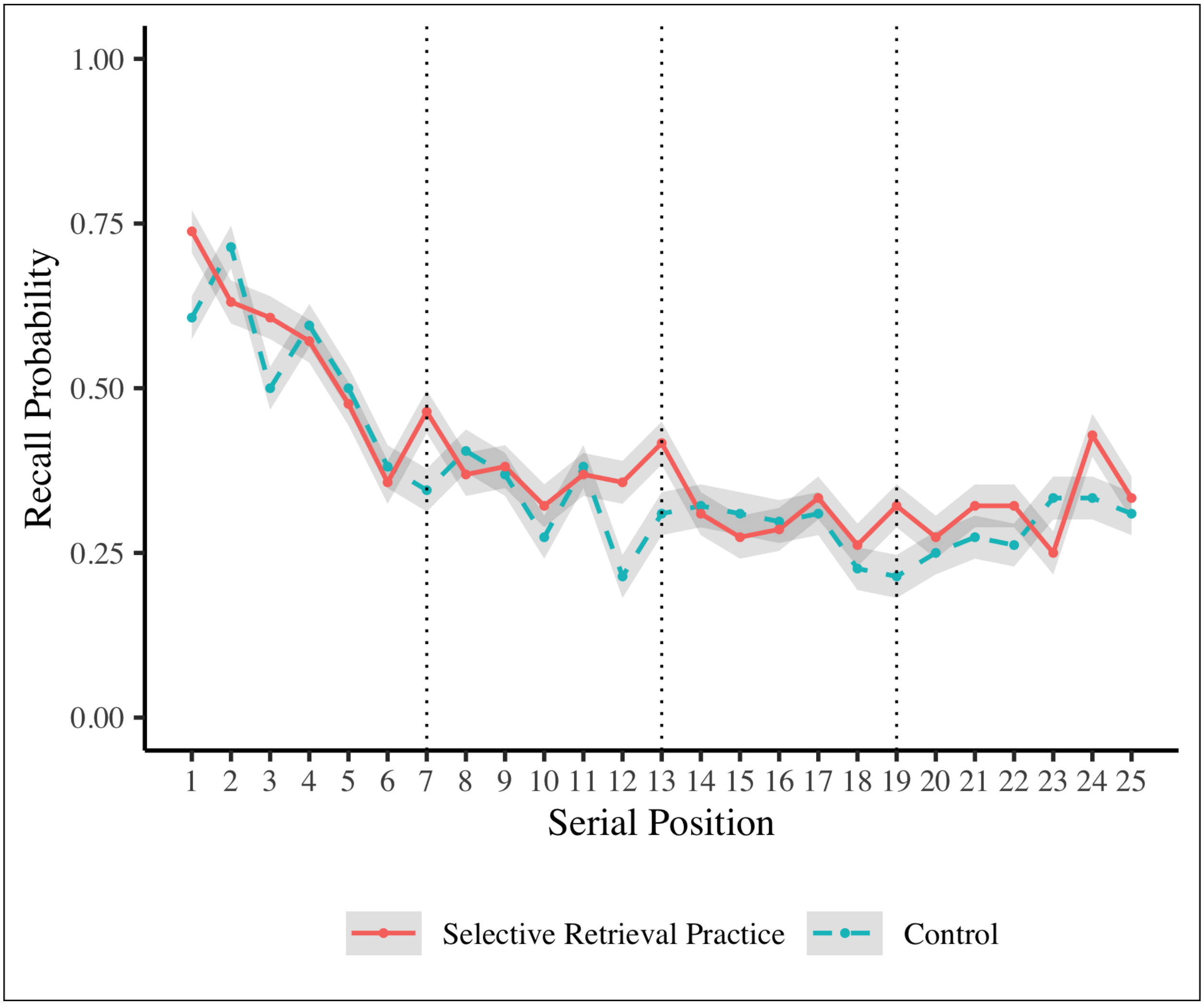

Figure 4.4: Recall probability as a function of serial position

Memory for target items

As expected, selectively practice recalling the target items – 7th, 13th, and 19th – words on the study list – improved participants' ability to later recall those words ($M = .40, SD = .28$) compared to the control condition ($M = .29, SD = .25$), $t(41) = 2.58, p = .014, d = 0.40$, 95% $CI[0.02, 0.21]$.

Memory for temporal neighbors of target items

Based on the assumptions from the episodic context account, retrieving a target item requires reinstating the associated context and (as a result) strengthening it. Since the temporal context tied to a retrieved item may also be associated with its immediate temporal neighbors: because temporal context changes slowly over time (Figure 4.3) and items studied close in time share a highly similar context. In our paradigm, the episodic context account would extend to predicting enhanced memory for temporal neighbors of the retrieved item whose associated context is highly similar. Thus, we compared memory for immediate temporal neighbors of target (7th, 13th, & 19th) items: 6th, 8th, 12th, 14th, 18th, and 20th items.

Contrary to our predictions based on the episodic context account, we found no differences between the selective retrieval condition ($M = .32, SD = .23$) than in the control condition ($M = .30, SD = .22$), $t(41) = 0.67, p = .507, d = 0.10$, 95% CI [-0.04, 0.09]. A Bayesian analysis with the same parameters as in previous experiments found a moderate support for the null hypothesis $BF_{01} = 5$.

Memory for forward temporal neighbors of target items

As mentioned above, it is also possible indirect benefits from temporal context reinstatement could be limited to forward temporal neighbors. Therefore, compared memory for immediate forward temporal neighbors (8th, 14th, and 20th) of the selected target items (7th, 13th, and 19th). Similarly from above, we found no differences between the selective retrieval condition (M = .32, SD = .25) than in the control condition (M = .33, SD = .27), $t(41) = 0.17, p = .869, d = 0.03$, 95% $CI[-0.10, 0.09]$. A Bayesian analysis with the same parameters as in previous experiments found a moderate support for the null hypothesis $BF_{01} = 5.88$.

Conditionalized analyses based on retrieval practice success

Memory for target items

Similarly from Experiments 1, 2, 3, and 4, in order to examine the effects of successful retrieval, we conducted an exploratory analyses in which we conditionalized participants' performance upon successful retrieval practice trials; that is, given that an item was successfully recalled during selective retrieval practice, what's the probability of recalling that the target item during final free recall? Because each participant had two of retrieval practice blocks and two of control blocks, we treated each block as yoked pairs. For example, if a participant successfully retrieved all two items (7^{th} and 13^{th}) during the first retrieval practice block and only retrieved one (19^{th}) of the items during the second retrieval practice block, their first control block was conditionalized to only include the same two items and the same single item for the second control block. By doing so, we aimed to control for biases that stemmed from failure to attention during study. Further confirming the presence of testing effect, participants' memory for target items conditionalized upon successful practice in the selective retrieval practice condition ($M = .72, SD = .37$) was significantly better than in the control condition ($M = .33, SD = .25$), $t(34) = 6.09, p < .001, d = 1.03$, 95% CI [0.26, 0.53].

Memory for temporal neighbors of target items

Similarly from above, we conditionalized participants' memory for forward and backward temporal neighbors of items that underwent successful retrieval practice. Once again, Contrary to our predictions based on the episodic context account, we did not find evidence for better recall of temporal neighbors from the selective retrieval practice condition ($M = .40, SD = .34$) as compared to the same items in the control condition ($M = .31, SD = .24$), $t(34) = 1.50, p = .144, d = 0.25$, 95% CI [-0.03,0.19].

Memory for forward temporal neighbors

We also examined conditionalized scores for forward temporal neighbors alone to account for the aforementioned forward asymmetry during recall. We again found no differences in participants' memory for the forward neighbors in the selective retrieval condition ($M = .39, SD = .38$) than in the control condition ($M = .34, SD = .31$), $t(34) = 0.63, p = .536, d = 0.11$, 95% CI[-0.11, 0.20].

Temporal organization during free recall

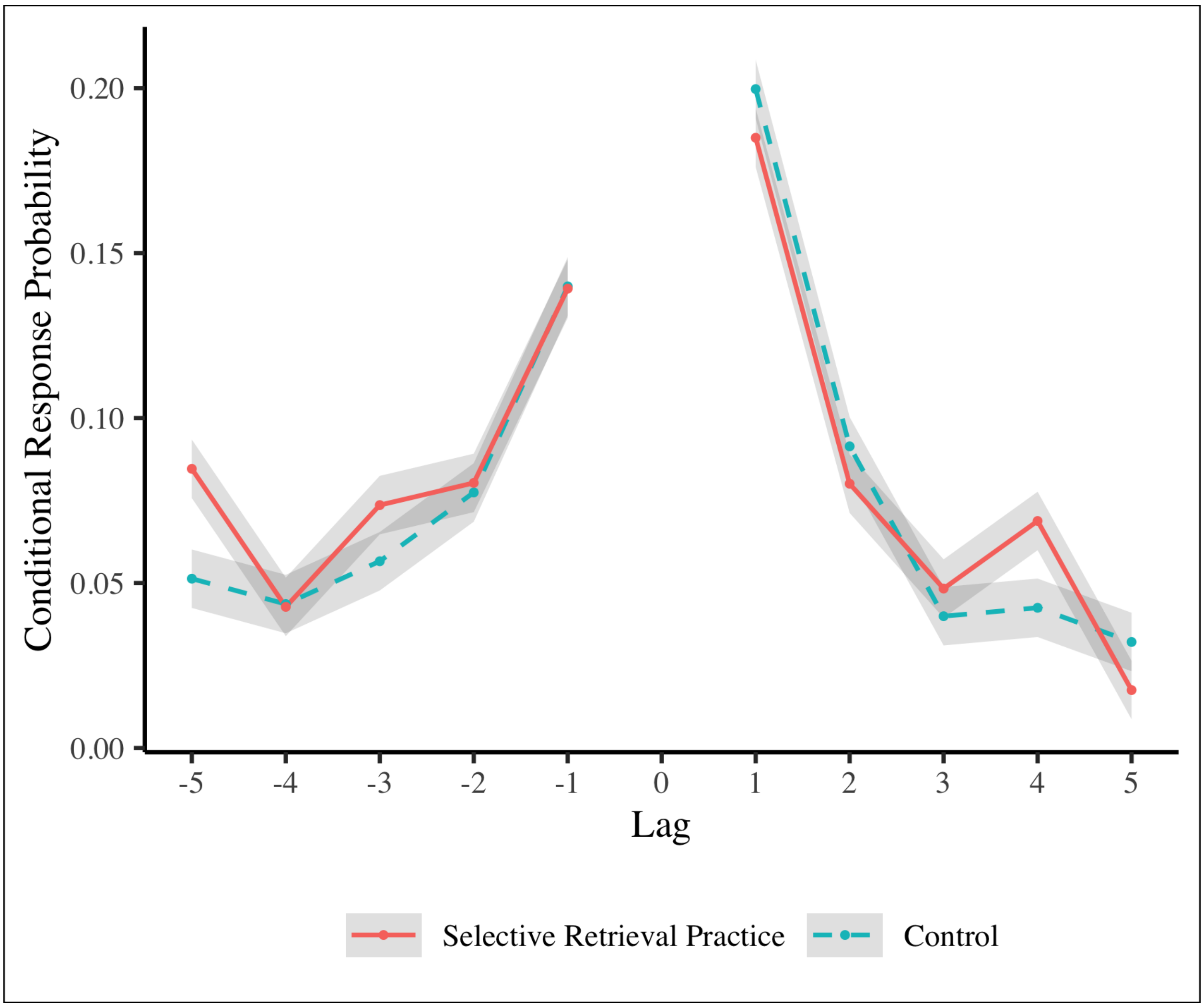

Figure 4.5: Temporal organization during free recall from Experiment 5.

Another measure of temporal context reinstatement is looking at temporal contiguity during free recall, a tendency for participants to recall items that were presented in nearby positions during the study phase. If temporal context is reinstated during retrieval practice, participants in the selective retrieval practice would exhibit stronger temporal contiguity for items around the target items. As a result, during the final free recall, we predicted that selective retrieval condition would lead grouping recalls by their temporal order around the target item (e.g., recalling 6^{th}, 7^{th}, then 8^{th} items or 12-13-14, or 18-19-20). However, as shown in Figure 4.5, no discernible differences in temporal contiguity was observed between the two conditions. Contrary to our prediction that participants would make more +1 and -1 transitions after retrieval practice as they group their recalls around the practiced target item (e.g., higher probability of recall of 6^{th} and 8^{th} item after recalling 7^{th} item in the retrieval practice condition), the probability of recalling a +1 and a -1 temporal neighbor in both conditions were nearly identical.

4.2.4 Discussion

In Experiment 5, we tested whether retrieval practice leads to reinstatement and strengthening of temporal context. Unlike Experiments 1, 2, 3, and 4, we examined the strength of context memory from the neighbors of retrieval practiced items based on the nature of temporal context that it changes slowly over time (Siegel & Kahana, 2014): if an item's associated temporal context is reinstated upon retrieval practice, its temporal neighbors – an item that was studied right before or after the retrieval practiced item – are also likely to be cued into memory because their associated temporal context is highly similar to the one that was reinstated and strengthened. In sum, we tested for temporal context reinstatement by examining whether the temporal neighbors of a retrieval practiced item indirectly benefits from retrieval practice. Similarly from Experiments 1, 2, 3, and 4, we did not find any evidence for reinstatement and strengthening of an item's context memory upon retrieval.

4.3 Experiment 6

Due to COVID-19 pandemic, data collection activities were interrupted and therefore we were unable to conduct Experiment 6. Therefore, the current section only includes procedures that we had planned for Experiment 6 and the predicted results.

4.3.1 Procedure

A visual overview of the procedure is shown in Figure 4.6.

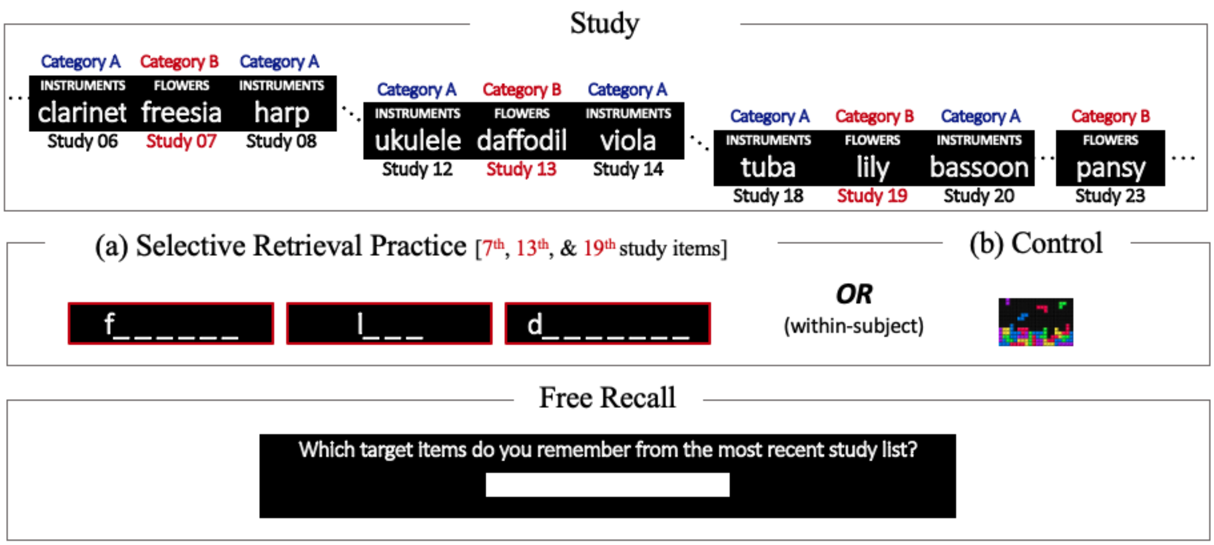

Figure 4.6: A schematic depiction of a single block from Experiment 6. In each block, participants will study a total of 25 words (18 critical items, 9 items from each of two categories, A and B). Participants will be given one of two tasks based on the assigned condition within each block. (a) Selective retrieval practice condition: after five minutes of tetris, participants will be prompted to recall three exemplars from category B, 7$^{\text{th}}$, 13$^{\text{th}}$, and 19$^{\text{th}}$ items from the previous study list. No feedback was given and participants played tetris again for five minutes after answering all three retrieval practice prompts. (b) Control condition: participants played tetris for ten minutes. Each block ended with a free recall phase.

4.3.2 Predicted results and analysis plan

Data analysis in Experiment 6 will follow the steps from Experiment 5 above.

Selective retrieval practice phase

As observed in Experiment 5, we aim to observe 30-40% accuracies during the selective retrieval practice phase such that participants successfully complete at least one or more items. If our pilot participants fail to accurately recall the target word for more than 50% of the time on average, we plan to enhance the retrieval cue by either (1) providing the first two letters of the target word or (2) providing the category membership along with the first-letter cue.

Memory for temporal neighbors of target items

We have two different predictions based on which type of context may be reinstated. First, if temporal context is reinstated during retrieval practice, the predicted results are the same as what we predicted in Experiment 4 in which temporal neighbors (in the forward direction) benefit from the retrieval practice event. What if *semantic* context is reinstated during retrieval practice instead of temporal context? In such a case, we predict the context reinstatement from selective retrieval of the three category B items will benefit participants' memory for the non-practiced category B items. Conversely, based on the retrieval-induced forgetting theory, it is also possible that memory non- practiced B category items will be suppressed as a result of selective retrieval practice of the three category B items. In such a case, we expect to observe the opposite where the non-practiced B category items show lower recall rates than other A category items in the same condition.

4.3.3 Discussion

Experiment 6 is designed to test the reinstatement of temporal or semantic context during retrieval. It is possible that in Experiment 6, we would not find any evidence for neither temporal nor semantic context reinstatement during retrieval based on our consistent findings from Experiments 1-5 where no evidence of context reinstatement was found. However, if semantic context is reinstated upon retrieval, we predict that participants' memory for semantic neighbors of retrieved items will be reduced after retrieval practice based on past findings from the retrieval-induced forgetting literature.

CHAPTER 5

General Discussion

5.1 Summary of Findings

In the last several decades, researchers have focused on examining the benefits of retrieval practice on memory. This extensive work has given us insights into the robustness of the effect and its universality across various experimental manipulations such as encoding modalities (e.g., auditory, visual), stimuli (e.g., foreign language vocabulary, paired associates, nonverbal stimuli), and study designs (e.g., delay length, with or without feedback after retrieval practice). Despite the wealth of evidence, the underlying mechanism remains unknown. As reviewed previously in Chapter 1, most proposed theories only provide conceptual explanation of the effect instead of a mechanistic one. The episodic context account by Karpicke, Lehman, and Aue (2014) is one that comes closer to specifying the mechanism by explaining the effect based the role of episodic context during memory search and recall: benefits of retrieval practice stem from *reinstating* and *strengthening* the context associated with an item upon retrieval. From a series of experiments, we directly tested predictions from the episodic context account: does retrieval practice strengthen memory for context?

In Experiments 1, 2, 3, and 4, we addressed this question by examining context memory from items that underwent successful retrieval practice. Participants studied 40 paired associates (e.g., Flower - pansy, average association strength between a cue and its target= 0.2) presented in one of four different colors. After study, half of the participants were prompted to recall the correct target for a given recall cue (e.g., pansy for Flower - _a_nsy) while the other half were asked to restudy the same paired associates. Importantly, both the retrieval practice cues and restudy items were displayed to participants without colors (white letters on a black background). In addition, while participants were informed about the item recall test as we instructed them to study each item carefully to prepare for a memory test at the end, they were not made aware of the color memory test. We compared participants' memory for source context (i.e., font colors) after retrieval practice and restudy of target items. Participants' recall performance for item's font colors (Figure 2.2) suggested that participants in both conditions showed some level of context memory that was significantly above the accuracy at chance-level (0.25 or 1/4). Contrary to our prediction based on the episodic account, however, participants' memory for the original study context after retrieval practice did not differ from that of restudy. More importantly, despite the absence of benefit toward context memory from retrieval practice, we still observed the testing effect: recalls were enhanced for items that underwent retrieval practice than restudy, even though memory of their associated source context did not differ between the two conditions. Based on our results from Experiments 1, 2, 3, and 4, we concluded that if the episodic context account is true, context reinstatement may be limited to features that are more inherent to encoding of an item such as semantic or temporal context.

Therefore, in Experiment 5, we tested for reinstatement and strengthening of temporal context during retrieval. Based on findings from the retrieval-induced facilitation literature, our main aim was to examine potential indirect effects of context reinstatement from retrieval practice. Temporal context changes slowly over time such that items studied near-by in time share a similar temporal context. If an item's temporal context is reinstated during recall, the item's temporal neighbors whose associated context is highly similar could also be cued into memory. This would produce an indirect testing effect or what had been observed as retrieval-induced facilitation of non-tested items (Chan et al., 2006). On the other hand, reinstating and strengthening the temporal context upon item recall could suppress one's memory for items tied to similar context as the cue is now strongly associated with a single item. Therefore, in Experiment 5, we tested for the presence of retrieval-induced benefit or forgetting from temporal context reinstatement. Similarly from Experiments 1, 2, 3, and 4, while we observed the testing effect for practiced items, we found no evidence of benefit or impoverishment from temporal context reinstatement.

5.2 Theoretical Implications

Results from Experiments 1, 2, 3, 4, and 5 suggest some limitations and constraints on the episodic context account proposed by Karpicke and colleagues (2014). Their account suggests that the benefits of recall practice stem from reinstating the episodic context of the original study event. Successful retrieval strengthens the context memory associated with the target item, and the retrieved context is updated to include features of the recall context. Because this strengthened and updated context memory becomes an effective retrieval cue for later recall attempts, items that receive retrieval practice become more memorable than those that were simply restudied.

While this account provides a tenable explanation as to why retrieval practice benefits memory, the account does not specify the type of information reinstated from the episodic context. Context, in general, is a broad concept, and can refer to many aspects of the encoded event including characteristics of the target item itself such as typeface, stimuli color, and mode of presentation, as well as other features of the episode such as the testing room, outside noises, one's mood or internal thoughts. In the current studies, we found no evidence that contextual details of the studied words themselves were strengthened following retrieval even though these contextual details were part of the encoding process and were remembered at above chance levels during the color memory test. It remains possible, however, that a more limited version of the episodic context account may be correct; one in which some but not all aspects of the episodic context are reinstated during retrieval practice.

5.2.1 Context Reinstatement during Retrieval: Only *Some*, but not All

Predictions from memory models as well as well-observed memory phenomena such as retrieval-induced forgetting and facilitation suggest that *some* context is reinstated during retrieval. However it is unclear exactly *what* type of context it is or *why* only those are chosen to be reinstated: Does it have to be an inherent characteristic such as an item's meaning? Or can it be a randomly assigned context within the experimental context? While it is not yet possible to define the key predictors toward context reinstatement given limited evidence, listed below are studies that did show reactivation of a specific type of context reactivation.

Relevance of Context

It is possible that reinstatement of context is beneficial only when the reinstated context features are relevant to the original encoding event. Supporting this idea is a recent study by Shin, Masís-Obando, Keshavarzian, Davé, & Norman (2020). As mentioned briefly in Chapter 1, the authors showed that context-dependent memory effect is observed only when the associated context is relevant to the encoding task. Replicating the main idea from a seminal study by Godden & Baddeley (1975) showing that words learned underwater were recalled better when tested underwater than on the land, Shin and colleagues (2020) used virtual reality environments to test if participants recalled more words when the context during testing matched the context during encoding. Participants were placed in one of two virtual environments – underwater or on Mars – in which they were presented with a total of 48 items. Upon encountering each item, participants were asked to judge whether the item would be useful for survival in the given environment. Either immediately or 24 hours after the study phase, participants were asked to recall all of the items they studied throughout the experiment. This retrieval test took place in one of two virtual environments such that some participants recalled the items in the same environment they had studied them while others were placed in a different environment. As expected, the results showed that items were better recalled when the context at retrieval and study matched. A critical finding from the study was that the context-dependent memory effects were only obtained for items that were judged to be useful for survival in the encoding environment. In other words, reactivating an item's associated study context during testing facilitated retrieval only when the item was deemed useful in the context it was presented in. These findings are consistent with the idea of incidental vs. intentional context from Bjork & Richardson-Klavehn (1989) mentioned above in the interim discussion (Chapter 2.5) as well as the encoding specificity principle (Chapter 1.2.1). In sum, these studies point out that reinstatement of context facilitates recall only when the reinstated context is relevant to the to-be-recalled item.

Spatial Context

Spatial context is one type of context that has been shown to be integrated and retrieved regardless of its relevance to the encoding task. Akan, Stanley, & Benjamin (2018) showed that participants had improved memory for each studied item's original spatial context (i.e. where each word was presented on the screen during study) after retrieval practice compared to restudy. Here, similar to our Experiments 1 – 4, participants were unaware that they would be asked to recall the locations of items as the retrieval practice task only involved item retrieval. Furthermore, the spatial location for each item were randomly assigned and were not relevant to any key characteristics of an item. Yet, contrary to our results, Akan, Stanley, & Benjamin (2018) did find strong evidence toward spatial context reinstatement and strengthening through retrieval practice.

Spatial memory is a unique type of context that has been shown to be encoded automatically in a change detection task (Beck, Angelone, Levin, Peterson, & Varakin, 2008). Like spatial context, it is possible that there are some context features that do not seem relevant to an item or the study event, but are in fact, integral to driving the benefits of retrieval practice on memory. Therefore, future studies should be designed to test the boundaries of relevance of a context feature to an item.

5.2.2 Alternative Theories of the Testing Effect

Given our findings, it is important to consider an alternative possibility that the retrieval practice advantage is in fact, unrelated to the successful retrieval of episodic contextual details. It could be that episodic contextual information from the study phase is sometimes recalled during the practice phase, but this has no bearing on the enhanced memorability of items that are successfully retrieved during practice. By this account, certain episodic contextual details (such as spatial position or list context) may simply be details that are more likely to be retrieved despite not explicitly being cued, and other contextual details (such as voice gender or font color) are less likely to be retrieved spontaneously. In other words, while the observation of enhanced memory for source details is consistent with the episodic context account of retrieval practice, it is also consistent with alternative theories in which source details are sometimes retrieved, but are not mechanistically involved in the retrieval practice advantage.

As reviewed in Chapter 1.1, several accounts other than the currently tested episodic context account have been suggested to explain the testing effect. Can any one of these accounts explain what we found in our experiments? Predictions that can be drawn from each account toward our current experiment are limited, however, due to the shortcomings that were previously discussed above. First, the three conceptual accounts of the testing effect – retrieval effort hypothesis, bifurcation account, and dual memory theory – can neither be supported nor refuted with our current results as no clear predictions can be made in the first place. For instance, the retrieval effort hypothesis does not specify what the effort entails. In our current experiments, even if we measured the *effort* as the time each participant took to successfully complete each retrieval practice trial and supposedly found a positive correlation between trial duration and memory, we would still be left with the same question: what happens during a difficult retrieval trial and how does it improve memory? The same problem applies to the bifurcation account (boost in memory from retrieval practice) and dual memory theory the term boost in the bifurcation account needs to be specified in order for us to be able to make specific predictions about the mechanism. measure and compare memory strengths after retrieval practice and restudy. The dual memory theory simply posits encoding variability as an explanation for the testing effect. While our overall pattern of results are consistent with what would be predicted from the dual memory theory (better recall after retrieval practice than restudy), the theory itself has been challenged for its limited nature as discussed above.

The most viable prediction consistent with our findings can be drawn from the primary and convergent retrieval model proposed by Hopper and Huber (2019). The main difference between primary and convergent retrieval model and the episodic context account lies in their assumed learning mechanisms. Under the episodic context account, the benefits of retrieval practice stem from better match between the target memory and its contextual cues during recall. In the primary and convergent retrieval model, this would be more analogous to enhanced primary retrieval (i.e., associations between context and item features) rather than convergent retrieval (i.e., association between item features and other item features). Furthermore, as mentioned above, the episodic context account would predict faster recall after successful retrieval practice based on the assumption that the item is more likely to be sampled from a set of contextually appropriate memories (Lehman, Smith, & Karpicke, 2014; Rohrer and Wixted, 1994, 1993). Hopper and Huber (2018) showed that retrieval practice with the same cue as the final test produced faster recall while no latency benefits were found when cued recall practice used cues that were different from those used on the final test. Such finding would be analogous to ours if context (i.e., font color) was used as a retrieval cue during the final recall test (instead of final free recall). Furthermore, the lack of stronger memory for context could be explained by the fact that font color or temporal context were not contextually appropriate memory in Experiments 1, 2, 3, 4, and 5.

5.2.3 Cue Binding Hypothesis

Expanding on the primary and convergent retrieval model by Hopper & Huber (2019), one possibility is a cue-binding hypothesis. Here, benefits of retrieval practice do not depend solely on the reinstatement of its episodic context – instead, the key mechanism lies in the binding of the retrieved item to its recall cues, regardless of the cue type. In other words, a recall cue is defined as any piece of information that can effectively guide the search process and lead to the correct target memory. Upon repeated retrieval, the association between this recall cue(s) and the retrieved item gets stronger, leading to faster recall as observed in Hopper & Huber (2018). The recall cue could be related to an item's features such as its temporal, semantic, and spatial context; or, it could be something completely unrelated to the item's inherent characteristics but somehow forged during retrieval such as a self-generated thought that functions as a reminder (Zhang & Tullis, 2021). This hypothesis is also in line with findings suggesting that active retrieval promotes binding of the retrieved content with other episodic elements (Shimamura, 2011; Bridge & Voss, 2015; Akan et al., 2018). To sum up, it may not be as important whether an item's recall cue is part of its original episodic context or not: instead, it could be *any* feature, as long as it can be cohesively linked to other cues to activate the target memory.

If a recall cue can be anything, why do existing research focus mostly on the role of temporal or semantic context during memory search and retrieval? It is possible that these are most often examined as critical factors for recall since temporal and semantic context features are most commonly present features in a lab-based experiment. In most laboratory memory studies, participants are asked to sit behind a computer and study a list of words presented on the screen without any specific guidance on how they should study these items. When seated in a quite room with no changes other than what is presented on the screen, features that reflect an item's presentation sequence or its meaning may be the most salient or memorable features to remember those items by.

5.3 Limitations: Correlation, Not a Causation

It is important to note that even if we did find any evidence of context reinstatement by observing stronger memory for context in the retrieval practice condition, our current experimental design cannot draw causal relationships between context reinstatement and the testing effect. In fact, the same limitations apply to other studies that also directly tested the episodic context account. The aforementioned study by Akan, Stanley, & Benjamin (2018) found stronger memory for spatial context in the retrieval practice condition, but similarly to ours, the study was not designed to test the causal relation between context reinstatement and the testing effect. While the results were sufficient to conclude a correlation between retrieval practice and enhanced memory for spatial context, the findings cannot attest to whether benefits of testing stems from reinstating and strengthening the original study context. Therefore, In order to test the validity of the episodic context account, future studies need to be designed to prove that context reinstatement is the *cause* of the testing effect.

Furthermore, even if the causal links were already established, the original description of the account needs to be revised to clarify the details on context reinstatement and update. While The original account mentions strengthening of the initial study context and updating it to include context at the time of recall, it is unclear how much of the *update* happens: does the memory for initial study context always remain the strongest? Or, will it be overwritten? Descriptions of the account from Karpicke, Lehman, and Aue (2014) infer that an item's association to its initial study context and context at retrieval would be equally strong. Furthermore, the authors also note that through repeated retrieval, an item would eventually be de-contextualized such that it is not longer tied to one specific context. While the description sounds plausible and is consistent with the idea of consolidation Tse et al. (2007), the authors do not make any predictions or make references to how decontextualization could then contribute to the consolidation processes. Furthermore, from the current version of the account, it is also unclear how memory for context is distributed as a function of retrieval frequency (i.e., would a participant have equally strong context memory for every retrieval event? If so, would it create any interference?) and how many retrieval attempts it may take to de-contextualize an item.

5.4 Conclusion and Future Directions

To master a new piece of music, musicians spend hours repeatedly practicing the same measures: over time, transitions that sounded disparate at first become a cohesive phrase as they become ingrained in one's muscle memory. Similarly, repeatedly recalling something from memory makes it stronger and become easier to recall it later. But *how* does this happen? Based on the role of temporal context as a guide for search and retrieval described in retrieved-context models of memory (Howard & Kahana, 2002), the episodic context account (Karpicke, Lehman, & Aue, 2014) suggest that the benefits of retrieval stem from reinstating and strengthening the original episodic context. In Experiments 1, 2, 3, 4, and 5, we tested a direct prediction from this account: can retrieval practice enhance memory for context?

From Experiments 1, 2, 3, and 4, we tested for reinstatement and strengthening of one type of episodic context, a source context using different font colors. While we observed enhanced memory for items after retrieval practice, this benefit was not accompanied by a stronger memory of the associated source context as compared to the restudy condition. In Experiment 5 and 6, we aimed to test the effects of retrieval practice on an item's original neighbors and semantic associates from the study list. In Experiment 5, we focused on temporal context: if retrieval practice of an item reinstates the temporal context associated with the item's local neighbors on the list, then the benefits of retrieval practice should spread to these neighbors as well. Our results from Experiment 5, however, did not find any support for the hypothesis that benefits of retrieval practice stem from the reinstatement of its original temporal context.

In conclusion, we have reliably shown that the extent of reinstatement and strengthening of the original episodic context may be a lot more limited than what was suggested in the original description of the account (Karpicke, Lehman, & Aue, 2014). Given that we still observed the testing effect in the absence of stronger context memory for source and temporal information, we suggest that perhaps only some, but not all aspects of the episodic context is reinstated during retrieval. As discussed previously, future research should focused on identifying the key features or characteristics of recall cues that are reinstated through retrieval. Furthermore, studies should be designed to test and establish the causal link between benefits of retrieval practice and context reinstatement in order to identify the true cognitive mechanism underlying the effect.

References

Abbot, E. E. (1909). On the analysis of the memory consciousness in orthography. *The Psychological Review: Monograph Supplements*, *11*(1), 127–158. doi: 10.1037/h0093013

Akan, M., Stanley, S. E., & Benjamin, A. S. (2018). Testing enhances memory for context. *Journal of Memory and Language*, *103*, 19–27. doi: 10.1016/j.jml.2018.07.003

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. *Journal of Experimental Psychology Monograph*, *80*(3 Pt. 2), 1–46.

Bäuml, K. H. (2002). Semantic generation can cause episodic forgetting. *Psychological Science*, *13*(4), 356–360. doi: 10.1111/j.0956-7976.2002.00464.x

Bäuml, K. H., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory and Cognition*, *32*(4), 610–617. doi: 10.3758/BF03195852

Bäuml, K. H., Aslan, A., & Abel, M. (2017). The Two Faces of Selective Memory Retrieval—Cognitive, Developmental, and Social Processes. *Psychology of Learning and Motivation*, *66*, 167–209. doi: 10.1016/BS.PLM.2016.11.004

Beck, M. R., Angelone, B. L., Levin, D. T., Peterson, M. S., & Varakin, D. A. (2008). Implicit learning for probable changes in a visual change detection task. *Consciousness and Cognition*, *17*(4), 1192–1208. doi: 10.1016/j.concog.2008.06.011

Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.), *Current issues in cognitive processes: The {tulane flowerree} symposium on cognition* (chap. 9). New Jersey: Lawrence Erlbaum Associates.

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

Bridge, D. J., & Voss, J. L. (2015). Binding among select episodic elements is altered via active short-term retrieval. *Learning and Memory*, *22*(8), 360–363. doi: 10.1101/lm.038703.115

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. doi: 10.1037/a0017021

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. doi: 10.1037/a0024140

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory and Cognition*, *34*(2), 268–276. doi: 10.3758/BF03193405

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

Chan, J. C., McDermott, K. B., & Roediger, H. L. (2006). *Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material* (Vol. 135) (No. 4). doi: 10.1037/0096-3445.135.4.553

DuBrow, S., Rouhani, N., Niv, Y., & Norman, K. A. (2017). *Does mental context drift or shift?* (Vol. 17). doi: 10.1016/j.cobeha.2017.08.003

Gardiner, F. M., Craik, F. I., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, *1*(3), 213–216. doi: 10.3758/BF03198098

Gates, A. (1917). Recitation as a factor in memorizing. *Archives of Psychology*(40), 104.

Geiselman, R. E., & Bjork, R. A. (1980). Primary versus secondary rehearsal in imagined voices: Differential effects on recognition. *Cognitive Psychology*, *12*(2), 188–205. doi: 10.1016/0010-0285(80)90008-0

Glover, J. A. (1989). The "Testing" Phenomenon: Not Gone but Nearly Forgotten. *Journal of Educational Psychology*, *81*(3), 392–399. doi: 10.1037/0022-0663.81.3.392

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and under water. *British Journal of Psychology*, *66*, 325–331.

Healey, M. K., Crutchley, P., & Kahana, M. J. (2014). Individual differences in memory search and their relation to intelligence. *Journal of Experimental Psychology: General*, *143*(4), 1553–1569. doi: 10.1037/a0036306

Healey, M. K., Long, N. M., & Kahana, M. J. (2019). *Contiguity in episodic memory* (Vol. 26) (No. 3). Springer New York LLC. doi: 10.3758/s13423-018-1537-3

Healey, M. K., & Uitvlugt, M. G. (2019). The role of control processes in temporal and semantic contiguity. *Memory and Cognition*, *47*(4), 719–737. doi: 10.3758/s13421-019-00895-8

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning \& Verbal Behavior*, *10*(5), 562–567.

Hopper, W. J., & Huber, D. E. (2018). Learning to recall: Examining recall latencies to test an intra-item learning theory of testing effects. *Journal of Memory and Language*, *102*, 1–15. doi: 10.1016/j.jml.2018.04.005

Hopper, W. J., & Huber, D. E. (2019). Testing the primary and convergent retrieval model of recall: Recall practice produces faster recall success but also faster recall failure. *Memory and Cognition*. doi: 10.3758/s13421-019-00903-x

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, \& Cognition*, *25*, 923–941.

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*, 269–299.

Jonker, T. R., Dimsdale-Zucker, H., Ritchey, M., Clarke, A., & Ranganath, C. (2018). Neural reactivation in parietal cortex enhances memory for episodically linked information. *Proceedings of the National Academy of Sciences*, 201800006. doi: 10.1073/pnas.1800006115

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory \& Cognition*, *24*, 103–109.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775. doi: 10.1126/science.1199327

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-Based Learning: An Episodic Context Account. In *Psychology of learning and motivation - advances in research and theory* (Vol. 61, pp. 237–284). Academic Press Inc. doi: 10.1016/B978-0-12-800283-4.00007-1

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. doi: 10.1016/j.jml.2011.04.002

Kubit, B. M., & Janata, P. (2021). Spontaneous mental replay of music improves memory for incidentally associated event knowledge. *Journal of Experimental Psychology: General*. doi: 10.1037/xge0001050

Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, *120*(1), 155–189. doi: 10.1037/a0030851

Liu, X. L., & Ranganath, C. (2021). Resurrected memories: Sleep-dependent memory consolidation saves memories from competition induced by retrieval practice. *Psychonomic Bulletin & Review*, 1–10. doi: 10.3758/s13423-021-01953-6

McDaniel, M. A., Cahill, M., Bugg, J. M., & Meadow, N. G. (2011). Dissociative effects of orthographic distinctiveness in pure and mixed lists: an item-order account. *Memory \& Cognition*, 1–12.

McDermott, K. B. (2021). *Practicing Retrieval Facilitates Learning* (Vol. 72). Annual Reviews Inc. doi: 10.1146/annurev-psych-010419-051019

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Memory*, *9*, 596–606.

Mensink, G. J., & Raaijmakers, J. G. (1988). A Model for Interference and Forgetting. *Psychological Review*, *95*(4), 434–455. doi: 10.1037/0033-295X.95.4.434

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.

Polyn, S. M., Erlikhman, G., & Kahana, M. J. (2011). Semantic cuing and the scale-insensitivity of recency and contiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 766–775.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156.

Polyn, S. M., & Sederberg, P. B. (2014). Brain rhythms in mental time travel. *NeuroImage*, *85*(2), 678–684.

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. doi: 10.1016/j.jml.2009.01.004

Pyc, M. A., & Rawson, K. A. (2010). *Why testing improves memory: Mediator effectiveness hypothesis* (Vol. 330) (No. 6002). American Association for the Advancement of Science. doi: 10.1126/science.1191465

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. doi: 10.1037/0033-295X.88.2.93

Racsmány, M., Bencze, D., Pajkossy, P., Szőllősi, , & Marián, M. (2021). Irrelevant background context decreases mnemonic discrimination and increases false memory. *Scientific Reports*, *11*(1), 6204. doi: 10.1038/s41598-021-85627-2

Rickard, T. C., & Pan, S. C. (2018). A dual memory theory of the testing effect. *Psychonomic Bulletin and Review*, *25*(3), 847–869. doi: 10.3758/s13423-017-1298-4

Roediger, H. L., Gallo, D. A., & Geraci, L. (2002). *Processing approaches to cognition: The impetus from the levels-of-processing framework.* doi: 10.1080/09658210224000144

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: basic research and implications for educational. *Perspectives on Psychological Science*, *1*(3), 181–210.

Roediger, H. L., & Karpicke, J. D. (2006b). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, *1*(3), 181–210. doi: 10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., & Karpicke, J. D. (2006c). Test-enhanced learning. *Psychological Science*, *17*(3), 249–255.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*(1), 45–48.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological bulletin*, *140*(6), 1432–1463. doi: 10.1037/a0037559

Rowland, C. A., & DeLosh, E. L. (2014). Benefits of testing for nontested information: Retrieval-induced facilitation of episodically bound material. *Psychonomic Bulletin and Review*, *21*(6), 1516–1523. doi: 10.3758/s13423-014-0625-2

Rowland, C. A., & DeLosh, E. L. (2015). Mnemonic benefits of retrieval practice at short retention intervals. *Memory*, *23*(3), 403–419. doi: 10.1080/09658211.2014.889710

Sederberg, P. B., Miller, J. F., Howard, M. W., & Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory \& Cognition*, *38*(6), 689–699.

Shimamura, A. P. (2011). Episodic retrieval and the cortical binding of relational activity. *Cognitive, Affective and Behavioral Neuroscience*, *11*(3), 277–291. doi: 10.3758/s13415-011-0031-4

Shin, Y. S., Masís-Obando, R., Keshavarzian, N., Dáve, R., & Norman, K. A. (2021). Context-dependent memory effects in two immersive virtual reality environments: On Mars and underwater. *Psychonomic Bulletin and Review*, *28*(2), 574–582. doi: 10.3758/s13423-020-01835-3

Siegel, L. L., & Kahana, M. J. (2014). A retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 755–764. doi: http://dx.doi.org/10.1037/a0035585

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342–353. doi: 10.3758/BF03197465

Spillers, G. J., & Unsworth, N. (2011). Variation in working memory capacity and temporal–contextual retrieval from episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1532–1539.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, *86*(2), 255–262.

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval questions and answers. *Experimental Psychology*, *56*(4), 252–257. doi: 10.1027/1618-3169.56.4.252

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., … Morris, R. G. M. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76–82.

Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science*, *2*(3), 67–70.

Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology*, *53*(1), 1–25. doi: 10.1146/annurev.psych.53.100901.135114

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*(3), 331–350. doi: 10.1037/0096-3445.132.3.331

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. doi: 10.1037/xlm0000379

Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory and Cognition*, *38*(8), 995–1008. doi: 10.3758/MC.38.8.995

Zhang, D., & Tullis, J. G. (2021). Personal reminders: Self-generated reminders boost memory more than normatively related ones. *Memory and Cognition*. doi: 10.3758/s13421-020-01120-7